杭州電子科技大学

硕 士 学 位 论 文

题　　目：基于 LBP-KNN 和 CNN-SVM 的人脸识别算法

研 究 生　Idelette Laure Kambi Beli (柯拉)

专　　业　信息与通信工程

指导教师　郭春生　副教授

完成日期　2018 年 3 月 14

杭州电子科技大学硕士学位论文

基于 LBP-KNN 和 CNN-SVM 的人脸识别算法

研 究 生： 柯拉

指导教师： 郭春生 副教授

2018 年 3 月 14 日

Dissertation Submitted to Hangzhou Dianzi University

for the Degree of Master

Enhancing Face Identification Based on LBP-KNN and CNN-SVM

Candidate: Idelette Laure Kambi Beli

Supervisor: Ass. Prof. Guo ChunSheng

March, 2018

杭州电子科技大学

学位论文原创性声明和使用授权说明

原创性声明

本人郑重声明：所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不含任何其他个人或集体已经发表或撰写过的作品或成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。

申请学位论文与资料若有不实之处，本人承担一切相关责任。

论文作者签名：　　　　　　　　　　　　日期：　　　年　　月　　日

学位论文使用授权说明

本人完全了解杭州电子科技大学关于保留和使用学位论文的规定，即：研究生在校攻读学位期间论文工作的知识产权单位属杭州电子科技大学。本人保证毕业离校后，发表论文或使用论文工作成果时署名单位仍然为杭州电子科技大学。学校有权保留送交论文的复印件，允许查阅和借阅论文；学校可以公布论文的全部或部分内容，可以允许采用影印、缩印或其它复制手段保存论文。（保密论文在解密后遵守此规定）

论文作者签名：　　　　　　　　　　　　日期：　　　年　　月　　日

指导教师签名：　　　　　　　　　　　　日期：　　　年　　月　　日

摘 要

人脸在我们社会交往中扮演着重要的角色，传递着我们自身信息。生物识别密码技术是一种 非常关键的安全技术，因其有着广泛的应用前景，在过去的几年里一直受到业内广泛的关注。人的面部表情有很多的变化（如：脸部老化、面部表情、明亮程度、不标准的姿势等），这些变化会导致脸部识别信息不准确，辨认身份能力较差。虽然人脸识别的技术上，已经有了很大的进展，同时也显示了非常精确的结果，但是在实际应用中，年龄不变的人脸识别仍然是系统应用中一个非常重要的挑战。

我们研究的目的是提供一种解决脸部识别问题的方法，这些问题收很多参数变化的影响，如姿势、明亮程度、年龄不变和面部表达等。为了解决这些问题，下一节将详细阐述不同的算法，来证明所提出模型的有效性。

为了证明在姿态变化、明亮程度和表达方面获取结果的可靠性，我们结合了两种算法：(a)鲁棒性 local binary pattern (LBP)，用于面部特征提取;(b) k-nearest neighbor (K-NN)进行图像分类。我们的实验已经在 CMU PIE (Carnegie Mellon University Pose, Illumination, and Expression) 的数据库上和 LFW (Labeled Faces in the Wild)数据集上进行了比对验证。结果证实我们提出的识别系统具有较高的性能，并提高了人脸相似性度量的值。

本研究对卷积神经网络(CNN)的深度学习和其有效的新用法进行了研究，特别是支持向量机(SVM)的方法，以解决年龄不变的人脸识别问题。我们把在不同时间段拍摄的主题的图像当作一个单一的集合，然后与其他主题的图像集合进行比较。使用卷积神经网络(CNN)去提取面部特征。在这里，我们发现使用 VGG-Face deep (神经网络)的一个 CNN 架构产生了高度的识别性和可互操作的特性，即使在不同的生物特征数据集(大的或小的，约束的或不约束的)中，也会对老化的变化产生强烈的影响。由于一个人的面部特征随着时间的推移而发生变化，这导致了大量的脸部信息发生了变化;因此，在提取的脸部信息特征时，要显示出高的跨类和低的内部类变化，否则使用子空间判别分类器的集合来对老化的数据集进行比较，必定会产生低泛化的错误。基于 SVM 的人脸识别方法实现了对图像的分类。对两种著名公共领域的老化数据集进行大量的实验研究发现:MORPH Album2 和 FGNET 证实了该方法的有效性。取得的结果表明，基于集合的识别性能优于基于单列的人脸识别方法。我们还发现，通过使用基于集合的识别，更容易被识别的是较年轻的主题，而不是较年轻的主题从较老的主题。

本研究的最终目的是在不同的参数（姿势、明亮程度、年龄不变和面部表达等）方面，呈现与研究活动相关的具体事实。介绍了过去十年中采用的主要方法，提出了一份综合的基准测试结果清单，最重要的是为今后的发展提供了路线，以便在人脸识别方面提供一些要求

和研究方向。

关键词：人脸识别; local binary pattern (LBP); k-nearest neighbor (K-NN); Support vector machine (SVM); convolutional neural networks (CNN), Deep Learning;

ABSTRACT

The human face plays an important role in our social interaction, conveying people's identity. Using the human face as a key to security, biometric passwords technology has received significant attention in the past several years due to its potential for a wide variety of applications. Faces can have many variations in appearance (aging, facial expression, illumination, inaccurate alignment and pose) which continue to cause poor ability to recognize identity. While very promising results and considerable progresses have been shown (made) on face recognition related problems, age-invariant face recognition still remains a major challenge in real world applications of face recognition systems with less research.

The purpose of our research work is to provide an approach that contributes to resolve face identification issues with large variations of parameters such as pose, illumination, age-invariant and expression. In order to resolve those problems, the next sections elaborate and explain carefully the different algorithms use to demonstrate the effectiveness of the proposed model.

For provable outcomes across pose, illumination, and expression, we combined two algorithms: (a) robustness local binary pattern (LBP), used for facial feature extractions; (b) k-nearest neighbor (K-NN) for image classifications. Our experiment has been conducted on the CMU PIE (Carnegie Mellon University Pose, Illumination, and Expression) face database and the LFW (Labeled Faces in the Wild) dataset. The proposed identification system shows higher performance, and also provides successful face similarity measures focus on feature extractions.

This study addition the novel use and effectiveness of Deep Learning in general and Convolutional Neural Networks (CNN), in particular with Support vector Machine (SVM) approach to face recognition subject for aging invariant. The images for each subject taken at various times are treated as a single set, which is then compared to sets of images belonging to other subjects. The facial features are extracted using a Convolutional Neural Networks characteristic of Deep Learning. Here, a CNN architecture using the VGG-Face deep (neural network) is found to produce highly discriminative and interoperable features that are robust to aging variations even across the different biometric datasets (large or small, constraint or unconstraint). The facial appearance of a person changes over time because of the aging process, which results in significant intra-class variations; therefore, the extracted features show high inter-class and low intra-class variability leading to low generalization errors on aging datasets using ensembles of subspace discriminant classifiers. SVM-based face recognition method then perfectly achieve the image classifications. Extensive experiments are conducted on two well-known public domain face aging datasets: MORPH Album2 and FGNET show the effectiveness of the proposed approach. The obtain outcomes show that, it is easier to recognize older subjects from younger ones rather than younger subjects from older ones.

The ultimate aims of this study are to present concrete facts related to research activities in facial identification across several parameters such as pose, illumination, age-invariant and expression. The adopted main methodologies during the past decade are presented, a comprehensive list of benchmark results and most importantly provide roadmaps for future trends are also highlighted, in order to give some requirements and research directions in facial recognition.

Keywords: Face Recognition; Face Identification; Local Binary Pattern (LBP); K-Nearest Neighbor (K-NN); Support Vector Machine (SVM); Convolutional Neural Networks (CNN), Deep Learning.

目　录

Chapter 1 Introduction

1.1 Problem Definition, Motivation and Objectives

Object recognition is a computer technology related to computer vision and image processing that deals with detecting and identifying humans, buildings, cars, etc., in digital images and video sequences. It is a huge domain including face recognition which basically has two different modes: verification and identification [1]. In this thesis, we focus on the identification basic mode.

A face is a typical multidimensional structure and needs good computational analysis for recognition. The overall problem is to be able to accurately recognize a person's identity and take some actions based on the outcome of the recognition process. Recognizing a person's identity is important mainly for security reasons, but it could also be used to obtain quick access to medical, criminal, or any type of records. Solving this problem is important because it could allow people to take preventive action, provide better service in the case of a doctor appointment, allow users access to a secure area, and so forth.

Face identification is the process of identifying a person in a digital image or video, and showing their authentication identity. Identification is a one-to-many matching process that compares a query face image against all the template images inside the face database in order to determine the identity of the query face. Identification mode allows both positive and negative recognition outcomes, but the results are much more computationally costly if the template database is large [2, 3]. Now, our goal is to determine which person inside the gallery—if any—is represented by the query face. More precisely, when a particular query image is submitted to the recognition system, the resulting normal map is compressed in order to compute its feature indexes, which are subsequently used to reduce the search to a cluster of similar normal maps selected through a visit in the k-d-tree [4].

Motivation

Computer vision is the ability of machines to see and understand what is in their surroundings. This field contains some methods for acquiring, processing and analyzing the images; those methods can be able to extract important information used by artificial systems. Most recently, in computer vision a lot of important research are conducted, especially in its major sub-domains such as object recognition, motion analysis or scene reconstruction. Object recognition plays an important role in computer vision. It is indispensable for many applications in the area of autonomous systems or industrial control and so on. Recently face recognition which is a one problem of object recognition, has started to receive a blooming attention and interest from the scientific community as well as from the general public (society).

Humans recognize a huge number of objects (faces, cars, buildings, toys…) in images with little effort, even when the image of the objects may vary in different viewpoints. In many different sizes / scales or even when they are translated or rotated. Objects can even be recognized when they are partially obstructed (blurred) from view. This task is still a challenge for computer vision systems. Recognition remains challenging in large part due to significant variations exhibited by real-world images.

Face recognition is a task that humans perform routinely and effortlessly in our daily lives. Wide availability of powerful and low-cost desktop and embedded computing systems has created an enormous interest in automatic processing of digital images in a variety of applications, including biometric authentication, surveillance, human computer interaction, and multimedia management. One of the reasons that face recognition has attracted so much research attention and sustained development over the past 30 years is its great potential in numerous government and commercial applications. In 1995, Chellappa et al. [5] listed a small number of applications of face recognition technology and described their advantages and disadvantages. However, they did not analyze any system deployed in real applications. Even the more recent review [6], where the set of potential applications has been grouped into five categories, did not conduct such an analysis. In 1997, at least 25 face recognition systems from 13 companies were available [7]. Since then, the numbers of face recognition systems and commercial enterprises have greatly increased owing to the emergence of many new application areas, further improvement of the face recognition technologies, and increased affordability of the systems.

Objectives

In the past several years, academia and industry have developed many research works and practical approaches to overcome face recognition issues, specifically in pattern recognition and computer vision domains [8]. Facial recognition is a difficult problem due to the morphology of the face that can vary easily under the influence of many factors, such as pose, illumination, expression and age invariant [9]. In addition, faces have similar form and the same local parts (eyes, cheekbones, nose, lips, etc.). Therefore, to enhance the ability of a system to identify facial images, we need to apply an efficient algorithm that can describe the similarity representation and distinctive classification properties of diverse subject images.

As we mentioned above, the couple of local binary patterns (LBP) and k-nearest neighbor (K-NN) (resolving pose, illumination and expression challenges) on the one hand, and the couple of convolutional neural network (CNN) and support vector machine (SVM) (resolving age invariant challenge) on the other hand are among the famous proposed solutions available today.

For a decade, LBP was only used for texture classification; now it is also widely used to solve some of the common face recognition issues. LBP has many important properties, such as its robustness against any monotonic transformation of the gray scale, and its computational simplicity, which makes it possible to analyze images in challenging real-time settings [10]. According to several studies [14, 16, 17] face recognition using the LBP method provides very good results, both in terms of speed and discrimination performance. A face image is first divided into small regions from which LBP histograms are extracted and then concatenated into a single feature vector. This feature vector forms an efficient representation of the face and can be used to measure similarities between images. The report describes the principles of the method and how exactly it can be implemented to perform face recognition.

K-nearest-neighbor (KNN) has been widely used in classification problems. KNN algorithm is then a method for classifying objects based on closest training examples in the feature vector; so, an object is well classified by a majority vote of its neighbors. The greater accuracy of k-nearest neighbor in image classification problems is highlighted here; it is commonly used for its easier interpretation and low calculation time [47,48]. KNN is based on a distance function that measures the difference or similarity between two instances; the standard Euclidean distance d(x, y) will be used for achieving the classification phase. However, the main aim of LBP and K-NN in this work is to extract features and classify different LBP histograms, respectively, in order to ensure good matching between the extracted features histograms and provide a greater identification rate.

In this study, we propose the use of deep Convolutional Neural Networks to resolve the age-invariant deep face features. Specifically, we extract the age-invariant deep features from convolutional features by a carefully designed fully connected layer. This paper mostly copes with aging variation at the feature extraction phase and leverages a deep learning approach for automatic feature extraction using cascaded convolutional neural networks (CNN). Automatic feature extraction using CNN [66 - 68] can provide highly interoperable descriptors that are robust to aging variations and add much flexibility when deployed and used in biometric systems. At the feature extraction phase, the goal is to derive image descriptors that are robust to intra-class aging variation. Inspired by the belief that the face image is a combination of an age-specific component and an identity-specific component, we expect that the resulting deep feature reduces the variations caused by aging process as much as possible. Ideally, we want the feature containing only identity-related component to perform well in age invariant face recognition.

To achieve the age-invariant face recognition, we base our method on the machine learning system of Support Vector Machines (SVMs) classifier. Recently, more sophisticated classifiers, such as support vector machines (SVM) [70, 71], have been shown to be able to further improve the classification performance of some algorithms subspace features. Given any two classes of vectors, the aim of support vector machines is to find one hyperplane to separate the two classes of vectors so that the distance from the hyperplane to the closest vectors of both classes is the maximum. The hyperplane is known as the optimal separating hyperplane. SVM excels at two-class recognition problem and outperform many other linear and non-linear classifiers. SVM classifier is subsequently used to recognize each subject within different face age.

1.2 Ethic and Society Implications

As we mentioned above, face recognition has recently received a blooming attention and interest from the scientific community as well as from the general public or society. The interest from the society is mostly due to the recent terrorism and criminal events over the world, which has increased the demand for useful security systems policy. Facial recognition applications are far from limited to security systems as described above. Face recognition systems identify people by their face images. In contrast to traditional identification systems, face recognition systems establish the presence of an authorized person rather than just checking whether a valid identification (ID) or key is being used or whether the user knows the secret personal identification numbers (PINs) or passwords. The security advantages of using biometrics to check identification are as follows. It eliminates the misuse of lost or stolen cards, and in certain applications it allows PINs to be replaced with biometric characteristics, which makes financial and computer access applications more convenient and secure. In addition, in situations where access control to buildings or rooms is automated, operators may also benefit from improved efficiency. Face recognition systems are already in use today, especially in small database applications. In the future, the targeted face ID applications will be large-scale applications such as e-commerce, student ID, digital driver licenses, and even national ID.

As one of the most nonintrusive biometrics, face recognition technology is becoming ever closer to people’s daily lives. Evidence of this is that in 2000 the International Civil Aviation Organization endorsed facial recognition as the most suitable biometrics for air travel [11]. Face recognition has a large-scale applications still face a number of challenges; in the table 1.1 we grouped them into 10 categories. These 10 categories are neither exclusive nor exhaustive, and for each category some of the example applications are also listed. The last category inside the table, called “Others,” includes future applications and some current applications.

Table 1.1. Different applications categories of Face recognition.

Category	Examples of Application Scenarios
Face ID	Driver licenses, entitlement programs, immigration, national ID, passports, voter registration, welfare registration
Access Control	Border-crossing control, facility access, vehicle access, smart kiosk and ATM, computer access, computer program access, computer network access, online transactions access, long distance learning access, online examination access, online database access.
Security	Terrorist alert, secure flight boarding systems, stadium audience, scanning, computer security, computer application security, database security, file encryption, intranet security, Internet security, medical records, secure trading terminals.
Surveillance	Advance video surveillance, nuclear plant surveillance, park surveillance, neighborhood watch, power grid surveillance, CCTV

	control, portal control.
Smart Cards	Stored value security, user authentication.
Law Enforcement	Crime stopping and suspect alert, shoplifter recognition, suspect tracking and investigation, suspect background check, identifying cheats and casino undesirables, post-event analysis, welfare fraud, criminal face retrieval and recognition.
Face Databases	Face indexing and retrieval, automatic face labeling, face classification.
Multimedia Management	Face-based search, face-based video segmentation and summarization, event detection.
Human Computer Interaction (HCI)	Interactive gaming, proactive computing.
Others	Antique photo verification, very low bit-rate image & video transmission, etc…

1.3 Literature Survey

1.3.1 Pose, Illumination and Expression Face Recognition

Over the past decades, there have been many studies and algorithms proposed to deal with face identification issues. Basically, the identification face is marked by similarity; authors in [10] measured the similarity between entire faces of multiple identities via Doppelganger List. It is claimed that the direct comparisons between faces are required only in similar imaging conditions, where they are actually feasible and informative. In the same way, Madeena et al. [12] presented a novel normalization method to obtain illumination invariance. The proposed model can recognize face images regardless of the face variations using a small number of features.

In [13], Sandra Mau et al. proposed a quick and widely applicable approach for converting biometric identification match scores to probabilistic confidence scores, resulting in increased discrimination accuracy. This approach works on 1-to-N matching of a face recognition system and builds on a confidence scoring approach for binomial distributions resulting from Hamming distances (commonly used in iris recognition).

In 2015, Pradip Panchal et al. [14] proposed Laplacian of Gaussian (LoG) and local binary pattern as face recognition solutions. In this approach, the extracted features of each face region are enhanced using LoG. In fact, the main purpose of LoG is to make the query image more enhanced and noise free. In our opinion, authors should use a Gaussian filter before applying LoG, since the combination of these two algorithms would provide better results than the ones obtained. Following the same way, authors in [15, 16] also used LBP technique. In [15], the face recognition performance of LBP is investigated under different facial expressions, which are anger, disgust, fear, happiness, sadness, and surprise. Facial expression deformations are challenging for a robust face recognition system; thus, the study gives an idea about using LBP features to expression invariant. Further, authors in [16] implemented LBP and SSR (single scale retinex) algorithms for recognizing face images. In this work, lighting changes were normalized and the illumination factor from the actual image was removed by implementing the SSR algorithm. Then, the LBP feature extraction histograms could correctly match with the most similar face inside the database. The authors claimed that applying SSR and LBP algorithms gave powerful performance for illumination variations in their face recognition system.

Bilel Ameur et al. [17] proposed an approach where face recognition performance is significantly improved by combining Gabor wavelet and LBP for features extraction and, K-NN and SRC for classification. The best results are obtained in terms of time consumption and recognition rate; the proposed work also proved that the system efficiency depends on the size of the reduced vector obtained by the dimension reduction technique. However, Dhriti et al. [18] revealed the higher performance and accuracy of K-NN in classification images. In the same way as [18], authors in [19] used K-NN as the main classification technique and bagging as the wrapping classification method. Based on the powerful obtained outcomes, the proposed model demonstrated the performance and capabilities of K-NN to classify images.

Nowadays, research is not only focalized on face recognition in constrained environments; many authors also are trying to resolve face recognition in unconstrained environments.

The works [20, 21, 22] proposed a convolutional neural network (CNN) as a solution of the face recognition problem in unconstrained environments. Deep learning provides much more powerful capabilities to handle two types of variations; it is essential to learn such features by using two supervisory signals simultaneously (i.e., the face identification and verification signals), and the learned features are referred to as Deep IDentification-verification features (DeepID2) [20]. The paper showed that the effect of the face identification and verification supervisory signals on deep feature representation coincide with the two aspects of constructing ideal features for face recognition (i.e., increasing inter-personal variations and reducing intra-personal variations), and the combination of the two supervisory signals led to significantly better features than either one of them individually. Guosheng Hu et al. [21] presented a rigorous empirical evaluation of CNN based on face recognition systems. Authors quantitatively evaluated the impact of different CNN architectures and implementation choices on face recognition performances on common ground. The work [22] proposed a new supervision signal called center loss for face recognition task; the proposed center loss is used to improve the discriminative power of the deeply learned features. Combining the center loss with the softmax loss to jointly supervise the learning of CNNs, the discriminative power of the deeply learned features can be highly enhanced for robust face recognition.

1.3.2 Age Invariant Face Recognition

Most existing age-related works for face image analysis focus on age estimation [23]–[25] and age simulation [26]–[28]. In recent years, researchers have started to focus on face recognition across age. One of the approaches is to construct 2D or 3D aging models [28]–[30] to reduce the age variation in face matching. Authors in [30] made an assumption with similar faces age in similar ways for all individuals. They introduced Aging pattern Subspace (AGES) and constructed a representative subspace by defining a sequence of a particular individuals face images sorted in time order. For a perfect construction with a minimum reconstruction error, they used a projection in subspace with the help of proper aging pattern for a previously unseen face image. Thus they would be able to determine the age of the image from the aging pattern. However, models usually rely on strong parametric assumptions, accurate age estimation, as well as clean training data, and therefore they do not work well in unconstrained environments. In [31], Wu *et al.* proposed to use a relative craniofacial growth model to model the face shapes for cross-age face recognition and it yields good performance on FG-NET dataset. However, their approach requires age information to predict the new shapes, which is not always available. Some other works focus on discriminative approaches. The work [32] used gradient orientation pyramid (GOPs) with SVM for face verification across age progression (to verify faces with large age differences). *The* [33] used multi-feature discriminant analysis for close-set face identification. Gong *et al.* [34] proposed to separate the feature into identity and age components using hidden factor analysis. They propose a novel coding framework, which encoding the low-level feature of a face image with an age-invariant reference space.

Shape and texture play a vital role and draw intensive interest in the fields of authentication systems and facial recognition [35]. They offered useful features for classifying face images related to the same subject and also discriminating between face images of different subjects. While facial aging often appears as changes of shape in human faces because of the craniofacial growth from babyhood to the teenage years [36], they appear more commonly in the form of textural variations, such as changes in skin color, wrinkles and other skin artefacts, during adulthood years. Zhao et al. [37] developed a shape transformation model to capture the delicate deformation of facial features that are exaggerated by aging. Their model implicitly accommodates the physical features and geometric trends of human face muscles. Subsequently, they developed a texture transformation based on the image gradient function that identified wrinkles on the face and other changes on the skin, regularly experienced at diverse ages. However, the proposed face aging model did not account for facial hair, and therefore, could not address hair loss. The work [38] constructed an aging model for individuals (probabilistic model) based on a Gaussian mixture model that combined the texture and shape information. The system performed well when tested on individuals in the age range between 18 and 69, but when the subject was less than 18 years old, the cumulative accuracy of the system degraded. Lanitis et al. [29] developed models that were statistical to perform facial aging simulation. They developed a model that combined information of both intensity and shape to characterize the images of faces. Since the system was trained using a comparatively small number of images, it learned to simulate age effects only in the ways exhibited in the narrow training set. Consequently, the system failed when the subject appeared in a new image representing a different aging pattern than the aging patterns of the subjects in the database. Moreover, they used local aging models although the process of aging can be different for different age groups.

The methods SIFT in [39], MLBP in [40] with a random sampling based fusion framework are combined in [33] to improve age-invariant face recognition performance; here, [41] is used as features and a variation of random subspace LDA approach (RS-LDA) [42] is then used for the classification. The [43] proposes a probabilistic model with two latent factors: an identity factor that is age-invariant and an age factor affected by the aging process, then the observed appearance can be modeled as a combination of the components generated based on these factors. Recently, [44] proposes a new deep Convolutional Neural Network (CNN) model for age-invariant face verification, authors also introduce two tricks to overcome insufficient memory capacity issue and to reduce computational cost. The work [44] develops a new maximum entropy feature descriptor (MEFA) that encodes the microstructure of facial images into a set of discrete codes in terms of maximum entropy and a new matching method named identity factor analysis (IFA) which is used to estimate the probability that two faces have the same underlying identity. The methods [44] and [45] have notably improved the performance of age-invariant face recognition recently.

Sethuram et al. [46] and Lanitis et al. [47] used the Active Appearance Models (AAMs), but authors in [46] combined it with Support Vector Machines (SVMs) and Monte-Carlo simulation. The combination of these techniques was to obtain a high accuracy aging model; for its accomplishment, two types of experiments were reported. In the first experiment they showed, that when the probe face progresses in age, the face recognition rate decreases. In the second experiment, they artificially aged the probe face in order to augment its age and match it with one face who has the same age present inside the gallery. Finally 32 percent rank1 accuracy was obtained by using PCA based face recognizer. In [47], they used a statistical face model for studying the age estimation problem. Firstly, they extracted AAM parameters from facial images which were marked with 68 points, then a Genetic Algorithm was then applied to build and optimize an aging function.

Yan et al. [48] introduced the concept of coordinate patches and GMMs. They used images of individuals and encoded them as ensembles of overlapped spatially flexible patches (SFPs). These SFPs were modelled with Gaussian Mixture Models (GMMs), then the local features were extracted by using the 2D discrete cosine transform (DCT). The local features were integrated with the patches and an estimate of the age for the individual was then obtained based on a maximum likelihood estimator obtained from total spatially flexible patches (SFPs) of the hypothetical age.

1.4 Resume

This section enumerates our approach and the different existing methods of face recognition. These methods are among the famous algorithms used to resolve the problem of pose, illumination, expression variations and age invariant. There are several algorithms available in the literature that can solve enumerate problems. Most of these papers address how to automatically extract features from face images and classify them, and focus on how to enhance the face recognition system. Researchers adopt different ways to automatically extract features and classify the images; some use a universal state-of-art algorithms and some use a specific new algorithms for the resolution.

The thesis is structured as follows. Chapter 2 deals with the fundamental concept of face recognition. Chapters 3 and 4 describe the theory of our proposed framework and the practical implementation and evaluation of the obtained results. Finally, we conclude the work in Chapter 5.

Notice that, our experiments have been conducted on an Intel Core i5-2430M CPU 2.40 GHz Windows 10 machine with 6 GB memory, and implemented in MATLAB R2016b.

Chapter 2 Face Recognition Fundamental

2.1 History of Face Recognition

Face Recognition is the process where the brain recognizes, understands and interprets the human face. The face is essential for the identification of others; the face expresses and reveals significant social information like intention, attentiveness, and communication. The first real attempts to develop semi-automated facial recognition systems began in the late 1960's and early 1970's, and were based on geometrical information.

The first semi-automated facial recognition programs were created by Woody Bledsoe, Helen Chan Wolf, and Charles Bisson in back to the 1960's [49]. Their programs required the administrator to locate features such as the eyes, ears, nose, and mouth on the photograph. It then calculated distances and ratios to a common reference point which was then compared to reference data.

In early 1970, Goldstein et al [50] used 21 specific subjective markers, such as hair color and lip thickness, to automate the recognition. The measurements and locations needed to be manually computed, causing the program to require a lot of labor time.

The number of research with faces used as stimuli has increased dramatically over the past decades. Kirby and Sirovich [51] (1988) applied principle component analysis, a standard linear algebra technique, to the face recognition problem. In 1991, Turk and Pentland [52] bring a novelty in Kirby's technology; they discovered that while using the Eigen faces techniques, the residual error could be used to detect faces in images. Although the approach was constrained by environmental factors, it created significant interest in furthering development of automated face recognition technologies.

However, for encouraging the development of face recognition algorithms and technology, the Face Recognition Technology Evaluation (FERET) [53] was sponsored by the Defense Advanced Research Products Agency (DARPA) from 1993 to 1997. FERET propelled face recognition from its infancy to a market of commercial products. Another evaluations built upon the work of FERET named the Face Recognition Vendor Tests (FRVT) [54] were performed in 2000, 2002, and 2006. The two goals of FRVT were: to assess the capabilities of commercially available facial recognition systems and to educate the public on how to properly present and analyze results.

2.2 Face Recognition System

Face recognition is a task that humans perform routinely and effortlessly in their daily lives. It is also, the ability of a computer to scan, store and recognize human faces for use in identifying people. Face recognition technology is the least intrusive and fastest biometric technology, it works with the most obvious individual identifier – the human face. Facial recognition analyzes the characteristics of a person's face images input through a digital image or a video frame from a video source. It measures the overall facial structure, including distances between eyes, nose, mouth, jaws, cheekbones and so on. These measurements are retained in a database and used as a comparison when a user stands before the camera. Every face has distinguishable landmarks, and it can have approximately 80 nodal points. Some of these nodal points are: Distance between the eyes，Width of the nose，Depth of the eye sockets，The shape of the cheekbones，The length of the jaw line, etc.… The following figure shows an example of 68 facial landmarks.

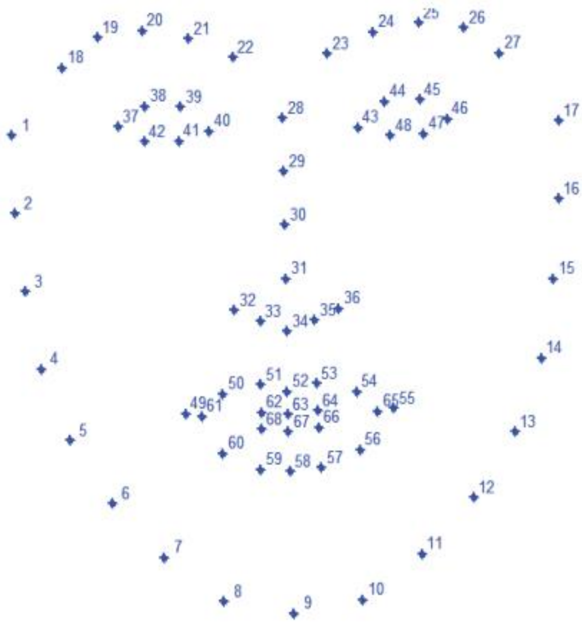


Figure 2.1. The 68 landmarks of the human face.

Facial recognition system can detect faces in images, quantify their features, and then match them against stored templates in a database. Face recognition measures and matches the unique characteristics for the purposes of its different two modes, which are identification and verification or authentication.

Verification mode seeks to answer the question “Am I who I say I am?” Also called one-to-one matching system, it compares a query face image against a template face image whose identity is being claimed. Under a verification system, an individual presents himself or herself as a specific person; The system checks his or her biometric against a biometric profile that already exists in the database linked to that person's file in order to find a match.

Identification mode seeks to answer the question "Who I am?" or "Who is this person?" Also called one-to-many matching system, it compares a query face image against all the template images in the database to determine the identity of the query face. Identification systems are different from verification systems because an identification system seeks to identify an unknown person, or unknown biometric. Verification systems only need to compare the presented person or biometric to a person or biometric reference stored in the database (system), so that they can generate results more quickly and are more accurate than identification systems, even when the size of the database increases. The purpose of our thesis is not a comparative study between those systems or modes, it is the evaluation of identification mode through the different factors which influence the face.

2.3 Different Face recognition challenges

This section elucidates the recent issues and challenges occurred in face recognition which significantly alter human face.

2.3.1 Pose Variation issues

It is a head's movements which can be described by the egocentric rotation angles that is pitch, roll and yaw, or a camera changing point of views could lead to substantial changes in face appearance and/or shape and generate intra−subject face's variations. Figure 2.2 illustrates some examples of variation poses in LFW dataset.

Figure 2.2. Illustration of pose variations around egocentric angles.

Some proposed solutions to the pose problem depend greatly on the particular application. Pentland et al [55] used view-based approach as an example; their method consisted to compare faces in a similar pose using a set of pose-specific classifiers or similarly measures. The problem with this approach is that, it requires multiple views of each gallery face and it requires good pose estimation of the query (probe) face to know which pose-specific classifier and gallery image to use for comparison.

Usually, the training data used by face recognition systems are frontal view face images of individuals. Frontal view images contain more specific information of a face than profile or other pose angle images. The problem appears when the system has to recognize a rotated face using this frontal view training data. User need together multiple views of an individual in a face database. The pose problem has been divided into three categories: (a) The simple case with small rotation angle; (b) The most commonly addressed case, when there is a set of training image pairs, frontal and rotated images; and (c) The most difficult case, when training image pairs are not. Therefore, those different variations making automated face recognition across pose a difficult task.

2.3.2 Illumination Variation issues

Besides face pose, illumination is the next most significant factor affecting the appearance of faces. Illumination variation has enormously complex effects on the object image. In the image of a familiar face, changing the direction of illumination leads to shifts in the location and shape of shadows, changes in highlights, and reversal of contrast gradients. Every time many experiences show a remarkably good outcomes in face recognition under the lighting variations.

Ambient lighting changes greatly within and between days and among indoor (controlled) and outdoor (uncontrolled) environments. Due to the 3D structure of the face, a direct lighting source can cast strong shadows that accentuate or diminish certain facial features.

As a fundamental problem in image understanding literature illumination problem is generally quite difficult and has been receiving consistent attentions. For face recognition, many good approaches have been proposed utilizing the domain knowledge that is all faces belong to one face class. These approaches can be broadly divided into four types [56]: (1) heuristic methods including discarding the leading principal components; (2) image comparison methods where various image representations and distance measures are applied; (3) class based methods where multiple images of one face under a fixed pose but different lighting conditions are available; and (4) model-based approaches where 3D models are employed.

Mostly differences in images induced by illumination are larger than differences between individuals, it causing the misclassification of the input image identity inside the systems based on comparing images. The following figure illustrates different lighting variations.

Figure 2.3. Different lighting variations of one subject of CMU PIE Database.

2.3.3 Expression Variation issues

Facial expressions are the facial changes in response to a person's internal emotional states, intentions, or social communications. Facial emotion are categorized into six, namely: Anger ("combination of lowered brow, upper lid raiser, tighten lid, tightened lips"), Disgust ("combination of nose wrinkle, lip corner depressor, lower lips depressor"), Fear ("combination of inner/outer raised brow, lowered brow, upper lid raiser, tighten lid, stretched lips, jaw drop"), Happy ("combination of cheek raiser, lip corner puller"), Sad ("combination of inner brow raiser, brow lowered, lip corner depressor"), Surprise ("combination of inner/outer raised brow, upper lid raiser, jaw drop").

The changes in the facial expression can be either based on minor deformations in wrinkles/bulges or based on major deformations (in eyes, eye-brow, mouth, nose, etc.). Some of the feature extraction techniques and facial expression categorization includes, Geometric based and Appearance based, Action Unit (AU) of individual/group of muscles and Non-AU based, Local versus Holistic and so on. In geometric based methods, the position and deformation/displacement information of the facial components are considered, whereas appearance methods simply apply a filter. In the same order, during the 1960 and 1970, one system for parameterizing minimal facial actions was developed by psychologists trying to analyze facial expressions. The system was called the Facial Action Coding System (FACS) [57] and its purpose was to describe each facial expression as a combination of around 50 Action Units (AUs), where each AU represented the activation of one facial muscle. The FACS has been a popular tool not only for psychology studies but also for computerized facial modeling; the following figure represents the Face Animation Parameter Units (FAPU).

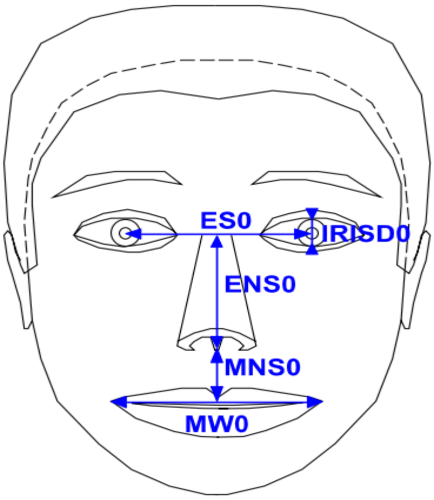


Figure 2.4. Face Animation Parameter Units (FAPU).

Following figure show the different expressions of one subject in LFW database.

Figure 2.5. Different expressions of subject.

2.3.4 Age Invariant issues

Age invariant face recognition (AIFR) is an important and a great challenge area of face recognition research. AIFR is an emerging research topic and is useful in a number of practical applications; in law enforcement for example, finding missing children and identifying or looking for criminal suspects based on their mug shots, under the identity requires recognizing photos across ages. The difficulty of this problem, to a great extent, arises from the fact that the face appearance of a person is subject to remarkable change caused by the aging process over time. First, the available face aging databases are usually collected from scanned images in different poses, illumination and expression. Meanwhile most face modeling methods require having face images with frontal pose, normal illumination and neutral expression to get the best fit results. Additionally, in order to have an exact model to represent the aging process, systems have to use a huge number of training images that are usually inefficient for the currently limited face aging databases. Second, forensic scientists proved that human face aging strongly depends on ethnicity and genders [58]. Although human faces have the same general manner under aging, each ethnic and gender group has distinct characteristics in face aging. Therefore, it is insufficient to assume that similar faces age in similar ways for each and every individual.

In recent years, some researches on age related face image analysis have been studied. Therefore, there is less research on age-invariant face recognition as compared to age modeling and age estimation; then the work that explicitly tackles age invariant face recognition is limited. Most existing methods on age invariant face recognition roughly fall into two categories: generative approaches and the discriminative approaches. Generative methods try to synthesis face images that match the target age before recognition. They try to construct a 2-D or 3-D generative model to compensate for the aging process in face matching. The discriminative models use facial features that are insensitive to age progression for achieving age-invariant face recognition. However, this approach cannot achieve a significant performance level if the two faces to be compared have a large age difference.

In definitive, matching facial images across ages is often necessary in real world applications; and as we explained below, it is still remained a great challenge problem. The faces of the same person can exhibit substantially different appearance at different ages, thus aggravating the difficulties. On the other hand, facial images of the same person also contain intrinsic features that are relatively stable across ages. Significant progress has been made in face recognition but AIFR still remains a major challenge in real world applications such as face recognition systems, in which age-related face image analysis has most traction. A typical AIFR approach is to use face modeling to synthesize and render face images to the same age as the gallery image prior to recognition. However, due to the strong parametric assumptions and the complexity of the algorithm, these methods are computationally expensive and the results are often unstable for real-world face recognition. Following figure represents two subject across different ages:

Figure 2.6. First range: Same Individual at different ages from FG-Net Database [78], Second range: Same Individual at different ages from MORPH Database [77].

2.3.5 Other related issues: Plastic/Cosmetic Surgery and Makeup

Nowadays, due to advances in technology, there are new emerging challenges for which the performance of face recognition systems degrades and plastic/cosmetic surgery and makeup are among of them.

Facial plastic surgery [80] is generally used for correcting facial feature anomalies or improving facial appearance, for example, removing birth marks, moles, scars and correcting disfiguring defects. These surgical procedures prove beneficial for some patients who suffering from structural or functional impairment of facial features, but some patients also misused such procedures in order to conceal their identities with the intent to commit fraud or evade law enforcement. Most popular facial plastic surgery is Rhinoplasty i.e. nasal surgery. It is used to re-enact the nose in cases involving birth defects, accidents where nose bones are dented and also to cure breathing problems caused due to the nasal structure. There is also the Blepharoplasty (eyelid surgery) used to remodel eyelids when over-growth of skin tissues on the eyelid causes vision problem.

Recent research has demonstrated the negative impact of makeup on automated face recognition. Makeup [81] is typically used to enhance or alter the appearance of an individual’s face. It has become a daily necessity for many human being. The cosmetic industry has developed a number of products, which can be broadly categorized as skin, eye or lip makeup. Makeup poses a challenge to automated face recognition due to its potential to substantially alter the facial appearance. For example, it changes the perceived facial shape and appearance, modifies contrast levels in the mouth and eye region, and alters skin texture. Such modifications can lead to large intra-class variations, resulting in false non-matches, where a subject’s face is not successfully recognized.

Therefore, we can assert that plastic surgery and makeup are the new and equally serious covariate that have sprung from our evolving cultures. They are also entwined with many ethical dilemmas in their use in biometric technology.

2.4 Resume

As we have seen along this section, the face recognition history has been revealed. It described the evolution of face recognition in our society and also detailed carefully its different issues and challenges. Face recognition is a process that recognizes any human face in an image. Face recognition is not a simple process as it is governed by lot of external and internal factors which affect the recognition system. Even if a subject's face is stored in the database, a disguise or even a minor change in appearance, like wearing glasses, blinking or closing eyes, different face position, wearing or growing a mustache and so on can often fool the system. Even an unusual facial expression and one individual at different age can confuse the system; different illuminations, plastic surgery also deform the faces significantly.

Chapter 3 Face Recognition using Local Binary Pattern And K Nearest Neighbor

This chapter discusses the first research methodology (LBP, KNN, and Gaussian Filter) and design used in the study including strategies. Each section below is clearly developed in order to explain each phase of our proposed algorithm. The choice of combination LBP-KNN and Gaussian filter was based on the faster execution time, the outcome results and especially the classification method. During our research, we have remarked that the classification method does not depend only on training data size but also on the extraction features algorithm (LBP). So, to better evaluate the performance of the proposed LBP features, our model uses the simple Euclidean Distance of the K Nearest Neighbor rule as the classifier.

3.1 Local Binary Pattern (LBP)

Recently, LBP-based approaches have been proposed to solve certain face recognition problems, such as illumination and expression variations. First proposed by Ojala et al. in 1996 [59], the LBP operator is a signified robust method of texture description; it is described as an ordered set of binary comparisons of pixel intensities between the center pixel and its surrounding pixels. LBP was originally defined for 3×3 neighborhoods, giving 8 bit codes based on the 8 pixels around the central one and representing the outcome as a binary number. LBP is derived for a specific pixel neighborhood radius R by comparing the intensities of P discrete circular sample points to the intensity of the center pixel (clockwise, counterclockwise), starting from a certain angle (as shown in Figure 3.1a). The comparison determines whether the corresponding location in the LBP of length M is “1” or “0”. The value “1” is assigned if the center pixel intensity is greater than or equal to the sample pixel intensity, and “0” otherwise (most commonly used P=8 with R=1); however, other values of the radius and sample numbers can be used (shown in 3.1b). If a sample point is located between pixels, the intensity value used for comparison can be determined by bilinear interpolation.

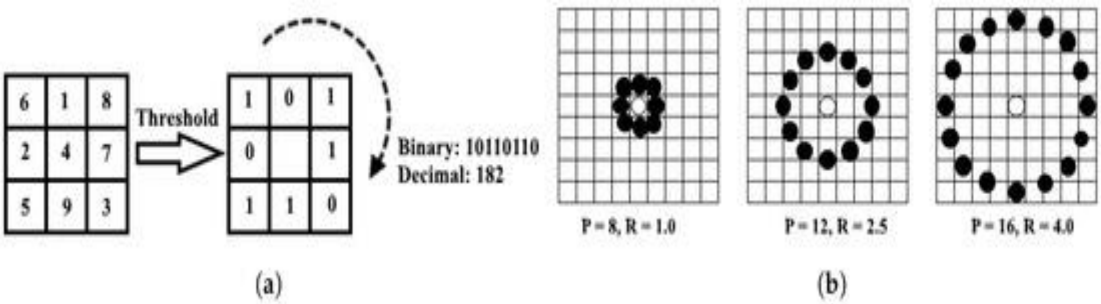


Figure 3.1. (a) The original local binary pattern (LBP) operator; (b) Circular neighbor-set for three

different values of P, R.

Uniform LBP is an important case of LBP. An LBP descriptor is called uniform if it contains at most two circular bitwise 0–1 and 1–0 transitions. Since the allotted binary string needs to be considered as circular, the occurrence of only one transition is not possible; this means a uniform pattern has no transitions or two transitions. For instance, 00,000,000, 11,111,111, 11,011,111, and 10,001,111 are uniform binary patterns with zero bitwise transitions and two bitwise transitions, respectively. P(P−1)+3 is a possible combination for uniform patterns with two bitwise transitions; it makes the work very easy compared to non-uniform patterns which have 2^p possible combinations. Instead of non-uniform binary patterns, there are two reasons for selecting uniform patterns. First, uniform LBP saves memory; for example, the number of possible patterns for a neighborhood of 8 pixels is 256 for standard LBP (non-uniform) and 59 for LBP^{u2} (u2 stands for using only uniform patterns), for 16 (interpolated) pixels is 65,536 for standard LBP and 243 for LBP^{u2}. The second reason is that it detects only the most important and useful features in the preprocessed images, such as corners, spots, edges, and line ends (Figure 3.2); thus, it can generate a more precise recognition rate and makes the process simpler and more effective.

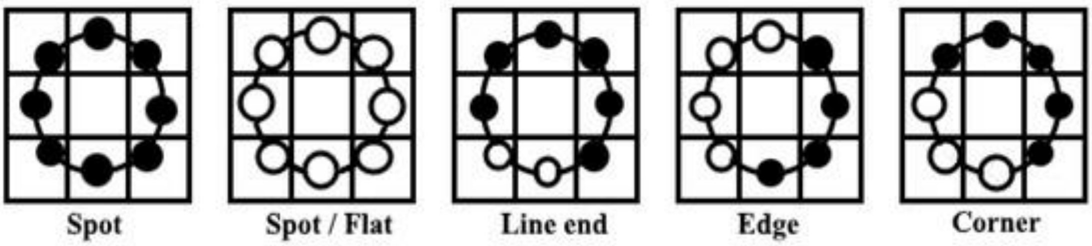


Figure 3.2. Different texture primitives detected by LBP^{u2}.

LBP Mathematical Approach

There are several methods for extracting unique and useful features from face images to perform face recognition; local binary pattern (LBP) is among the most popular ones, and it is also the most efficient and newest algorithm in that research field. It provides very good results in terms of both speed and discrimination performance.

The facial image texture is divided into several small blocks, from which the feature histogram (of each region) is constructed separately; therefore, the LBP histogram of each block will be combined to obtain a concatenated vector (a global histogram of the face). The similarity (distance) can then be measured by using the global histogram of different images. The global histogram of a facial image $f_l(\mathrm{x}, \mathrm{y})$ is represented by:

$$H_{i,j} = \sum_{x,y} I(f_l(x,y) = i) \tag{3.1}$$

Where $H_{i,j}$ is the global histogram and I is the LBP histogram of one block.

As described above, we designate patterns that have uniformity value of at most 2 as ’uniform’ and uses the following Equation:

$$LBP_{P,R}^{U2} = \begin{cases} \sum_{P=0}^{P-1} s(g_p - g_c) & if \quad U(\mathrm{LBP}_{P,R}) \leq 2 \\ P+1 & otherwise \end{cases} \tag{3.2}$$

Where,

$$U(LBP_{P,R}) = |s(g_{p-1} - g_c) - s(g_0 - g_c)| + \sum_{P=1}^{P-1} |s(g_p - g_c) - s(g_{p-1} - g_c)| \tag{3.3}$$

The U value is the uniformity value for at most 2. From definition exactly P+1 ’uniform’ binary patterns can occur in a circularly symmetric neighbor set of P pixels [61]. The above equation assigns a unique label to each of them, corresponding to the number of ’1’ bits in the pattern (0 -> P), while the ’non uniform’ patterns are grouped under the miscellaneous label (P+1). The different ULBP advantages and disadvantages are: (a) Advantages: Only “uniform” patterns are fundamental patterns of local image texture. The uniform LBP gives better performance than LBP due to statistical properties of these patterns. Lower dimensionality of features. (b) Disadvantages: No rotation Invariant.

The following figure shows and explains the blatant difference between LBP and ULBP.

If U≤2 it is uniform else it is non-uniform LBP and Uniform LBP has P(P−1)+3 output values.

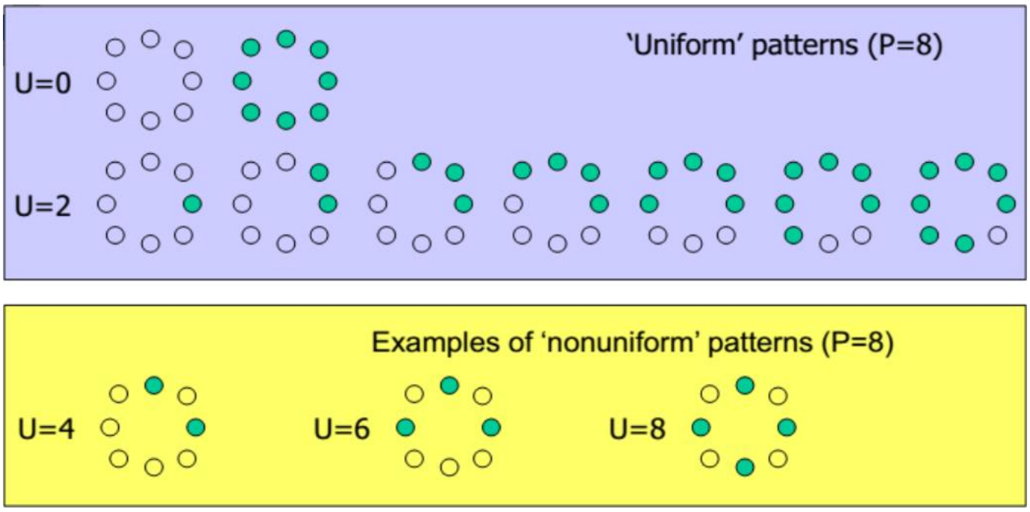


Figure 3.3. Difference between LBP and ULBP.

The current form of the LBP method is quite different from its basic version, a number of extensions have been developed during the past few years. An example of a recent extension is the assignment of weights to the regions. Therefore, there is still a lot of research going on to find more extensions and improve the robustness of the method.

3.2 Using K-Nearest Neighbor (KNN) to Classify a Face Image

K-nearest neighbor has been used in statistical estimation and pattern recognition since the beginning of 1970s as a non-parametric technique; nowadays, it is commonly used for object classification. K-NN is a type of lazy learning algorithm where the function is only approximated locally and all computation is deferred until classification. The K-NN classifier has been best suited for classifying persons based on their images, due to its lesser execution time and better accuracy than other commonly used methods such as hidden Markov model and kernel method. Some methods like support vector machine (SVM) and Adaboost algorithms have proved to be more accurate than K-NN classifier, but the K-NN classifier has a faster execution time and it is more dominant than SVM [60].

Choosing the optimal value for K firstly depends upon inspecting the specific dataset; so, the K value is estimated using the available training sample observations. In general, a large K value is more precise as it reduces the overall noise in the classification, but there is no guarantee because it makes boundaries between classes less distinct. Cross-validation is one way to retrospectively determine a good K value by using an independent dataset to validate the K value; a good K can also be selected by various heuristic techniques. Historically, the optimal K for most datasets has been chosen between 3 to 10; that produced much better results than 1-NN. In K-NN classification, the output is a class membership. An object is classified by a majority vote of its neighbors, with the object being assigned to the most common class among its k-nearest neighbors (K is a positive integer, typically small). The special case where the class is predicted to be the class of the closest training sample (K=1) is called the nearest neighbor algorithm.

The training sets are vectors in a multidimensional feature space, each with a class label. The training phase of the algorithm consists only of storing the feature vectors and class labels of the training samples. In the classification phase, K is a user-defined constant, and an unlabeled vector (a query or test face image) is classified by assigning the label which is most frequent among the K training samples nearest to that specific query face. That means an image in the test set is recognized by assigning to it the label of the closest face inside the training set, so that the distance is measured between them. A commonly used distance metric is the Euclidean distance, which is often chosen for determining the closeness between the data points in K-NN; a distance is assigned between all pixels in a dataset. The distance defined as the Euclidean distance between two pixels is given by:

$$D(\mathrm{a},\mathrm{b}) = \sqrt{\sum_{i=1}^{n} (\mathrm{a}_i - \mathrm{b}_i)^2} \tag{3.4}$$

KNN Mathematical Approach

***k*-Nearest Neighbors (*k*-NN)** is a simple machine learning algorithm that categorizes an input by using its *k* nearest neighbors [62]. KNN can be used for both classification and regression predictive problems. However, it is widely used in our thesis for classification problems. This paragraph answer a couple of questions like: How does the KNN algorithm work? Which method do we use for choosing K value? How do we choose the value K? And so forth.

Let's take a simple case to demonstrate a KNN analysis (algorithm) [63]. We consider here the task of classifying an unknown image (query image) among a plenty known images (gallery). The figure below depicts the instances with the red triangles (RT) and blue squares (BS) which represent the known image, and the green circle (GC) which is the query image. We intend to classify the green circle (GC). Let us begin with K = 3, we will now make a circle with GC as a center just as big as to enclose only three data points on the plane (refer to the solid line). In this case, the algorithm would return a red triangle (RT), since it constitutes a majority of the 3 neighbors. Likewise, with k = 5 (the dotted line), the algorithm would return a blue square; here, the choice became very obvious as all the votes from the closest neighbor went to BS. The choice of the parameter K is very crucial in this algorithm, the next paragraph will explain carefully how we can conclude that one considered value is the best choice of K.

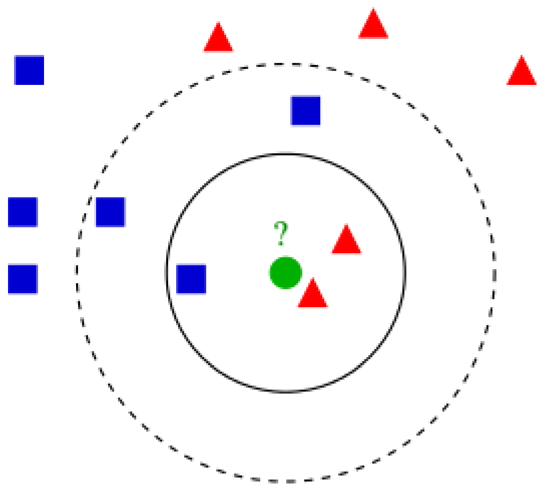


Figure 3.4. Example of KNN classification.

The best choice of k depends upon the data; generally, larger values of k reduce the effect of noise on the classification but make boundaries between classes less distinct. The previous example, given that all the training observation remain constant, with a given K value we can make boundaries of each class. These boundaries will segregate RT from BS (K = 3). The following figures present different boundaries separating two classes with different case of K values; therefore, we can see the effect of K value on the class boundaries.

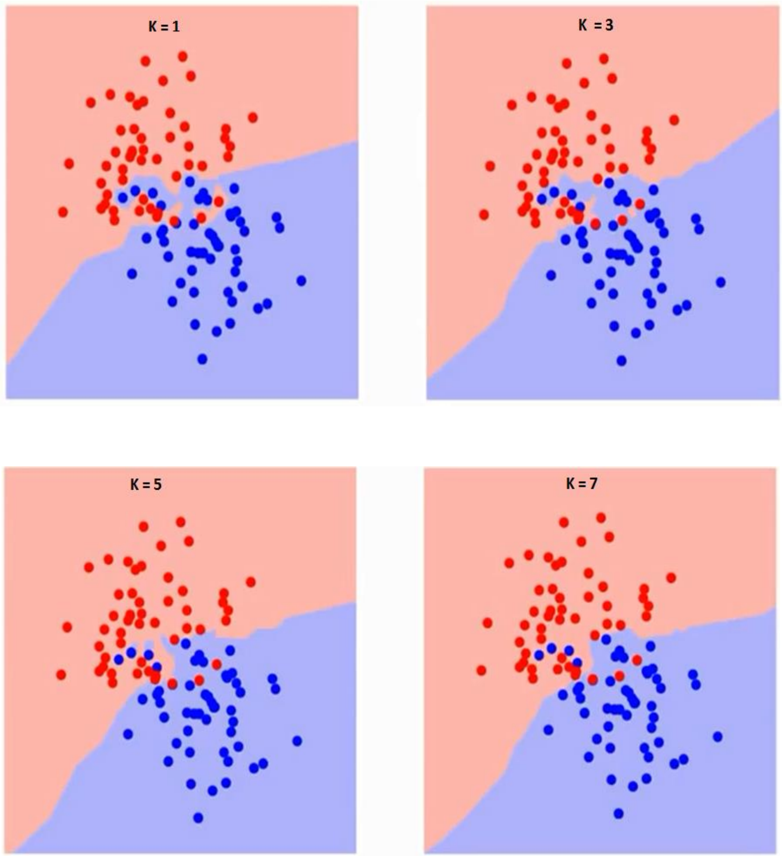


Figure 3.5. (a) K=1, (b) K=3, (c) K=5, (d) K=7.

You can remark that the boundary becomes smoother with increasing value of K. With K increasing to infinity it finally becomes all blue or all red depending on the total majority. The training error rate is one parameter we need to access on different K-value. The figure 3.6 is an example of the curve for the training error rate with varying value of K:

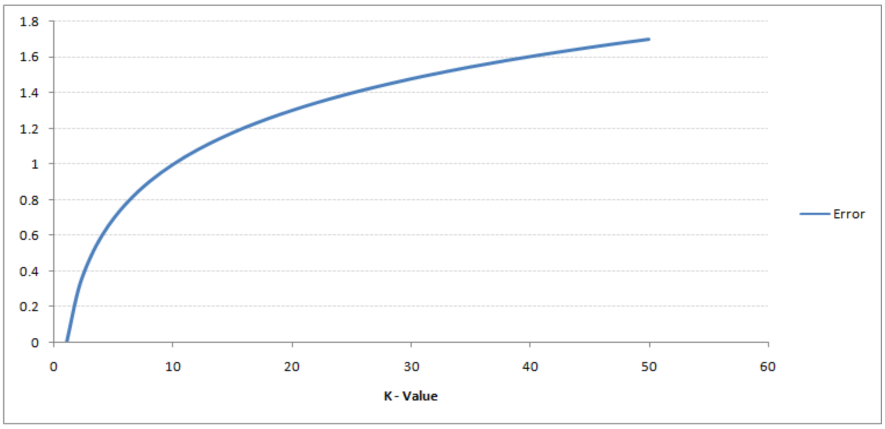


Figure 3.6. Training error rate curve.

We can see, the error rate at K=1 is always zero for the training sample. This is because the closest point to any training data point is itself. Hence, a good K can be selected by Cross-validation, by various heuristic techniques and so on. In this context, our good K value is selected by applying a K-fold cross-validation approach in order to estimate the optimum K.

Cross-validation [64] is a model assessment technique used to evaluate a machine learning algorithm's performance in making predictions on new datasets that it has not been trained on. It is a model evaluation method which is better than residuals. The problem with residual evaluations is that, they do not give an indication of how the learner will do when it is asked to make new predictions for data it has not already seen. Hence, Cross-validation can be a computationally intensive operation since training and validation is done several times. Because each partition set is independent, this analysis can be performed in parallel to speed up the process. This is done by partitioning a dataset and using a subset to train the algorithm and the remaining data for testing. Some of the data is removed before training begins, then when training is done, the data that was removed can be used to test the performance of the learned model on ``new" data. Because cross-validation does not use all of the data to build a model, it is a commonly used method to prevent overfitting during training. However, cross-validation can be distinguished on two common types, exhaustive and non-exhaustive cross-validation. Our interest here is to be focalize on K-fold Cross Validation which is one method of non-exhaustive type.

K-fold cross validation is one way to improve over the holdout method [65]. The data set is partitioned into k randomly chosen subsets (or folds), and the holdout method is repeated k times. Each time, one of the k subsets is used as the test set and the other k-1 subsets are put together to form a training set, so that the average error across all k trials is computed. A variant of this method is to randomly divide the data into a test and training set k different times. The advantage of doing this is that we can independently choose how large each test set is and how many trials our average over. The cross-validation process is then repeated k times (the folds), with each of the k subsamples used exactly once for validation data. The k results from the folds can then be averaged to produce a single estimation. The advantage of this method over repeated random subsets is that all observations are used for both training and validation, and each observation is used for validation exactly once. The disadvantage of this method is that the training algorithm has to be rerun from scratch k times, which means it takes k times as much computation to make an evaluation.

For example, setting k = 4 results in 4-fold cross-validation. In this case, we randomly shuffle the dataset into four sets e_0, e_1, e_2 and e_3, so that all of the sets are equal size (this is usually implemented by shuffling the data array and then splitting it in four). We then train on e_0 and test on e_1, e_2 and e_3, followed by training on e_1 and testing on e_0, e_2 and e_3; and so forth. Figure 3.7 shows properly the diagram of 4-fold cross-validation:

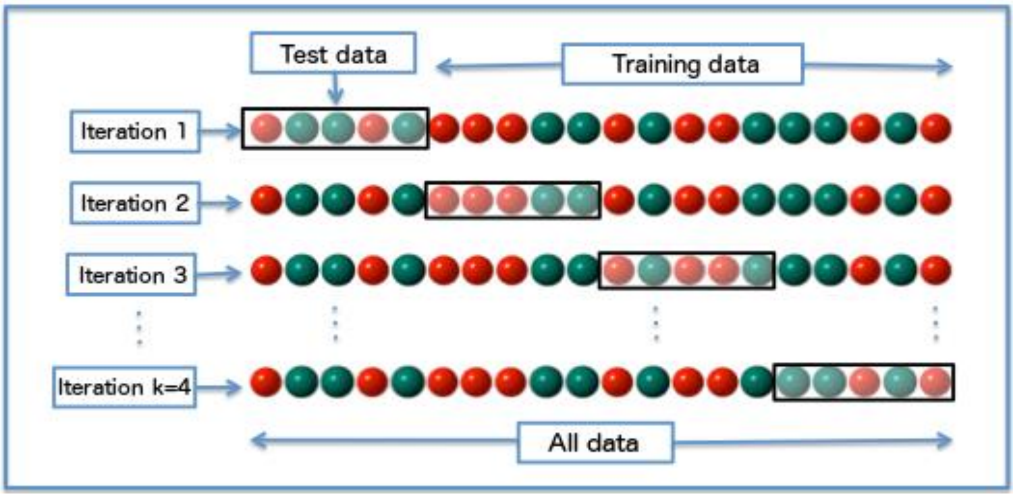


Figure 3.7. Diagram of K-fold Cross Validation with k = 4.

3.3 Gaussian Filter

The objective of this section is to understand the concept of image enhancement and to be able to apply commonly used image enhancement techniques to improve the visual interpretation of an image. Image enhancement deals with the procedures of making a raw image better interpretable for a particular application. The commonly used enhancement techniques are described here, which improve the visual impact of the raw remotely sensed data for the human eye. After enhancing the image or processed the image, the obtain result is more suitable for a particular application (sharpening or de-blurring an out of focus image, highlighting edges, improving image contrast, or brightening an image, removing noise). Image enhancement techniques can be classified in many ways. Contrast enhancement, also called global enhancement, transforms the raw data using the statistics computed over the whole data set. Examples are: linear contrast stretch, histogram equalized stretch and piece-wise contrast stretch. Contrary to this, spatial or local enhancement only take local conditions into consideration and these can vary considerably over an image. Examples are image smoothing and sharpening.

The sensitivity of the on-board sensors of satellites, has been designed in such a way that they record a wide range of brightness characteristics, under a wide range of illumination conditions. Few individual scenes show a brightness range that fully utilizes the brightness range of the detectors. The goal of contrast enhancement is to improve the visual interpretability of an image, by increasing the apparent distinction between the features in the scene. Although the human mind is excellent in distinguishing and interpreting spatial features in an image, the eye is rather poor at discriminating the subtle differences in reflectance that characterize such features. By using contrast enhancement techniques these slight differences are amplified to make them readily observable. Contrast stretch is also used to minimize the effect of haze. Scattered light that reaches the sensor directly from the atmosphere, without having interacted with objects at the earth surface, is called haze or path radiance. Haze results in overall higher digital number values and this additive effect results in a reduction of the contrast in an image. The haze effect is different for the spectral ranges recorded; highest in the blue, and lowest in the infrared range of the electromagnetic spectrum. The noise filtering is also a greater issue for solving the illumination conditions which seriously affect the images. As contrast enhancement, its goal is to remove the unnecessary information from the image while preserving the underlying structure and the original texture under diverse lightning conditions. In this work, we will based on noise filtering for resolving illumination conditions problem, and we will precisely focus on low-pass filter as a solution.

Applying a low pass filter has the effect of filtering out the high and medium frequencies and the result is an image which has a smooth appearance. Hence, this procedure is sometimes called image smoothing and the low pass filter is called a smoothing filter. It is easy to smooth an image, but the basic problem is to do this without losing interesting features; for this reason much emphasis in smoothing is on edge-preserving smoothing. There are many low-pass filters as linear smoothing filters, but the most important one is the Gaussian filter, which applies weights according to the Gaussian distribution. More aggressive than the mean filter, the Gaussian filter deals with random noise more effectively, but it still simply mixes the noise into the result and smooths indiscriminately across edges. Gaussian filters are ideal to start experimenting with filtering because their design can be controlled by manipulating just one variable- the variance σ. The key parameter σ controls the extent of the kernel and consequently the degree of smoothing (and how long the algorithm takes to execute). Gaussian filters are a class of low-pass filters, all based on the Gaussian probability distribution with this following mathematical function:

$$G(x,y)=e^{-\frac{(x^2+y^2)}{2\sigma^2}} \tag{3.5}$$

Where G(x,y) is the Gaussian low-pass filter of size x and y, with standard deviation σ (positive).

Gaussian filters have weights specified by the probability density function of a bivariate Gaussian, or Normal, distribution with variance σ^2, that is

$$w_{ij}=\frac{1}{2\pi\sigma^2}\exp\left\{\frac{-(\mathrm{i}^2+\mathrm{j}^2)}{2\sigma^2}\right\} \qquad for \quad i,j=-[3\sigma],\ldots,[3\sigma], \tag{3.6}$$

For some specified positive value for σ^2. Here 'exp' denotes the exponential function and [3σ] represents the 'integer part' of 3σ. Limits of ±3σ are chosen because Gaussian weights are negligibly small beyond them. Note, that the divisor of $2\pi\sigma^2$ ensures that the weights sum to unity (approximately), which is a common convention with smoothing filters. If σ^2 = 1, the array of weights is [78]:

$$w = \frac{1}{1000}\begin{pmatrix} 0 & 0 & 1 & 2 & 1 & 0 & 0 \\ 0 & 3 & 13 & 22 & 13 & 3 & 0 \\ 1 & 13 & 59 & 97 & 59 & 13 & 1 \\ 2 & 22 & 97 & 159 & 97 & 22 & 21 \\ 1 & 13 & 59 & 97 & 59 & 13 & 1 \\ 0 & 3 & 13 & 22 & 13 & 3 & 0 \\ 0 & 0 & 1 & 2 & 1 & 0 & 0 \end{pmatrix} \tag{3.7}$$

(For succinctness of notation, we show a 7 x 7 array to represent the weights w_{kl} for k, l = -*3, …., 3*, and we have specified the weights only to three decimal places -- hence the divisor of 1000 at the beginning of the array.)

The Gaussian Filter is used as a smoothing filter and it is applied by convolving an $n \times n$ image window with an $n \times n$ Gaussian kernel and obtaining a weighted sum [32]. A filter usually consists of a 3×3 array (sometimes called kernel) of coefficients or weighting factors. It is also possible to use a 5×5, a 7×7 or even a larger odd numbered array. The filter can be considered as a window that moves across an image and that looks at all the digital number values falling within the window. Each pixel value is multiplied by the corresponding coefficient in the filter. For a 3×3 filter, the 9 resulting values are summed and the resulting value replaces the original value of the central pixel; this operation is called convolution. Figure 3.8 illustrates the convolution of an image using a 3×3 kernel.

Figure 3.8. Apply Filter.

3.4 Proposed Face Recognition System

The proposed face identification system is based on the combination of the robust uniform local binary pattern and k-nearest neighbor. Face recognition is not a simple problem, since an unknown face image seen in the extraction phase is usually different from the face image seen in the classification phase. The main aim of this work is to solve the identification problem through face images which can vary easily under the influence of pose, illumination, and expression. The face image is divided into a grid of small non-overlapping regions, where the global LBP histogram of a particular face image is obtained by combining the histogram sequence of each non-overlapping region; explicitly, the global features are collected in single vector and therefore classified using the k-nearest neighbor algorithm. The Euclidean distance finds the minimum distance between histogram images. After comparing two individual histograms, if there is any similarity distance, it means they are related, and otherwise, not.

The Figure 3.9 below displays our process diagram.

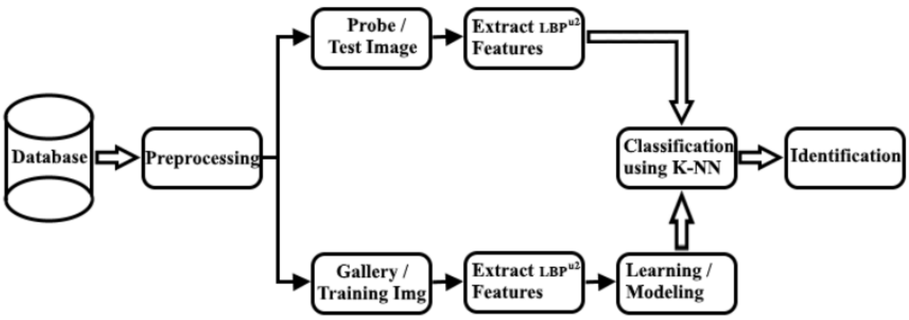


Figure 3.9. Diagram of the Process.

Our proposed system contains two principal stages before the junction:

Start:

- Face database
- Preprocessing

First stage:

- Input gallery images (training images)
- Collection of the extraction features using uniform LBP algorithm
- Learning or modeling via the LBP histogram

Second stage:

- Input the probe or query image (test images)
- Collection of the extraction features using uniform LBP algorithm

Junction:

- Classification using the K-NN algorithm with Euclidean distance

End:

- Identification process

Preprocessing phase

Consists of registering all images inside the database. The main aim of this phase is to improve the image data by suppressing unwanted distortions or enhancing important image features for further processing. Sometimes images have many lacks in contrast and brightness due to different limitations of imaging sub-systems and illumination conditions while capturing the image; techniques to resolve these issues include: contrast stretching, noise filtering, and histogram modification. Only noise filtering is applied in our work, after image registration. In definition, noise filtering is used to remove the unnecessary information from the image while preserving the underlying structure and the original texture under diverse lightning conditions. There are various types of filters available today, such as low-pass, high-pass, mean, median, etc.

Gaussian Filters Used as a Low Pass Filters

One of the major problems that face recognition has to deal with is variations in illumination. Many studies have been explored to reduce, normalize, and ameliorate the effect caused by illumination variations. A Gaussian filter used as a low-pass filter is an appropriate method for carrying out illumination reduction and remove the lighting changes; its main purpose is to suppress all noise in the image. Another important property of Gaussian filters is that they are non-negative everywhere; this is important because most 1D signals vary about $x = 0$, $(x \in \mathbb{R})$ and can have either positive or negative values. Images are different in the sense that all values of one image are non-negative $(x \in \mathbb{R}^+)$. Thus, convolution with a Gaussian filter guarantees a non-negative result, so the function maps non-negative values to other non-negative values $(f : \mathbb{R}^+ \to \mathbb{R}^+)$; the result is always another valid image. Digital images are composed of two frequency components: low (illumination) and high (noise). The Gaussian mathematical function implemented in this work is:

$$G(x, y) = e^{-\frac{(x^2+y^2)}{2\sigma^2}}$$

Where G(x,y) is the Gaussian low-pass filter of size x and y, with standard deviation σ (positive).

Feature Extraction Phase

The LBP algorithm is a method of damage reduction technology that represents a discrimination of an interesting part of the face image in a compact feature vector. When the pre-processing phase is achieved, the LBP algorithm is applied to the segments in order to obtain a specific feature histogram. A focus on the feature extraction phase is essential because it has an observable impact on recognition system efficiency. An important one is that feature extraction with LBP is a straightforward (real-time) process, for each image the feature vector can directly be constructed. The absence of the training aspect in LBP has a positive influence on the processing speed and the integration of the method in a new environment. Another important thing is that LBP is less sensitive to illumination variations in the images, because LBP eliminates offsets in pixel values; it is also less sensitive to rotation and scaling variations of the images. The selection of our feature extraction method is the single most important factor to achieve higher recognition performance; that is why we used uniform LBP to extract useful features as it generates a more precise recognition rate and makes the process simpler and more effective. The recent version of Matlab (from 2015a to now) has a build function appropriate for Uniform LBP named “extractLBPFeatures”; and this function is very helpful concerning the selection of the LBP parameters (a neighborhood P of size R surrounding the center pixel). Because the parameters P and R influence the performance of LBP operator, then the application of suitable neighbor-sets for different values of (P, R) needs to be done with utmost care.

Learning or Modeling Phase

Learning or modeling via LBP histogram is used to fit a model of the appearance of face images in the gallery, so that we can be able to know the discrimination between the faces of different subjects inside the database. In order to improve processing time, the extracted distance vectors are sorted in increasing order. In our framework, the learning step forms tightly packed conglomerates of visual feature histograms at detailed scales. These are determined by a form of configuration feature set, implying that the processing part reveals the similarity between features histograms. The characteristics of the processing part will be made explicit during matching in the classification phase.

K-NN Classifier

KNN algorithm is a method for classifying objects based on closest training examples in the feature space. K-NN is the simplest of all machine learning and classification algorithms, and stores all available cases and classifies new cases based on a similarity measure. Training process for this algorithm only consists of storing feature vectors and labels of the training images. In the classification process, the unlabeled query point is simply assigned to the label of its k nearest neighbors. KNN is a k-related algorithm; its classification accuracy is sensitive to the value of k. Thus, a natural thought is designing a learning algorithm to automatically search a best k value for KNN. To achieve this goal, the most direct method is try various k values and choose the best one. Based on this idea, the value K is used to perform classification by computing the simple histogram similarities. In this context, our good K value is selected by applying a K-fold cross-validation approach in order to estimate the optimum K. Further, each image of a set of visual features will find the best matching feature set between the test and all the training images. Typically the object is classified based on the labels of its k nearest neighbors by majority vote. If k=1, the object is simply classified as the class of the object nearest to it. When there are only two classes, k must be an odd integer. However, there can still be tie when k is an odd integer when performing multiclass classification. After we convert each image to a vector of fixed-length with real numbers, we used the most common distance function for KNN which is Euclidean distance D(x, y) to achieve the classification phase.

Algorithm 1: Proposed Algorithm

1. Initialize temp = 0
2. For each image I inside the database
3. Apply Gaussian low pass filter (G)
4. Divide the database into training and test sets
5. For each image inside the sets

   Extract LBP features

   End For
6. k-nearest neighbours (K value) are then found by analyzing the Euclidean distance matrix
7. Find similarity between LBP histograms
8. The most common class label is then assigned to the successfully recognized image

3.5 Test Results for CMU PIE and LFW

CMU PIE Database

Between October 2000 and December 2000 Terence Sim, Simon Baker, and Maan Bsat collected a database of over 40,000 facial images of 68 people. Using the Carnegie Mellon University 3D Room (controlled environment) they imaged each person across 13 different poses, under 43 different illumination conditions, and with 4 different expressions; they call this database the CMU Pose, Illumination, and Expression (PIE) database [73]. The purpose of PIE database is to evaluate the face recognition systems, it may also be used for facial feature detection, face pose estimation, and facial expression recognition. The following table 3.1 illustrates the different properties of CMU PIE database:

Table 3.1. Properties of CMU PIE database.

Properties	Descriptions
# of subjects	68
# of images/videos	41,368 images
Static/videos	static
Single/Multiple faces	Single
Gray/Color	Color
Resolution	640*486
Face pose	13 pose angles in vertical and horizontal

Facial expression	4 facial expressions: neutral, blinking (or eyes closing), smiling and talking
Illumination	43 different illumination conditions
Accessories	Glasses
Ground truth	Some feature point data Identification of subjects Measured locations of camera Head pose Facial expression labels Illumination positions
Additional materials	Background images

LFW Database

LFW (Labeled Faces in the Wild) [74] is a database of face photographs designed for studying the problem of unconstrained face recognition. LFW dataset contains 13,233 web-collected images from 5749 different identities, with large variations in poses, expressions and illuminations. In this work we used LFWcrop[75] which is a cropped version of the LFW dataset, keeping only the center portion of each image (i.e. the face). In the vast majority of images almost all of the background is omitted. LFWcrop was created due to concern about the misuse of the original LFW dataset, where face matching accuracy can be unrealistically boosted through the use of background parts of images (i.e. exploitation of possible correlations between faces and backgrounds).

Experiments

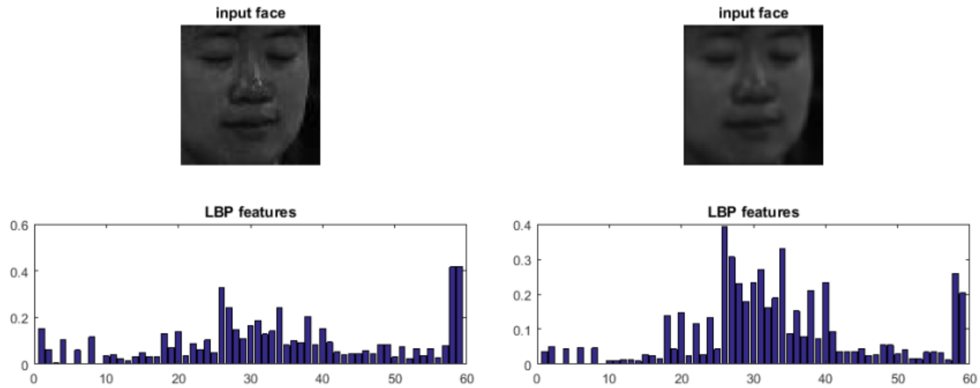
To verify the robustness and optimum of our method, experiments were carried out on two huge databases: CMU PIE and LFW. The performance of our proposed algorithm showed a powerful identification rate on the CMU PIE dataset.

Our face databases are very influential and common in studies of face recognition across pose, illumination, and expression variations. We used the image of the five nearly frontal poses (P05, P07, P09, P27, P29), a subset of 11,560 face images with 170 images per person on the CMU PIE dataset; and used around 6000 face images on the LFW dataset.

Firstly, we preprocessed our database due to the different illumination variations, and then applied the Gaussian filter before feature extractions in order to remove noises in the image to get a real LBP histogram of each image. Euclidian distance calculates the distance matrix between two images so that the image can be classified by a majority vote of its neighbors.

In our framework, we showed the performance of the Gaussian filter used as low-pass filter, which is an appropriate method for noise filtering. Here, the filter size used was different for each dataset.

For the CMU PIE dataset we used 3×3 as the size with σ = 2, and 5×5 as the size with σ = 1 for the LFW dataset. The higher filter size on LFW is due to the fact that it is an unconstrained or uncontrolled environment database and each image contains much more noise than images in a constrained or controlled environment (CMU PIE). Thus, the calculation and application of Gaussian parameters must be done with utmost care. After applying the filter, we obtain an enhanced image without noise; it is important to note that the Gaussian filter has the same role inside our databases. The filter size and σ are the same for all images inside a specific database. For instance, figure 3.10 illustrates: figure 3.10a is the image before applying the Gaussian filter (Original image), associated with its corresponding LBP^{u2} histogram; figure 3.10b is the image after applying the Gaussian filter (Filtered image), with the corresponding LBP^{u2} histogram. The Gaussian filter removes all of the undesirable artifact (noise). Thus, we obtained an unmistakable image compared to figure 3.10a. Moreover, the histograms in figure 3.10a and figure 3.10b are different; in figure 3.10b, applying the filter is beneficial to get higher and more precise features (real histogram image without noise) than figure 3.10a. Therefore, Gaussian's difference can increase the visibility of edges and other details present in a digital image.



(a) Original Image **(b)** Filtered Image

Figure 3.10. LBP histograms comparison.

Figure 3.11 reveals the identification results of four people across the pose, illumination, and expression variations and accessories (wearing glasses). As we can see, all the subjects were correctly matched. Particularly, the subject in Figure 3.11a has a correct matching even in a reverse image, with incomplete face appearance and lighting change. Whereas, subjects in Figure 3.11b–d, displayed correct matching with different facial expressions: blinking and wearing glasses, talking, and smiling with lighting change, respectively.

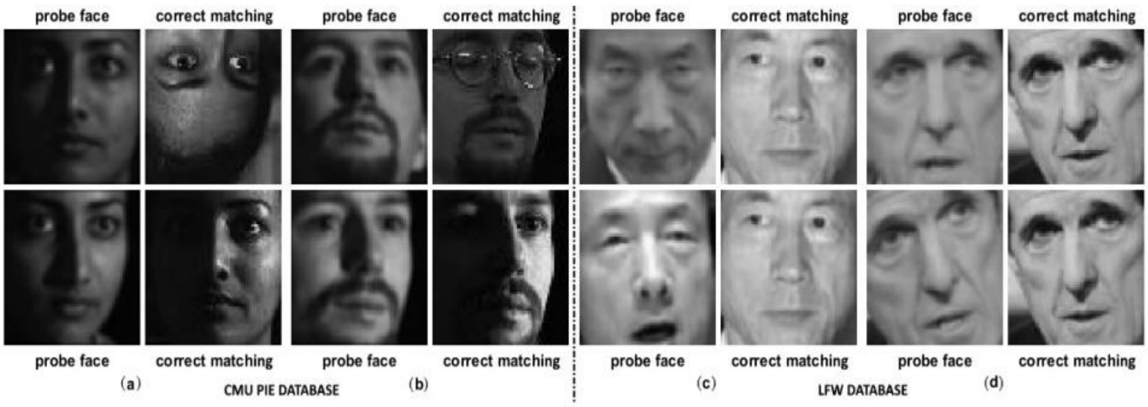


Figure 3.11. Correct identification. (a,b): CMU PIE; (c,d): LFW.

Finally, in Figure 3.12, the incorrect matching is less distinguishable—especially for the subject in Figure 3.12a,b, where the resemblance between probe image and gallery image (incorrectly matched) is extremely close. However, there are some cases where the failures are very blatant (Figure 3.12b,c), since the displayed images are chosen randomly inside the different sets.

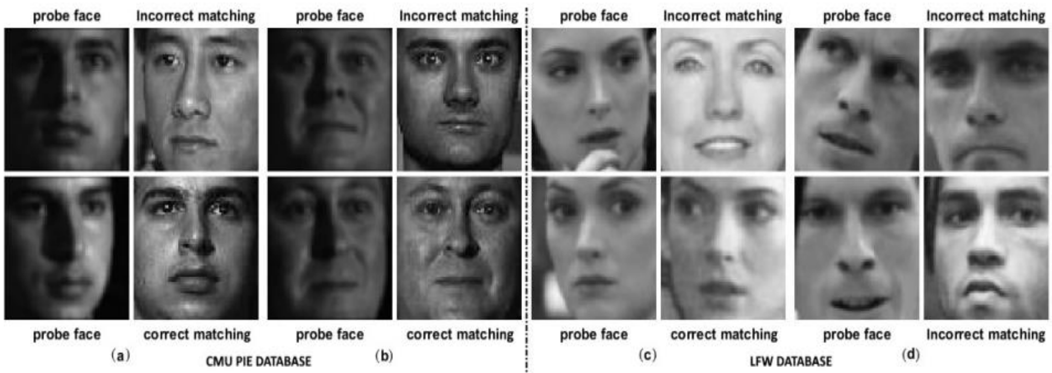


Figure 3.12. Incorrect vs. correct matching. (a,b): CMU PIE; (c,d): LFW.

For overall results, Table 3.2 describes the different outcomes obtained during the experiments. Maximum accuracy and powerful performance were achieved by implementing $LBP^{u2}_{22,4}$ K=4 on CMU PIE (99.26%) and $LBP^{u2}_{14,4}$ K = 4 on LFW (85.71%).

Table 3.2. Different Outcomes.

Databases	LBPu2P,R	K	Identification Rate (%)
CMU PIE	22,4	4	99.26
LFW	14,4	4	85.71

Table 3.3 and Table 3.4 describe the comparison of our results against many existing ones in both controlled and unconstrained environments, respectively.

Table 3.3. Controlled environment.

Method	Accuracy (%)

LBP (Chi Square Distance) [15]	82.33
LOG & LBP [14]	80
Lower-Order PZM [12]	97.75
Proposed Method	**99.26**

Table 3.4. Unconstrained environment.

Method	Accuracy (%)
DeepID2 [20]	99.15
Network Fusion + JB [21]	87.63
Model C [22]	99.28
Proposed Method	**85.71**

The novelty in this research effort is that the combination of LBP, K-NN algorithms, and Gaussian filter is applied to increase and enhance our face identification rate. Furthermore, our method proved that the performance of the proposed model can be validated using one controlled environment database (CMU PIE). In order to reinforce our experiments, we used one unconstrained database (i.e., LFW). The obtained result shows that our proposed algorithm compared to the innovative solutions produced approximatively the same results.

Chapter 4 Face Recognition using Convolutional Neural Network and Support Vector Machine

This chapter discusses the second approach (CNN and SVM), elaborates the proposed method and explains carefully the experiment results. Here the goal of CNN is to extract the features from images, in addition it has to perform "end-to-end learning" that means perform a classification task but it did not achieve this task as we had hoped. The execution time was very slow and the obtained results were around 20% less than the final ones. Then, in order to achieve and perform the classification phase we used the CNN features as input to SVM; the end process is detailed step by step below. Therefore, our experimental results have showed that the choice of SVM has a significant impact on performance.

4.1 Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN or ConvNet) is defined in machine learning, like a class of deep, feed-forward artificial neural networks that has successfully been applied to analyzing visual imagery. Convolutional neural networks were inspired by biological processes and are variations of multilayer perceptronsdesigned to use minimal amounts of preprocessing.

Convolutional neural networks (CNNs) expand on traditional neural networks by including both fully-connected hidden layers and locally-connected layers Known as convolutional layers [66] [67] [68]. In the traditional neural network, each hidden layer node is fully-connected to all nodes in the preceding layer. For large raw input images, full-connectivity becomes computationally intensive and does not take advantage of the local correlations in natural images. Inspired by the biology of human vision, convolutional neural networks take advantage of the spatial correlation in natural images and restrict the connectivity of each node to a local area known as its receptive field. In addition, convolutional neural networks use parameter sharing, pooling, and dropout to greatly reduce the number of parameters ("features") learned by the CNN. Convolutional networks is also common to see other types of layers such as pooling, activation, and normalization (Rectified Linear Units) layers. The following figure shows the architecture of a simple convolutional neural network consisting of two convolutional layers, two pooling layers, and three fully connected layers.

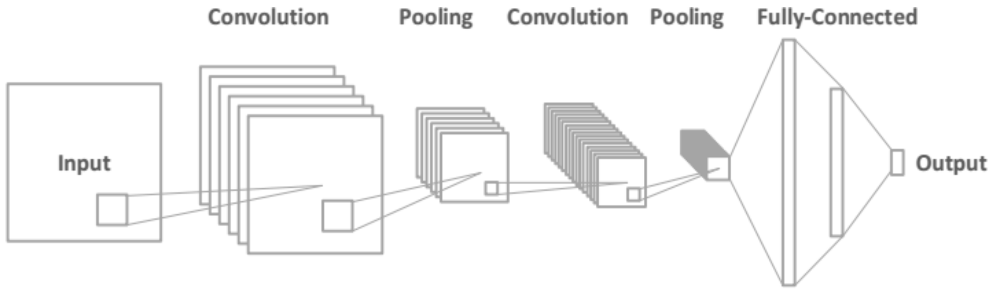


Figure 4.1: Convolutional neural network composed of convolution, pooling, and fully connected layers.

Recently, CNN has successfully recommended to resolve many Machine Learning application problems. One of the first applications of convolutional neural networks (CNN) is perhaps the LeNet-5 network proposed for resolving the optical character recognition. Compared to modern deep CNN, their network was relatively modest due to the limited computational resources of the time and the algorithmic challenges of training bigger networks. Though much potential laid in deeper CNN architectures (networks with more neuron layers), they have become recently prevalent, following the dramatic increase in both the computational power (due to Graphical Processing Units), the amount of training data readily available on the Internet, and the development of more effective methods for training such complex models. Other recent and notable example is the use of deep CNN for image classification on the challenging Image-net benchmark. Deep CNN has also additionally been successfully applied to applications including human pose estimation, face parsing, automated email spam detection, facial key point detection, stock market prediction, speech recognition, contextual online advertising, action classification and more.

In spite of the efficient architecture of convolutional neural networks, they still require a large number of images for training; so, using a small dataset for training leads to over-fitting. This requirement restricts the use of convolutional neural networks for classification purposes only to large datasets. However, recent research has shown that the output of the locally-connected convolutional layers produces highly discriminative descriptors. Consequently, a convolutional neural network pre-trained on a large image dataset can be used as a feature extractor for other closely related datasets. In the same vein, convolution layers play the role of feature extractor but they are not hand designed. Convolutional layers are able to extract the local features because they restrict the receptive fields of the hidden layers to be local. Convolution filter kernel weights are decided on as part of the training process. Here below is described the working process of Convolutional Neural Network:

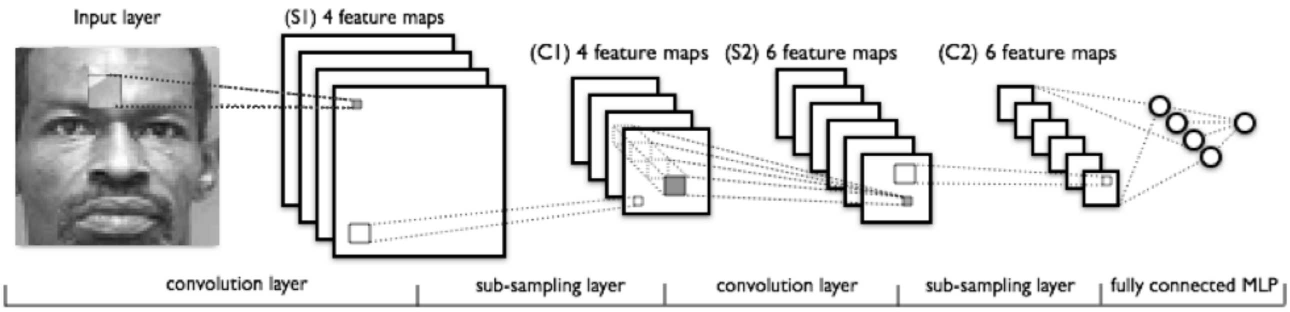


Figure 4.2: the working process of Convolutional Neural Network

From left to right in the above image, you can observe:

- The real input image that is scanned for features, with the light rectangle passes over it that represents the filter.
- The Activation maps are arranged in a stack on the top of one another, one for each filter that you use; the larger of the rectangle is 1 patch to be down-sampled.
- The activation maps are condensed via down-sampling.
- A new group of activation maps generated by passing the filters over the stack that is the first down-sampling.
- The second down-sampling condenses the second group of activation maps.
- A fully connected layer designates the output with 1 label per node.

There are several advantages to using Convolutional Neural Networks for our project. First, this type of network is invariant to small movements. Second, it's the fact that the neural network extracts a set of facial characteristics (feature maps) for each class during the process of training, keeping their relative position in space. We can change the architecture of convolutional network, controlling the number of layers, their size, and the number of feature maps for each layer.

Several CNN architectures have been proposed in the literature survey and some have been shown to produce better outcomes than the most advanced state-of-the-art recognition methods. In our experiments, we focus on the implementation of the VGG-Face CNN which is a deep convolutional neural network based on the VGG-Net architecture (VGG-Very-Deep-16 CNN architecture) [68]. VGG-Face network is composed of a sequence of convolutional, rectified linear unit (ReLu), pool, and fully connected (FC) layers. The convolutional layers use filters of dimension three while the pool layers perform subsampling with a factor of two. Since training can require extensive computational resources and large amounts of training data, training deep convolutional neural networks remain difficult from scratch; if such resources are not available, one can use a VGG network's activations layers as feature extractors. While the VGG-Face CNN can only identify the subjects in its training dataset, it can however be used as a feature extractor for any arbitrary face image by running the image through the entire network, and then extracting the output of the first fully-connected layer. The extracted feature is a highly discriminative, compact, and interoperable encoding of the input image. Once the features are well extracted from the FC-1 layer of the VGG-Face CNN, they can be used for training and testing arbitrary face classifiers as will be carefully explained and shown in the proposed method section (each phase is well defined in that section); then, the found features can be used for both face identification and face verification (in this case we used it for face identification). Table 4.1 provides the details on the VGG-Face CNN layers. The volume column represents the width, height, and depth or each layer, respectively. The parameters column shows the number of parameters learned in each layer.

Table 4.1. VGG-Face CNN layers.

Layer (1-11)	Volume	Parameters	Layer (12-22)	Volume	Parameters
INPUT	224×224×3	0	CONV3-512	28×28×512	(3×3×256)×512 = 1,179,648
CONV3-64	224×224×64	(3×3×3)×64 = 1,728	CONV3-512	28×28×512	(3×3×512)×512 = 2,359,296
CONV3-64	224×224x64	(3×3×64) ×64 = 36,864	CONV3-512	28×28×512	(3×3×512)×512 = 2,359,296
POOL2	112×112×64	0	POOL2	14×14×512	0
CONV3-128	112×112×128	(3×3×64)×128 = 73,728	CONV3-512	14×14×12	(3×3×512)×512 = 2,359,296
CONV3-128	112×112×128	(3×3×128)×128 = 147,456	CONV3-512	14×14×512	(3×3×512)×512 = 2,359,296
POOL2	56×56×128	0	CONV3-512	14×14×512	(3×3×512)×512 = 2,359,296
CONV3-256	6×56×256	(3×3×128)×256 = 294,912	POOL2	7×7×512	0
CONV3-256	6×56×256	(3×3×256)×256 = 589,824	FC-1	1×1×4096	7×7×512×4096 = 102,760,448
CONV3-256	6×56×256	(3×3×256)×256 = 589,824	FC-2	1×1×4096	4096×4096 =16,777,216
POOL2	28×28×256	0	FC-3	1×1×2622	4096×2622 =10,739,712

4.2 Support Vector Machine (SVM)

Support vector machine (SVM) is a non-linear classifier which is often reported as producing superior classification results compared to other methods. The support vector machine is also a training algorithm for learning the categorization and regulation of the regression from the related data. This algorithm has been suggested by a Russian researcher named Vapnik Vladimir in 1965s as the most famous trainee's categorization , and has been also recovered by Vapnik and Corinna Cortes in 1995 [69, 70] for the nonlinear mood coming from the Statistical Learning Theory being organized and arranged on the operational risk minimization process. This method is one of the most fairly newest approaches that has been innovated in the recent years in comparison to the traditional methods such as Perceptron Neural Nets. The support vector machines take in to consideration the operational risk as the aim variable and calculate the optimized value. Recently, SVM becomes popular because of its success in handwritten digit recognition, it is now regarded as an important example of "kernel methods", one of the key area in machine learning. The idea behind the method is to non-linearly map the input data to some high dimensional space, where the data can be linearly separated, thus providing great classification (or regression) performance. One of the bottlenecks of the SVM is the large number of support vectors used from the training set to perform classification (regression) tasks. The main purpose of the support vector machine is to obtain the function $F(x)$ as a determinant of the hyperplane, that means find the optimal separating hyperplane which maximizes the margin of the training data. There have been many various hyperplane that are able to separate the data, but the main question is: what is the optimal separating hyperplane do we have to choose or select? The training concept of the pictures being categorized into the higher dimensions in a one space is not unique. The main distinction of the related algorithm is subjected to the selection this hyperplane.

Support Vector Machine (SVM) is a supervised machine learning algorithm which can be used for both classification and regression challenges; however, it is mostly used in classification problems. Following here is the representation of different cases of the Support Vector Classifier (SVC):

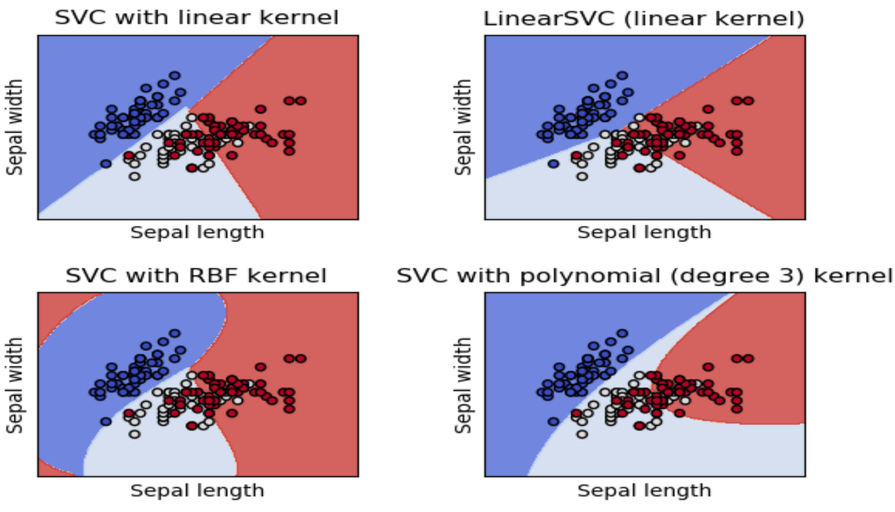


Figure 4.3. Support Vector Classifier (SVC).

Support vector machines (SVMs) are formulated to solve a classical two class pattern recognition problem. For a two-class classification problem, the goal is to separate the two classes by a function. Consider N points that belong to two different classes,

$$\{(\mathrm{x}_i, \mathrm{y}_i)\}_{i=1}^{N} \qquad and \qquad y_i = \{+1, -1\}, \tag{4.1}$$

Where x_i is an n-dimension vector and y_i is the label of the class that the vector belongs to. SVM separates the two classes of points by a hyperplane,

$$w^T x + b = 0 \tag{4.2}$$

Where x is an input vector, w is an adaptive weight vector, and b is a bias. The functional margin of the hyperplane is represented as,

$$\begin{cases} \overline{\quad}T\overline{\quad} \\ w_0 x_i + b_0 \geq +1 \qquad y_i = +1 \\ \overline{\quad}T\overline{\quad} \\ w_0 x_i + b_0 \leq -1 \qquad y_i = -1 \end{cases} \tag{4.3}$$

For a given w_0 and b_0, the geometrical distance of a point x from the optimal hyper-plane is,

$$d(w_0, b_0, x) = \frac{|w_0 x + b_0|}{\|w_0\|} \tag{4.4}$$

The goal of the SVM is to find the parameters w_0 and b_0 for the optimal separating hyperplane to maximize the geometrical margin, , i.e. the distance between the hyperplane and the closest point of both classes. Hence the hyperplane that optimally separates the data is the one that minimizes $\frac{2}{\|w_0\|}$

subject to the constraints $y_i(\mathrm{w} . \mathrm{x}_i + \mathrm{b}) \geq 1, \forall_i$. The solution to this optimization problem is found through the maximization of the dual Lagrangian,

$$\Phi(\mathrm{w}) = \frac{1}{2}\|w\|_2^2 = \frac{1}{2}(\mathrm{w} . \mathrm{w}), \tag{4.5}$$

$$Q(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j (\mathrm{x}_i . \mathrm{x}_j), \tag{4.6}$$

$$\alpha_i \geq 0, \qquad \sum_{i=0}^{N} \alpha_i y_i = 0, \tag{4.7}$$

With respect to Lagrange multiplier α_i , subject to the constraints,

In the solution, only a small number of α_i is none zero, each of which corresponds to one training data point. These data points are called support vectors since they lie on the margin border. These support vectors are therefore the only data points that appear in the resulting hyperplane, i.e.

decision function, SVM decision function is presented by the following formula:

$$f_l(\mathrm{x}) = \sum_{i=1}^{m} y_i \alpha_i \langle x, x_{si} \rangle + b, \tag{4.8}$$

Where each x_{si} represents a support vector and m is the number of support vectors. Each test vector x is then classified by the sign of f(x).

The solution can be extended to the case of nonlinear separating hyperplanes by a mapping of the input space into a high dimensional space, x → Φ(x). The key property of this mapping is that the function Φ is subject to the condition that the dot product of the two functions $\Phi(x_i) \cdot \Phi(x_j)$ can be rewritten as a kernel function $K(x_i, x_j)$. The decision function above then becomes,

$$f(\mathrm{x}) = \mathrm{sgn}(\sum_{i+1}^{m} \alpha_i y_i . k(\mathrm{x}, \mathrm{x}_i) + \mathrm{b}) \tag{4.9}$$

Where x is the input vector, α and y are the weights of the support vectors, having y as positive or negative class mark (+1 or -1), b is the bias and K (. , .) the kernel function.

These induce sparseness in the solution and give rise to efficient approaches to optimization. Once a decision function is obtained, classification of an unseen example x amounts to checking on what side of the hyperplane the example lies.

The SVM approach is highly modular, allowing domain specific selection of the kernel function used. Some advantages and disadvantages of SVM are detailed as follow:

ADVANTAGES: There are many folds advantages of using the supervised learning approach of Support Vector Machine (SVM). They are very effective when we have very high dimensional spaces; they are still very effective when number of dimensions becomes greater than the existing number of samples. SVM uses a subset of training point also known as support vectors to classify different objects hence it is memory efficient. Support Vector Machines are versatile, for different decision function we can define different kernel as long as they provide correct result. Depending upon our requirement and application we can choose types of kernel which is most productive for our application.

DISADVANTAGES: The disadvantage of SVM is that if the number of features is much greater than the number of samples, the method is likely to give poor performances (avoid over-fitting in choosing Kernel functions and regularization term is crucial). SVM gives efficient result for small training samples as compared to large ones. SVMs do not directly provide probability estimates, so these must be calculated using indirect techniques. Also, we can have Non-traditional data like strings and trees as input to SVM instead of featured vectors. Lastly selecting appropriate kernel for the project is a big issue which depends upon user’s requirement.

4.3 Proposed Aging Face Recognition

Preprocessing Phase

The pre-processing phase here comprised converting the color input images into 8-bit grey-scale images, locating the eyes manually, normalizing (scaling and rotating) the images geometrically in such a way that the centers of the eyes were localized at predefined positions, cropping the face parts of the images and resizing the cropped area to a standard size and finally, normalizing the face images photo metrically by eliminating their mean and scaling their pixels to unit variance. Cropping an image means cutting out or trimming unneeded portions of an image; it is also the removal of the outer parts of an image to improve framing, accentuate subject matter or change aspect ratio.

Because we face on two unconstraint environment databases, first of all we detect facial image from the input image, identify facial landmarks, and align the face to remove the influence of the outer parts of an image in order to improve framing (crop the image). For detecting the face, a robust algorithm that can detect faces in real time named Viola-Jones algorithm [79] has been used. It works by selecting the prominent Haar-like features that differentiate between face and non-face using Adaboost algorithm; Cascading classifier is introduced to reject quickly non-face images and improve the detection rate. After detected the faces, then all the face images inside the MORPH and FGNET databases are properly and globally cropped to a standard size of 200 × 200 pixels according to the 5 facial landmarks (two eyes, nose and two mouth corners) by similarity transformation. Figure 4.4 shows the crop images of both databases:

Figure 4.4. Example of cropped images of two subjects in FGNET and MORPH databases respectively.

Feature Extraction

The one important step in face recognition system is the extraction of the feature matrix. A typical feature extraction algorithm tends to build a computational model through some linear or nonlinear transform of the data so that the extracted feature is as representative as possible.

Deep convolutional neural network-based feature extractors are typically distinguished according to whether the filters (i.e., the convolution kernels) employed are learned (i.e., determined from a training data set through optimization) or pre-specified (i.e., chosen a priori, possibly taking into account structural properties of the data set). We used the VGG-Face CNN provided by the MatConvNet toolbox [71] for feature extraction. By definition, MatConvNet is an implementation of Convolutional Neural Networks (CNNs) for MATLAB. The toolbox is designed with an emphasis on simplicity and flexibility. It exposes the building blocks of CNNs as easy-to-use MATLAB functions, providing routines for computing linear convolutions with filter banks, feature pooling, and many more. In this manner, MatConvNet allows fast prototyping of new CNN architectures; at the same time, it supports efficient computation on CPU and GPU allowing to train complex models on large datasets such as ImageNet ILSVRC (Large Scale Visual Recognition Challenge).

The VGG-Face network described in section above has a deep architecture composed of 3×3

convolution layers, 2 × 2 pooling layers, ReLu layers, and 3 fully-connected layers. The VGG network is originally trained in the purpose to perform classification rather than feature extraction; so that, 4096-dimensional descriptors are extracted from the activation of the first fully connected layer (FC-1). To extract features from an image, the image is firstly preprocessed and fed to the CNN as a multidimensional array of pixel intensities. Concerning the RGB images, the input is a 224 × 224 × 3 array, while for grayscale images the input is a 224 × 224 × 1 array. Each convolutional layer performs a filtering operation on the preceding layer resulting in an activation volume which in turn becomes the input of the following layer. Pooling is used here throughout the network to reduce the number of nodes by down sampling the activation maps using the max operation. The fully connected layers of the network were used for learning the classification function. The feature descriptors are extracted from the output of the first fully-connected layers (FC-1), thus L2-normalized by dividing each component by the L2-norm of the feature vector; then the normalized features are used for training and testing in the purpose to achieve the identification process.

SVM Classification

In pattern recognition and machine learning, classification is the problem of identifying which of a set of class a new observation belongs, on the basis of a training set of data containing observations (or instances) whose class is known. In this study the classification of the known faces from the unknown faces is implemented using SVM. The SVM is a non-probabilistic binary linear classifier which classifies the test face input in to either known or unknown face. After the images were preprocessed and extracted features, they would present in the large representation space. Thus, they would be projected into the Sub-space in order to analysis easily and reduce dimensions of image's feature. We demonstrate our SVM-based algorithm on identification application. In identification, the algorithm is presented with an image of an unknown person; the algorithm reports its best estimate of the identity of an unknown person from a database of known individuals. In a more general response, the algorithm will report a list of the most similar individuals in the database.

Our implementation uses SVM MATLAB Toolbox [72] as the underlying SVM classifier. We encapsulate its stateless functionality in an object-oriented manner to work in an incrementally trained interactive environment. This avoids having to supply the set of training examples in its entirety before any classification can proceed and allows the user to augment the training data. It also enables us to export the entire state of a trained SVM classifier for later use. Hence, data gathered across several training sessions is preserved and can be re-used for classification. In addition, it allows for convenient combination of training data from multiple subjects to accomplish independent image classification; then, the user requests for training examples to be gathered at discrete time intervals and provides an image for each subject. This is combined with the displacements output by the feature extraction phase and added as a new example to the training set; the SVM is then retrained. Unseen aging face to be classified pass the same feature extraction process and are subsequently assigned the image of the corresponding face aging that most closely matches their displacement pattern by the SVM classifier. Most computational overhead resides in the training phase. However, due to the fact that the training set is interactively created randomly by the user and hence limited in magnitude and that the image training examples are of constant and small size, overhead is low for typical training runs. This is also aided by the sparseness of the SVM solution, manifested by the fact that the number of support vectors which define the decision surface only increases sub-linearly as more examples are added to the training data. Because evaluation of an SVM decision function on unseen input essentially amounts to checking which of the two subspaces defined by a separating hyperplane a point lies in, classification overhead is negligible. This allows our approach to perform classification both directly upon user request and continuously in real time for every image inside the dataset, with the current result being constantly reported back to the user.

4.4 Test Results for MORPH Album 2 and FG-Net

MORPH Album 2

MORPH (Craniofacial Longitudinal Morphological Face Database) [76] is the largest publicly available longitudinal face database. It is actively being used in over 30 countries. The MORPH data corpus embraces thousands of facial images of individuals across time, collected in real-world conditions (it is an uncontrolled collection or dataset). MORPH has two separate datasets or albums: Album 1 and Album 2.

Album 1 contains digital scans of many photographs of different individuals taken between October 26th 1962 and April 7th 1998 which we refer to as acquisition dates. Those acquisition dates correspond to increasing ages for individuals in the database; and the range of the dates varies between 46 days to 29 years after the earliest photograph. A stats sheet for Album1 is available to the public and this album only contains 1690 face images from 625 different subjects. Album 2 contains longitudinal digital photographs collected also over several years. The album is still evolving and images are acquired quarterly. A subset of this album is available for academic researchers and contains about 78,000 face images of 20,000 different subjects captured under different ages (ranging from 16 to 77). Both albums include metadata for race, gender, date of birth, and date of acquisition. Eye coordinates for the sets are also available upon request. Comparing to the MORPH Album 1 dataset, the MORPH Album 2 dataset has two desired attributes: (i) very large number of subjects, and (ii) large number of face images captured under different ages. Thus, we use in this study the MORPH Album 2 dataset.

FGNET Database

The FG-NET database [77] is a publicly available database that has been widely used for evaluating face recognition across aging algorithms. The database has facial images collected at ages in the range from 0 to 69. The study used the FG-NET database in the experiments for validation since it is by far the largest face aging database that covers such a wide age range. Moreover, the FG-NET database provides annotated facial landmarks, as well as age information for each image. In this database, there are 1002 images of 82 subjects. Table 4.2 illustrates the age range distribution of the face images in the FG-NET face aging database. The table shows the percentages occupied by each age span from the total number of images in the database. About 65% of the images were from children (with ages < 18) and around 34% of the images were from the adults (with ages = 18). For each facial image, there were 68 hand labelled landmarks representing the facial shape. The distribution of the number of images and subjects in different age ranges is then shown in Table 4.3.

Table 4.2. Age spans distribution in the FG-NET database.

Age Range	FG-NET
0 - 9	37.03
10 – 19	33.83
20 – 29	14.37
30 – 39	7.88
40 – 49	4.59
50 – 59	1.5
60 – 69	0.8
70 - 77	0

Table 4.3. Distribution of the number of images and subjects in different age ranges.

Age Range	0 - 5	6 - 10	11 - 15	16 - 20	21- 30	31 - 40	41 - 50	51 - 60	61 - 70
No of images	233	178	164	155	143	69	39	14	7
No of subjects	75	70	71	68	84	35	22	8	4

In addition, the images of this dataset have different qualities since some images are taken contemporarily in color while some others are captured a long time ago in gray. Although the samples included in the FG-NET dataset are limited, this dataset has a very wide age range and has become a general benchmark for comparison among different approaches and is still widely used to evaluate the performance of different age invariant algorithms.

Experiments

To compare our proposed age-invariant face recognition method with other standard methods, we conducted our experiment using the MORPH Album 2 and the FGNET Aging Databases, which are in the most widely used datasets having the largest age range. They contain about 78,000 face images of 20,000 different subjects captured under different ages (ranging from 16 to 77) and 1,002 face images of 82 people with ages ranging from 0 to 69 respectively. However, all images inside the different databases are cropped to contain the face region only, with a size of 200×200, and are aligned based on both eyes.

After image preprocessing as described in section above, we extract 4096-dimensional feature vectors from each image inside the both databases separately and respectively, then train the proposed suite of classifiers. We divide randomly the images as follow, 80% of the images of each subject are used for training phase and 20% are used for testing phase.

We divide each subject’s image into a multitude of different patch variations and generated the CNN feature vectors for each patch; and we extract the deep features from the output of the FC1 layer. For all experiments, the final representation of a testing face is obtained by concatenating its original face features and its horizontally flipped features. We process every combination of patches and then classify them using the SVM classifier. While on each image is extracted 4096-dimensional feature vectors use for experiments. For each feature vector, we test each subject in one-to-many (1:N) setup; in (1:N) setup, one image of a subject considering as probe set is compared to multiple images of that subject in the gallery set. Therefore, images are partitioned as we described above (80% for gallery and 20% for probe images). So, the score is computed by the SVM classifier of two features; the scheme allocates a score for each image feature depending on the closeness of the match to the desired image.

The performance reported is the average performance found for each database, table 4.4 and table 4.5 represent the FGNET results and the MORPH results respectively. Figures 4.5 and 4.6 describe the comparative diagram between existing methods and our own. The best performances are 88.2% (for FGNET database) and 94.3% (for MORPH database); they were achieved using an assembling multiple CNNs, each of which features are extracted from a well-aligned human face. These features are then concatenated to train SVM for attribute recognition in order to achieve maximum performance. However, it is straightforward to adapt this method to face attributes, since face parts can be well-aligned by landmark points; it has the additional benefit of requiring significantly less time to train with good performance. In conclusion, the performance on MORPH is better than FG-NET due to lower intra-class variability of the dataset.

Table 4.4. FGNET Results.

Method	Accuracy
Discriminative Model (Li et al.) [33]	47.5%
Park et al [28]	37.4%
HFA (Gong et al.) [34]	69.0%
MEFA (Gong et al.) [45]	76.2%
Ours	**88.2%**

Table 4.5. MORPH Results.

Method	Accuracy
Generative Model (Park et al.) [28]	79.8%
Discriminative Model (Li et al.) [33]	83.9%
HFA (Gong et al.) [34]	91.1%
CARC (B.C. Chen et al.)[43]	92.8%
Deep CNN (Li et al.)[44]	93.6%
MEFA (Gong et al.) [45]	94.5%
Ours	**94.3%**

You can remark in table 4.5 that, the outcome of our method is a little lower than Gong et al. [45] work which used the combined features; because the difference between the different outcomes is very close, then we can consider that the obtained results of both works are approximatively the same.

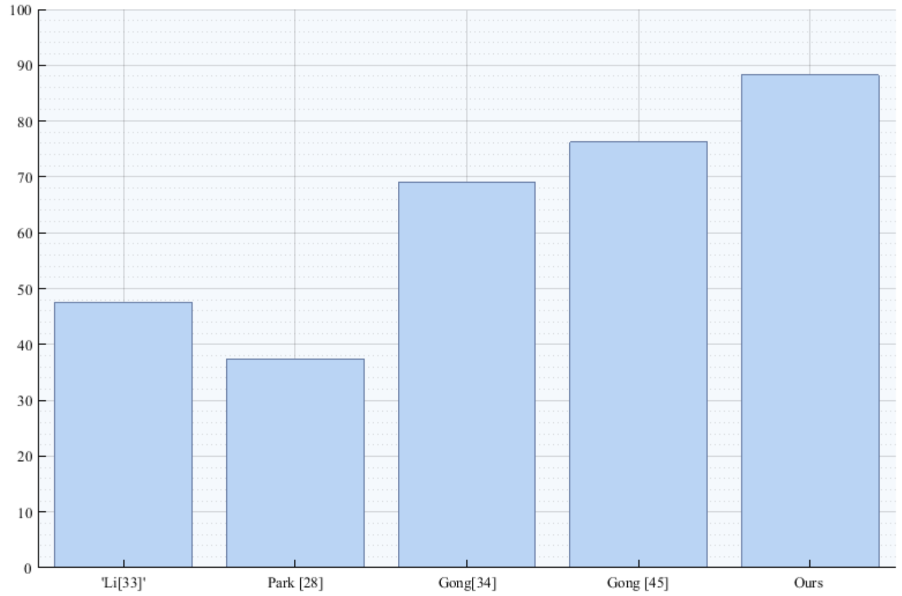


Figure 4.5. FGNET comparative diagram results.

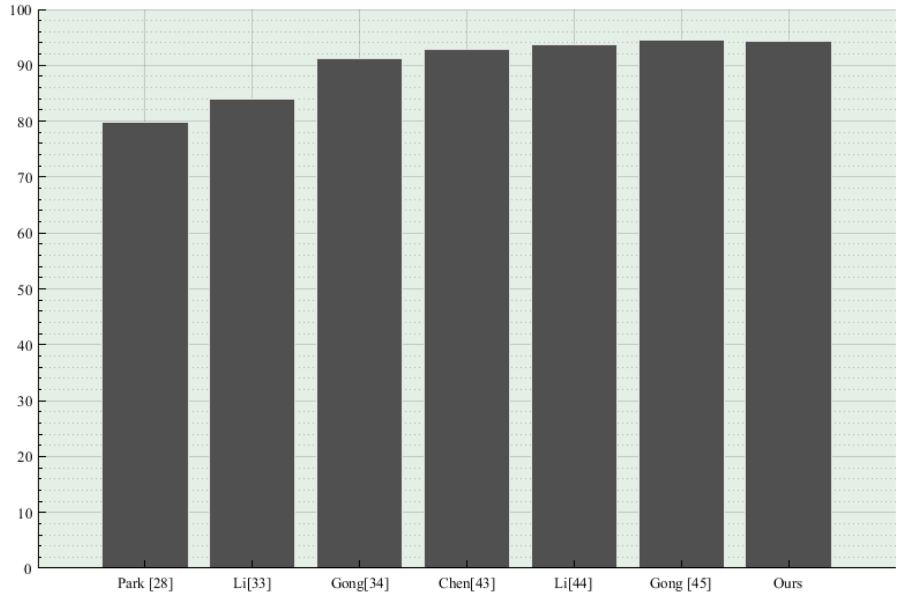


Figure 4.6. MORPH comparative diagram results.

Chapter 5 Conclusion

The face plays a major role in our social intercourse in conveying identity, and the human ability to recognize faces is remarkable. Goldstein (1983) (as cited in Chung & Thomson, 1995) stated that: "The face is the most important visual stimulus in our lives probably from the first few hours after birth, definitely after the first few weeks". The most difficult problem for today's face recognition systems is to deal with face variation factors. This study proposes a novel approach for pose, illumination and expression variations in face recognition by predicting facial features at different image changes. LBP and KNN achieve powerfully the feature extraction phase and the classification phase respectively on two different datasets (controlled environment database: CMU and uncontrolled environment database: LFW). This study also advances a novel approach for age-invariant face recognition using automatic, highly discriminative and interoperable deep learning driven CNN descriptors across two uncontrolled environment biometric datasets. It illustrates the feasibility and utility of a pre-trained CNN for automatic feature extraction yielding performance on the challenging FG-NET and MORPH datasets on the one hand with SVM classification image on the other hand.

Concerning the PIE (pose, illumination and expression variations), in this study, the face image is first divided into several blocks, from which features are extracted using local binary patterns (LBP), then the global feature histogram of each face is constructed. Identification is performed using k-nearest neighbor (K-NN) classifier in the computer feature space Euclidean distance (D) as similarity measure. Before extracting features, we applied a Gaussian filter to the images in order to remove noise and normalize illumination variations; this made LBP extraction easier to correctly match the probe image with other images inside the database. The experiments showed that $LBP^{u2}_{22,4}$ with K=4 achieved the maximum accuracy (99.26% on CMU PIE database). The simulation results indicate that the LBP features and K-NN classifier form a strong base for facial identification on unconstrained environment databases (85.71% on LFW dataset). Therefore, the unconstrained environment outcomes are opened for further analysis and may be improved upon.

This paper also addresses the challenge of face recognition subject to aging by using an approach based on deep learning and Support Vector Machine classifier; therefore, extensive experiments have been conducted on two public domain face aging datasets (MORPH Album 2 and FGNET) to confirm the effectiveness of our approach. We leverage a pre-trained Convolutional Neural Network to extract compact, highly discriminative and interoperable feature descriptors. The SVM classifier was found to be most robust to face variability due to aging and has a significant impact on top performance for classifying the images. We evaluated the performance of one-to-many matching (identification) and the experimental results showed that the set work is better than the singleton one for aging invariant. The best performance reported for our approach is reflected in generalization due to transfer learning and local processing due to the combined use of CNN and SVM classifier. The obtain results also show that it is easier to recognize older subjects from younger ones rather than younger subjects from older ones.

Acknowledgement

This dissertation would not have been possible without the guidance and the help of several individuals who in one way or another contributed and extended their valuable assistance in the preparation and completion of this study. After an intensive period of two years and half, today is the day: writing this note of thanks is the finishing touch on my dissertation. It has been a period of intense learning for me, not only in the scientific arena, but also on a personal level. Writing this dissertation has had a big impact on me. I would like to reflect on the people who have supported and helped me so much throughout this period. I sincerely wish to express my appreciation for the valuable help, which I received during the completion of the thesis by the following:

Foremost I offer my sincerest gratitude to my supervisor, Dr Guo Chun Sheng, who has supported me throughout my Master study and research with his patience, enthusiasm and knowledge whilst allowing me the room to work in my own way. I could not have imagined having such kind of advisor and mentor for my Master study, but thanks GOD I had him and I learnt too much from him.

A very special thanks goes to my best friend, mi amor, my beloved husband Raphael Ngnie who has been my inspiration as I hurdle all the obstacles in the completion of this research work and during my curriculum. Thanks for drafting my figures, proof reading this dissertation, and so forth; for his immense love, his motivation, his enthusiasm, his dedication, and his encouragement. Even when I were peaky and did not want to wake up in the morning, he allowed me to not give up.

In addition, I would like to thank my tutors, Mr. & Mrs. Mbombo and Professor Dr. Aloy, for their valuable guidance. You definitely provided me with the tools that I needed to choose the right direction and successfully complete my Master study.

My sincere thanks also goes to Dr. Wei (Anne) and Mrs. Wu (Joy), for offering me the full Scholarship opportunities in their university especially in the account of Department of International Students. Thanks Joy for believing in me without knowing my real skills. Indeed I did not deceive them because I did my best to be among the best student of my promotion.

I would also like to thank the entire Zhang family, Pascal Djembissi, Cyril Tchouthoua whose direct and indirect support helped me to complete my project in time. Thanks for their assistance and being such a good company.

Last but not the least, I would also like to thank my parents, my entire family and family in-law for their wise counsel and sympathetic ear, you are supporting me spiritually throughout my life and you are always there for me. Finally, there are my friends. We were not only able to support each other by deliberating over our problems and findings, but also happily by talking about things other than just our papers.

The authors gratefully acknowledge support for this work from the National Natural Science Foundation of China (No. 61372157). Guo Chun Sheng served as Principal Investigator for the grant at Hangzhou Dianzi University, Department of Information and Communication Engineering.

Finally I thank my God, my good Father, for letting me through all the difficulties. I have experienced Your guidance day by day JESUS. You are the one who let me finish my degree. I will keep on trusting You for my future and forever and ever. Thank you, Lord.

Thank you very much, everyone!

References

[1] Ahonen, T.; Hadid, A.; Pietikainen, M. Face Recognition with Local Binary Patterns; Springer: Berlin/Heidelberg, Germany, 2004.

[2] Mau, S.; Dadgostar, F.; Lovell, B.C. Gaussian Probabilistic Confidence Score for Biometric Applications. In Proceedings of the 2012 International Conference Digital Image Computing Techniques and Applications (DICTA), Fremantle, WA, Australia, 3–5 December 2012.

[3] Jones, M.J. Face Recognition: Where We Are And Where To Go From Here; IEEJ. Trans. on Elect., Information and Systems Vol. 129 (2009) No. 5 P 770-777.

[4] Abate, A.F.; Nappi, M.; Ricciardi, S.; Sabatino, G. One to Many 3D Face Recognition Enhanced Through k-d-Tree Based Spatial Access; Candan, K.S., Celentano, A., Eds.; Springer-Verlag Berlin Heidelberg 2005; pp. 5–16.

[5] R. Chellappa, C. Wilson, and S. Sirohey. Human and machine recognition of faces: a survey. Proceedings of the IEEE, 83-5:704–740, 1995.

[6] W. Zhao, R. Chellappa, A. Rosenfelda, and J. Phillips. Face recognition: a literature survey. Technical Report, CS-TR4167R, University of Maryland, 2000.

[7] C. Bunney. Survey: face recognition systems. Biometric Technology Today, pages 8–12, 1997.

[8] Phillips, P.J.; Moon, H.; Rizvi, S.A.; Rauss, P.J. The FERET Evaluation Methodology for Face-Recognition Algorithms. IEEE Trans. Pattern Anal. Mach. Intell. 2000, 22, 1090–1104. [Google Scholar] [CrossRef]

[9] Jones, M.J. Face Recognition: Where We Are and Where to Go from Here. IEE J. Trans. Elect. Inf. Syst. 2009, 129, 770–777. [Google Scholar] [CrossRef]

[10] Rahim, A.; Hossain, N.; Wahid, T.; Azam, S. Face Recognition using Local Binary Patterns (LBP). Glob. J. Comput. Sci. Technol. 2013, 13, 469–481. [Google Scholar]

[11] FaceIt-Hist. http://www.identix.com/company/comp history.html.

[12] Sultana, M.; Gavrilova, M.; Yanushkevich, S. Expression, Pose, and Illumination Invariant Face Recognition using Lower Order Pseudo Zernike Moments. In Proceedings of the 2014 Intertional Conference of Computer Vision Theory and Applications (VISAPP), Lisbon, Portugal, 5–8 January 2014.

[13] Mau, S.; Dadgostar, F.; Lovell, B.C. Gaussian Probabilistic Confidence Score for Biometric Applications. In Proceedings of the 2012 International Conference Digital Image Computing Techniques and Applications (DICTA), Fremantle, WA, Australia, 3–5 December 2012.

[14] Panchal, P.; Patel, P.; Thakkar, V.; Gupta, R. Pose, Illumination and Expression invariant Face Recognition using Laplacian of Gaussian and Local Binary Pattern. In Proceedings of the 5th Nirma University

International Conference on Engineering (NUiCONE), Ahmedabad, India, 26–28 November 2015.

[15] Khorsheed, J.A.; Yurtkan, K. Analysis of Local Binary Patterns for Face Recognition Under Varying Facial Expressions. In Proceedings of the Signal Processing and Communication Application Conference 2016, Zonguldak, Turkey, 16–19 May 2016.

[16] Ali, A.; Hussain, S.; Haroon, F.; Hussain, S.; Khan, M.F. Face Recognition with Local Binary Patterns. Bahria Univ. J. Inf. Commu. Technol. 2012, 5, 46–50.

[17] Ameur, B.; Masmoudi, S.; Derbel, A.G.; Hamida, A.B. Fusing Gabor and LBP Feature Sets for KNN and SRC-based Face Recognition. In Proceedings of the 2nd International Conference on Advanced Technologies for Signal and Image Processing—ATSIP, Monastir, Tunisia, 21–23 March 2016.

[18] Manvjeet Kaur, D. K-Nearest Neighbor Classification Approach for Face and Fingerprint at Feature Level Fusion. Int. J. Comput. Appl. 2012, 60, 13–17.

[19] Ebrahimpour, H.; Kouzani, A. Face Recognition Using Bagging KNN. In Proceedings of the International Conference on Signal Processing and Communication Systems (ICSPCS'2007), Gold Coast, Australia, 17–19 December 2007.

[20] Sun, Y.; Wang, X.; Tang, X. Deep Learning Face Representation by Joint Identification-Verification. In Advances in Neural Information Processing Systems; MIT Press: Cambridge, MA, USA, 2014; pp. 1988–1996.

[21] Hu, G.; Yang, Y.; Yi, D.; Kittler, J.; Christmas, W.; Li, S.Z.; Hospedales, T. When Face Recognition Meets with Deep Learning: An Evaluation of Convolutional Neural Networks for Face Recognition. In Proceedings of the IEEE International Conference on Computer Vision Workshops 2015, Santiago, Chile, 7–13 December 2015.

[22] Wen, Y.; Zhang, K.; Li, Z.; Qiao, Y. A Discriminative Feature Learning Approach for Deep Face Recognition. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 October 2016; Springer: Berlin/Heidelberg, Germany, 2016.

[23] G. Mu, G. Guo, Y. Fu, and T. S. Huang. Human Age Estimation using Bio-inspired Features. In *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2009, pp. 112–119.

[24] A. Montillo and H. Ling. Age Regression from Faces using Random Forests. In *Proc. IEEE Int. Conf. Image Process.*, Nov. 2009, pp. 2465–2468.

[25] K.-Y. Chang, C.-S. Chen, and Y.-P. Hung. Ordinal Hyperplanes Ranker with Cost Sensitivities for Age Estimation. In *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2011, pp. 585–592.

[26] J. Suo, X. Chen, S. Shan, and W. Gao. Learning Long Term Face Aging Patterns from Partially Dense Aging Databases. In *Proc. IEEE Int. Conf.Comput. Vis.*, Sep.–Oct. 2009, pp. 622–629.

[27] J. Suo, S.-C. Zhu, S. Shan, and X. Chen. A Compositional and Dynamic Model for Face Aging. *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 3, pp. 385–401, Mar. 2010.

[28] U. Park, Y. Tong, and A. K. Jain. Age-Invariant Face Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*,

vol. 32, no. 5, pp. 947–954, May 2010.

[29] A. Lanitis, C. J. Taylor, and T. F. Cootes. Toward Automatic Simulation of Aging Effects on Face Images. *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 4, pp. 442–455, Apr. 2002.

[30] X. Geng, Z.-H. Zhou, and K. Smith-Miles. Automatic Age Estimation based on Facial Aging Patterns. *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 12, pp. 2234–2240, Dec. 2007.

[31] T. Wu and R. Chellappa. Age Invariant Face Verification with Relative Craniofacial Growth Model. In *Proc. Eur. Conf. Comput. Vis.*, 2012, pp. 58–71.

[32] H. Ling, S. Soatto, N. Ramanathan, and D. W. Jacobs. Face Verification across Age Progression using Discriminative Methods. *IEEE Trans. Inf. Forensics Security*, vol. 5, no. 1, pp. 82–91, Mar. 2010.

[33] Z. Li, U. Park, and A. K. Jain. A Discriminative Model for Age Invariant Face Recognition. *IEEE Trans. Inf. Forensics Security*, vol. 6, no. 3, pt. 2, pp. 1028–1037, Sep. 2011.

[34] D. Gong, Z. Li, D. Lin, J. Liu, and X. Tang. Hidden Factor Analysis for Age Invariant Face Recognition. In *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 2872– 287.

[35] Tefas, A., C.K., Pitas, I. Using Support Vector Machines to Enhance the Performance of Elastic Graph Matching for Frontal Face Authentication. IEEE Trans. Pattern Anal. Mach. Intell., 2001, 23, pp. 735–746.

[36] Mark, L.S., Shaw, R.E., Pittenger, J.B. Natural Constraints, Scales of Analysis, and Information for the Perception of Growing Faces. In Alley, T.R., (Ed.); 'Social and applied aspects of perceiving faces', (Lawrence Embaum Associates, Mahwah, NJ, 1978).

[37] Zhao, W., Chellappa, R., Phillips, P.J., Rosenfield, A. Face Recognition: a Literature Survey. ACM Comput. Surv. (CSUR), 2003, 35, (4), pp. 399–458.

[38] Gayathri, M., Chandra, K. Age Invariant Face Recognition using Graph Matching. Proc. 4th IEEE Int. Conf. on Biometrics: Theory Applications and Systems (BTAS), IEEE, 2010, pp. 1–7.

[39] C. Geng and X. Jiang. Face Recognition using Sift Features. IEEE International Conference on Image Processing (ICIP), pp. 3313–3316, 2009.

[40] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, pp. 971–987, 2002.

[41] D. Chen, X. Cao, F. Wen, and J. Sun. Blessing of Dimensionality: High-Dimensional Feature and its Efficient Compression for Face Verification. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3025–3032, 2013.

[42] X. Wang and X. Tang. Random Sampling lda for Face Recognition. IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 259–265, 2004.

[43] B.C. Chen, C.S. Chen, and W.H. Hsu. Cross-Age Reference Coding for Age-Invariant Face Recognition and Retrieval. Computer Vision ECCV. Springer International Publishing, vol. 8694, pp. 768–783, 2014.

[44] Y. Li, G. Wang, L. Lin, and H. Chang. A Deep Joint Learning Approach for Age Invariant Face Verification.

In Computer Vision. Springer Berlin Heidelberg, pp. 296–305, 2015.

[45] D. Gong, Z. Li, D. Tao, J. Liu, and X. Li. A Maximum Entropy Feature Descriptor for Age Invariant Face Recognition. IEEE Conference on Computer Vision and Pattern Recognition, pp. 5289–5297, 2015.

[46] Syambas, N. R. and U. H. Purwanto (2012). Image Processing and Face Detection Analysis on Face Verification based on the Age Stages. Telecommunication Systems, Services, and Applications (TSSA), 2012 7th International Conference on.

[47] Lanitis, C. Taylor, and T. Cootes. Modeling the Process of Ageing in Face Images. In the Seventh IEEE Intl Conf. on Computer Vision (ICCV), Kerkyra, volume 1, pages 131136, 1999.

[48] H. Ling, S. Soatto, N. Ramanathanan, and D. Jacobs. A Study of Face Recognition as People Age. IEEE International Conference on Computer Vision, Rio De Janeiro, Brazil, October 2007.

[49] Animetrics. (2008) *Biometrics and Facial Recognition*. Retrieved from http://www.animetrics.com/technology/frapplications.html

[50] Goldstein, A. J., Harmon, L. D., & Lesk, A. B. Identification of Human Faces. Proceedings of the IEEE, 59(5), 748-760, 1971.

[51] Sirovich, L. & Kirby, M. A low-dimensional procedure for the characterization of human faces. Journal of Optical Society of America, 4(3), 519-524, 1987. Retrieved from http://www-flare.cs.ucl.ac.uk/staff/S.Prince/4C75/Sirovich87.pdf

[52] Turk, M. A. & Pentland, A. P. Face Recognition using Eigenfaces. Proceedings of the IEEE, 586-591 1991. Retrieved from http://www.cs.tau.ac.il/~shekler/Seminar2007a/PCA%20and%20Eigenfaces/eigenfaces_cvpr.pdf

[53] Phillips, P. J. et al. The FERET Evaluation Methodology for Face-Recognition Algorithms. IEEE Transactions on PAMI, 22(10), 1090-1104 2000. Retrieved from http://ieeexplore.ieee.org.ezproxy.umw.edu:2048/xpls/abs_all.jsp?arnumber=879790&tag=1

[54] National Science and Technology Council. (2006). *Face Recognition*. Retrieved from http://www.biometrics.gov/Documents/FaceRec.pdf

[55] A. Pentland, B. Moghaddam and T. Starner. View-based and Modular Eigenspaces for Face Recognition. IEEE Conf. on Comp. Vis. And Patt. Rec., pp. 84-91, 1994.

[56] W. Zhao. Robust Image Based 3D Face Recognition. PhD Thesis, University of Maryland, 1999.

[57] P. Ekman, W. V. Friesen. Facial Action Coding System. Consulting Psychologists Press, Palo.

[58] K. Taylor. Forensic Art and Illustration. CRC Press, Boca Raton, 2008.

[59] Ojala, T.; Pietikäinen, M.; Harwood, D. A comparative study of texture measures with classification based on feature distributions. Pattern Recognit. 1996, 29, 51–59.

[60] Manvjeet Kaur, D. K-Nearest Neighbor Classification Approach for Face and Fingerprint at Feature Level Fusion. Int. J. Comput. Appl.(0975–8887) Volume 60 (12), 13-17.

[61] T. Ojala, M. Pietikäinen, and T. Mäenpää. Multiresolution Gray-Scale and Rotation Invariant Texture

Classification with Local Binary Patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 7, pp. 971–987, Jul 2002.

[62] K-Nearest Neighbors. Retrieve from http://www.scholarpedia.org/article/K-nearest_neighbor

[63] K-Nearest Neighbors Algorithm. Retrieve from https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm

[64] Cross Validation (Statistics). Retrieve from https://en.wikipedia.org/wiki/Cross-validation_(statistics)

[65] Ronald Christensen. Thoughts on Prediction and Cross-Validation. Department of Mathematics and Statistics University of New Mexico May 21, 2015. Retrieve from http://www.math.unm.edu/~fletcher/Prediction.pdf

[66] Le Cun, Y., Bottou, L., Bengio, Y. and Haffner. Gradient Based Learning Applied to Document Recognition. Proceedings of the IEEE, 86, 2278-2324 P. (1998). https://doi.org/10.1109/5.726791

[67] Krizhevsky, A., Sutskever, I. and Hinton, G. Imagenet Classification with Deep Convolutional Neural Networks. In NIPS (2012).

[68] Parkhi, O.M., Vedaldi, A. and Zisserman, A. Deep Face Recognition. Proceedings of the British Machine Vision Conference (BMVC) (2015). https://doi.org/10.5244/c.29.41

[69] V. N. Vapnik. The Nature of Statistical Learning Theory. Springer-Verlag New York, Inc., New York, NY, USA, 1995.

[70] V. N. Vapnik. Statistical Learning Theory. Wiley, New York, NY, USA, 1998.

[71] Vedaldi, A. and Lenc, K. (2015) MatConvNet, “Convolutional Neural Networks for MATLAB. Proceedings of the 23rd ACM International Conference on Multimedia, Brisbane, 26-30 October 2015, 689-692. Retrieved from http://dx.doi.org/10.1145/2733373.2807412.

[72] SVM MATLAB Toolbox Retrieved from https://cn.mathworks.com/help/stats/support-vector-machine-classification.html

[73] Terence Sim, Simon Baker, and Maan Bsat. The CMU Pose, Illumination, and Expression (PIE) Database. Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition (FGR’02), 2002 IEEE.

[74] Labeled Faces in the Wild Homepage. Retrieved from http://vis-www.cs.umass.edu/lfw/

[75] LFWcrop Face Dataset. Retrieved from http://conradsanderson.id.au/lfwcrop/.

[76] Karl Ricanek Jr and Tamirat Tesafaye. MORPH: A Longitudinal Image Database of Normal Adult Age-Progression. IEEE 7th International Conference on Automatic.

[77] FG-NET Aging Database. Year: 2002. Retrieve from www-prima.inrialpes.fr/FGnet/

[78] Book: Chapter 3 FILTERS Retrieved from http://www.bioss.ac.uk/people/chris/ch3.pdf

[79] P. Viola and M. J. Jones. Robust real-time face detection. International Journal of Computer Vision, vol. 57, no. 2, pp. 137–154, 2004

[80] H.S. Bhatt, S. Bharadwaj, R. Singh, and M. Vatsa. Face Recognition and Plastic Surgery: Social, Ethical and Engineering Challenges.

[81] C. Chena , A. Dantcheva b , A. Ross. An ensemble of patch-based subspaces for makeup-robust face recognition. Elsevier B.V. Year: 2015.

Appendix

Published paper:

[1] Laure, K. Beli; Chunsheng, G. Enhancing Face Identification Using Local Binary Patterns and K-Nearest Neighbors. MDPI: Journal of Imaging 2017.