Facial Age Classification Using Geometric Ratios and Wrinkle Analysis

Shima Izadpanahi

Submitted to the
Institute of Graduate Studies and Research
in partial fulfillment of the requirements for the Degree of

Doctor of Philosophy
in
Computer Engineering

Eastern Mediterranean University
June 2014
Gazimağusa, North Cyprus

Approval of the Institute of Graduate Studies and Research

Prof. Dr. Elvan Yılmaz
Director

I certify that this thesis satisfies the requirements as a thesis for the degree of Doctor of Philosophy inComputer Engineering.

Prof. Dr. IşıkAybay
Chair, Department of Computer Engineering

We certify that we have read this thesis and that in our opinion it is fully adequate in scope and quality as a thesis for the degree of Doctor of Philosophy in Computer Engineering.

Asst. Prof. Dr. Önsen Toygar
Supervisor

Examining Committee

1.Prof. Dr.GözdeBozdağıAkar
2. Prof. Dr.Hakan Altınçay
3. Prof. Dr.A.EnisÇetin
4. Assoc. Prof. Dr. Hasan Demirel
5. Asst. Prof. Dr. Önsen Toygar

ABSTRACT

Age group classification is the process of automatically determining an individual's age range based on features extracted from facial image. It plays an important role in many real-life applications such as age specific human computer interaction, forensic art, access control and surveillance monitoring, person identification, data mining and organization, and cosmetology. In this thesis, we propose facial age classification approaches based on local and global descriptors extracted through feature selection methods.

This thesis proposes two different methods on facial age classification. The first proposed method is a novel and efficient age group classification approach that combines holistic and local features extracted from facial images. These combined features are used to classify subjects into several age groups in two key stages. First, geometric features of each face are extracted to construct a global facial feature. Support Vector Classifier (SVC) is used to classify the facial images into several age groups using computed facial feature ratios. Then, local facial features are extracted utilizing subpattern-based Local Binary Patterns (LBP) to classify adults. These combined features are used to classify subjects into six major age groups. The superiority of subpattern-based LBP over Principal Component Analysis (PCA) and Subspace Linear Discriminant Analysis (subspace LDA)techniques is presented.

The second proposed method presents geometric feature-based model for age group classification of facial images. The feature extraction is performed considering significance of the effects that age has on facial anthropometry. In this context,

Particle Swarm Optimization (PSO) technique is used to find optimized subset of geometric features. Age Classification on these features is evaluated using SVC. Wrinkle feature analysis is also applied to classify adult images. The facial images are categorized into seven major age groups.

The effectiveness and accuracy of the proposed age classification are demonstrated with the experiments that are conducted on two publicly available databases namely Face and Gesture Recognition Research Network (FGNET) and Iranian Face Database (IFDB). The experimental results show significant improvement of the proposed methods compared to the state-of-the-art models.

Keywords: Age group classification, feature extraction, Local Binary Patterns, Particle Swarm Optimization.

ÖZ

Yaş grubu sınıflandırması, bir insanın yaşını yüz resminden çıkarılan özniteliklere bağlı olarak otomatik bir şekilde belirleme işlemidir. Bu işlem gerçek hayattaki, insan tanıma, veri madenciliği ve organizasyonu, ve kozmetik gibi alanlarda önemli bir rol oynar. Bu tezde, öznitelik seçici yöntemlerle yerel ve evrensel metodları kullanıp yüz resimlerinden yaş sınıflandırması yapan yaklaşımlar önerilmiştir.

Bu tezde, insan yüzünden yaş sınıflandırma yöntemi üzerine önerilen iki yaklaşım vardır. İlk yöntem yeni ve etkili yaş grubu sınıflandırması için yüz resimlerinden bütünsel ve yerel öznitelikler çıkaran bir yaklaşımdır. Bu birleştirilmiş öznitelikler iki aşamada birçok yaş grubuna göre sınıflandırma yapmak için kullanılırlar. İlk aşamada, evrensel bir yüz özniteliği oluşturmak için herbir yüz resiminin geometrik öznitelikleri çıkartılır. Hesaplatılan yüz öznitelik oranları ve Destekçi Vektör Sınıflandırıcı (SVC) kullanılarak yüz resimleri birçok yaş grubuna göre sınıflandırılır. Daha sonra yetişkinleri sınıflandırmak için alt-örüntüye dayalı Yerel İkili Örüntü (LBP) yöntemiyle yerel yüz öznitelikleri çıkartılır. Bu birleştirilmiş öznitelikler kişileri altı ana yaş grubuna göre sınıflandırır. Alt-örüntüye dayalı LBP yönteminin Ana Bileşenler Analizi ve altuzay Doğrusal Ayırtaç Analizi yöntemlerine göre daha iyi oluğu gösterilmiştir. Önerilen ikinci yöntem yüz resimleri üzerinde geometrik öznitelikler kullanarak yaş grubu sınıflandırması yapan bir yaklaşımdır. Öznitelik çıkarma işlemi yaşlanmanın yüzdeki belirgin etkileri gözönünde bulundurularak yapılmıştır. Bu bağlamda, Parçacık Sürü Optimizasyonu (PSO) tekniği kullanılarak en iyi geometrik öznitelikler kümesi bulunmuştur. Bu öznitelikler, yaş sınıflandırma işleminde SVC yardımıyla değerlendirilirler. Ayrıca,

yetişkilerin yaş sınıflandırması için kırışıklık öznitelik analizi de uygulanmıştır. Bu işlemlerin sonucunda yüz resimleri yedi ayrı yaş grubuna ayrılmıştır.

Önerilen yaş sınıflandırma yöntemlerinin etkisi ve başarımı “Face and Gesture Recognition Research Network”(FGNET) ve“Iranian Face Database” (IFDB) veritabanları kullanılarak yapılan deneylerle gösterilmiştir. Deney sonuçları, önerilen yöntemlerin literatürdeki diğer yaklaşımlara göre daha iyi sonuç verdiğini göstermiştir.

Anahtar Kelimeler: yaş grubusını flandırması, öznitelik çıkarma, YerelİkiliÖrüntü, Parçacık Sürü Optimizasyonu.

ACKNOWLEDGEMENT

I would like to extend my sincerest gratitude to my supervisor,Asst. Prof.Dr. Önsen Toygar for her generous guidance in expanding my knowledge in Image Processing and her constructive advice and enthusiasm to share her insight and wisdom.

I also thank Assoc.Prof.Dr. Hasan Demirel and Prof.Dr. Hakan Altınçayfor their constructive guidance and support during the editing of the thesis.I am also indebted to my instructors in the Department of Computer Engineering at Eastern Mediterranean University for their guidance, and support throughout my studies.

This work is dedicated to my father, Kavous Izadpanahi and my mother, Shahnaz Bahmanyaras an indication of their significance in this study as well as through my entire life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS/ ABBRIVATIONS

(P, R) sampling points on a circle of radius R

$LBP_{(P,R)}$ value of the notable center pixel

g_p gray values of 8 equally spaced pixels on a circle of radius R

g_c gray value of the center pixel of a local neighborhood

R a circle of radius R

$\chi^2(X, Y)$ Chi Square distance

(X, Y) histogramsX and Y

n number of data vectors

xi data vectors

p number of columns

X data matrix

$m[j]$ mean along each dimension j

N number of training images

B matrix of size nxp

h column vector off all 1s

C covariance matrix of size pxp

B^T Transpose of matrix B

D MxM diagonal matrix of eigenvalues of C

V matrix of p column vectors

V^{-1} transpose of V

W eigenvectors of the projection matrix

B^T transpose of matrix B

S_w within-class scatter matrix

cov_j	covariance of class j
S_b	between-class scatter matrix
p_j	fraction of data belonging to class j
m_j	mean vector of class j
$D_{Manhattan}$	Manhattan distance between point P1 and P2
P1	point with coordinates (x1; y1)
P2	point with coordinates (x2; y2)
$D(x, y)$	Manhattan distance between two vectors X, Y
P_L	Most sideward point of the cheek bone(left)
M_L	Most sideward point at the angle of the lower jaw (left)
C	Lowest point of chinchin
M_R	Most sideward point at the angle of the lower jaw (right)
P_R	Most sideward point of the cheek bone (right)
E_L	Middle point of left eye
E_{ML}	Medial hinge of the eyelid (left)
E_M	Mid-point between two eyes
E_{MR}	Medial hinge of the eyelid (right)
E_R	Middle point of right eye
E_{LR}	Sideward hinge of the eyelid (right)
N	Tip point of nose
N_L	Most sideward point on the wing of the nose (left)
N_R	Most sideward point on the wing of the nose (right)
L_L	Most sideward point of the lips (left)
L_R	Most sideward point of the lips (right)
T	Top point of the head

L	Midline point where the lips
L_{LM}	Middle point of the lower margin of the lower lip
E_{LL}	Sideward hinge of the eyelid (left)
E_{LML}	Lowest point of the margin of the lower eyelid (left)
E_{HL}	Highest point of the eyelid (left)
D(A,B)	Euclidean distance between point “A” and point “B”
$R1_x$	Biometric ratios set
Vi	velocity for ith particle
P	size of the population
W	inertia weight,
c_1	acceleration constants
c_2	acceleration constants
$rand_1()$	random numbers
M	number of geometric ratios
Pbest	position giving the best fitness value of any particle
gbest	index of the best particle among all the particles in the population
Xi	ith particle
Wrinkle density	Wrinkle density in region ‘‘R’’
$Edge_R$	edge in region ‘‘R’’
$TotalPixel_R$	total pixel in region ‘‘R’’
HCI	Human Computer Interaction
ECRM	Electronic Customer Relationship Management
FGNET	Face and Gesture recognition research NETwork
IFDB	Iranian face database
LBP	Local Binary Patterns

PCA	Principal Component Analysis
subspace LDA	Subspace Linear Discriminant Analysis
AI	Artificial Intelligence
SVC	Support Vector Classifier
AAMs	Active Appearance Models
AGES	AGingpattErn Subspace
LBPH	Local Binary Patterns Histogram
FERET	Facial Recognition Technology
GOP	gradient orientation pyramid
LDA	Linear Discriminant Analysis
AG	Age Group
HE	Histogram Equalization
MVN	Mean-Variance Normalization technique
FLL	Facial landmarks location
AGC	age group classification
GFA	Geometric Feature Analysis
WFA	Wrinkle Feature Analysis

Chapter 1

INTRODUCTION

1.1 Motivation

Age is a natural factor affecting human faces. Age related research has been considered in problems such as facial age classification(age based estimation) and age invariant face recognition and facial age simulation [1-10].The general topic of facial image processing had been studied by many researchers. A number of researchers study facial age group classification and automatic human age estimation. Automatic human age estimation has become one of the most attracting topics in recent years due to its crucial role in many areas such as, age specific human computer interaction, indexing of face images according to age groups,machine learning, computer vision,development of automatic age progression methods andidentifying the progress of age perception by humans[1-4].There are many factors that make the process of age classification and age estimation even more challenging. The most important challenges are lack of proper large data set that contains enough images at different age groups,and illumination and pose variation across different human face images in datasets. The researchers show the inaccuracy in age estimation even by human observation [3, 11 and 12]. Although the calendar age that an individual has is fixed, the biological age is not. The procedure of aging can make a human to be some years younger or older than how a person ages based on date of birth.As such, even doctors make mistake in estimating age, up to 10 years. Increasing number of age classes,inadequate training face

images, personalized and temporal aging patternsare other factors that make the classification crucial.

The face expresses important relevant information for a person including gender, age, identity,ethnic background, emotions. Person's identity and ethnicity does not change through aging process. The same can be said for a person's gender while the age does not remain the same.As such, age classification is harder, being a temporal property.

Generally the human life is classified into one of three age periods: formative years (childhood),young adulthood, senior adulthood[1- 4, 11-13]. Since these three age groups do not share the identical optimal distinguishable features, the level of complication of classification increases. The facial transformation is considerablydifferent duringchildhood. During formative years, the cranium's shapegrowsin a relatively short time and the processes focus on changes on cranium'sdevelopment and ratios among facial features.

The problem of age group classification from facial images is an actively growing area of research and has not been studied in depth yet. There are many factors that make the age classification crucial.The most important challenges are1) lack of proper large data set that contains enough images at different age groups,2) variation in illumination and pose among face images in datasets. 3)the inaccuracy in age estimation even by human observation [3, 11, 12]. 4) Increasing number of age classes,5)inadequatenumber of training images,6) personalized and temporal aging patterns.

Age classificationmethods contain training and test sets to produce a model that can arrange the input facial images in classes according to their age. The aim is to use an algorithm that determines an individual's age range considering features extracted from facial images.Face age simulation involves the variation of shape and texture of a subject's face, which is useful in creating the age- progressed or age-regressed images of an individual. The goal in age simulation is the precise prediction of the face of a given imageduring growth procedure.

Age simulationmethods yield significantcontributions to several problems but have someserious complications due to nature of human growth. First, the process of aging is not the same for different people. As a result, the estimated future face using age simulation techniques may completely differ from the real one. Second, many other factors influence the process of human aging including abrupt changes in weight, harsh weather, stress, and other personal bad habits. As a result, generative methods are needed for an approximation of the age, which we cannot estimate at all. These approaches areavoiding the age simulation and trying toachieve a coarse estimate of an individual's age instead of precise age.

Age group classification techniques have become very important topics in recent years as a result of its significant role in many applications. It plays an essential role in various real-life applications including age specific human computer interaction, forensic art, security control and surveillance monitoring, person identification, human machine interaction, data mining and organization, and electronic customer relationship management. Typical applications for each category mentioned above are as follows:

1.1.1 Forensic art

The forensic art is the artistic method that can be used in a law enforcement investigation for identification of a wanted or suspected person to investigate how the person would look like today [14]. This person can be a criminal, a missingchildren [3,1,15] or an unidentified deceased person. The artist uses age progression to change and improve the outdated photosautomatically or manually. The natural age of the person can be estimated considering all gathered information on the suspect, such as personal habits, occupations, weight gain or loss, facial hair, hair dye, medical history, etc.Age progression systems require proper data related to the current age of a subject. As a result, an automatic age recognition system can play a significant role in enhancing the efficiency of the forensic artist.

1.1.2 Age-based security control

Surveillance systems and security control issues play an important role in our everyday life. An age determination system can analyze shopping customers and pedestrians with the help of a monitoring camera. For instance, to prevent those less than legal age from entering clubs or alcohol shops or stop teenagers buying cigarettes and alcohols from vending machines. Another example is to restrict children to ensure that they do not access to unsuitable materials such as some web pages or restricted films [2, 3]. In some countries, the connection between a specific age range and credit card fraud on Automated teller machines are found. Therefore,using camera systems to monitor and recognize the age of people can be very useful. Age recognition system is useful to provide accurate estimation of the subject’s agelooking for access to a particularfield[16].

1.1.3 Age Specific Human Computer Interaction

Human Computer Interaction (HCI) investigates a person and a computer in aggregation to increase the communication between users and machines. It is to study how people design, implement and use interactive computer systems and how machines affect users, organizations and society.Individuals of different age periods have quite different expectation while interacting with computers since they require different information. If computers could estimate the age of the user, the class of interaction mayaccustomedbased on user's age. For this purpose, age estimation system can play an important role for estimating an individual's age and automatically adjust the suitable user interface to fulfill the requests of a specific age range. For instance, use of clickable icons for young children who cannot read and write well as an interface or, considering text with larger font size for adults with low vision.

1.1.4 Data mining and organization

Automatic e-photo album sorting and retrieval of consumer photographs [5] is possible using age estimation systems. Photo groups can be selected by users from the auto-organized album based on the facial traitsclassification.

1.1.5 Electronic Customer Relationship Management (ECRM)

The ECRM [17] is a management approach for efficiently managing customers and communicating with them personally based on the information collected from them. Customers belonging to different age classes have different habits and preferences. Companies should collect and analyze sufficient data from customers of different age groups to respond directly to their needs in order to obtain more profits in marketing. For example, a mobile phone company wants to find out the interest and preferences of different age groups of customers to their new mobile products. The

advertisers might want to find out what percentage of customers is interested in particular advertisements. A shoe shop owner desires to findout which category of people is attracted in their products and providing customized services accordingly.

Obviously, this task is challenging and suffers from certain limitations. First, it requires establishing long-term customer relationships. Second, it is not cost beneficial as it needs a large amount of cost input. Finally, it violates customer's privacy. For this purpose, an automatic age estimation system can be established to capture customers snapshot and tag a human photo automatically according to the age range of the individual face.

1.1.6 Person Verification

In identification application, given two photos of a subject belonging to different age ranges, the certainty in verifying the identity of the individual after several years of age gap is investigated. [18-20] For example, this task is used in passport picture verification and border security [18] that involve photographs, where each photo is belonging to the same individual taken at different years apart.

1.2 Challenges

The literature review shows that the classification of images based on age suffers from certain limitations. First, either the data sets are small, or not publicly available. Second, the exact information of age labels is not available making age assignment subjective. Third, the age groups are selected by authors based on their objectives. Finally, the complexities are increased by conditions like inter-personal changes (i.e. ethnicity, gene), intra -personal variation (i.e. head position, face emotions), external factors (i.e. health condition, living location) and variation in acquiring image (i.e. illumination, image resolution).

Additional essential issue for developing an age classification systems is to consider the significance of age range. The characteristics of aging progress are not same for people belonging to not the same age ranges; therefore a method accomplished to deal with a certain age group may not be appropriate to utilize in different age groups.

This motivated us to present a new age group classification method without aforementioned problems. We would like to emphasize that the problem of having a proper method for achieving more particular age categories is quite a challenging problem. Hence, we focus on more particular age groups in our study.

1.3Proposed Contributions to Facial Age Classification

We have two different contributions to facial age classification adopting feature extraction techniques.

1.3.1 Effect of LBP on Facial Age Classification of Adult Faces

In this research, we propose an advanced and new age classification technique based on the features of human face progression that combines holistic geometric feature models and local binary patterns, to improve the accuracy rate of age classification. A main challenge we concerned in our study is how to achieve efficiency in both feature extraction and classifier training. We address this challenge using two publicly available databases namely FGNET[22],IFDB[21]. We carefully select different schemes for feature extraction that perform feature extraction for different age categories efficiently. This allows us to extract more accurate features prior to classification. Basically, the geometric features extracted from input facial images are used to discriminate young faces from adults. Later, the facial imagesidentified as adults are sent to the adult age-estimation part which utilizes subpattern-based Local

Binary Patterns (LBP) to classify adults. Experimental results show that the separation of input images into different age groups with similar characteristics helps the extraction and development of a more accurate age estimation system. The superiority of subpattern-based LBP over holistic methods such as PCA and subspace LDA techniques is illustrated as well.

Our approach is new in the following four ways. First, age classification precision is significantly enhanced using a combination of the recommended features. It appears that both the geometric features and local binary pattern approaches have only been applied individually, but, to the best of our knowledge, no study combining them considered in the earlier researches. As a result, for the aging feature extraction we used a new combination of global and local features as this method compensates for weaknesses found in using separate global and local features. This allows us to extract more accurate features prior to classification.

Secondly, the geometric feature extraction was performed with the understanding of the effect of age on facial anthropometry. Consequently, a new set of geometric features are extracted and proposed in order to improve the performance. Thirdly, we tried to choose a fast and efficient classifier for each stage of the proposed method.

Finally, the literature review shows that the classification of images based on age suffers from certain limitations. Either the datasets are small, or they are not publicly available. The exact information of age labels is not available making age assignment subjective. Moreover, the age groups are selected by authors based on their objectives. This motivated us to present a new age range classification method without aforementioned problems. The accuracy of combining geometric and local

facial features is demonstrated using two standard and publicly available databases for age estimation namely FGNET and IFDB.

Since the problem of having proper method for having more specific age ranges is still unsolved, we focus on more particular age groups in our study.The experimental results clarifies that the performance of the presented approach is significantly improved compared to the state-of-the-art approaches that used these aging databases.

1.3.2 Facial Age Classification with Geometric Ratios and Wrinkle Analysis

In this research, we introduce a novel age group classification approach using frontal face images taken from two large and publicly available databases namely IFDB [21] and FGNET [22]. For age group classification we use geometrical ratios and wrinkles in the face images as features. Geometric ratios are computed considering the facial measurements and the size of main face features to separate and discriminate young faces from adults. Using geometric ratios of the face instead of distances between the face features prevents the dependency on the image scale. Particle swarm optimization is an optimization technique that is used to find the optimal set of geometric ratios by creating particles that search a space of various ratios of the input facial image while attempting to produce the best accuracy rate on the data. The adult face images go through wrinkle analysis to be further classified into precise age groups. The age separation ability of the individual features (geometric and wrinkle features) and various combinations of the features are evaluated using SVC [23].

The literature review shows that the classification of images based on age suffers from certain limitations. First, either the data sets are small, or not publicly available.

Second, the exact information of age labels is not available making age assignment subjective. Finally, the age groups are selected by authors based on their objectives. This motivated us to present a new age group classification method without above mentioned problems.

Since the problem of having proper method for having more specific age ranges is important, we focus on more particular age groups in our study. The experimental results show that the performance of the presented feature selection method is significantly improved and increased the classification accuracy.

1.4 Thesis Overview

This thesis describes two different contributions to facial age classification. Chapters are ordered as follows: Chapter 2 provides a detailed review of the related studies,and focus on some of the difficulties that investigators face in this domain. The method presented is principally a classification approach of facial images after extraction of the features. Therefore, holistic and local approaches on extraction of the features are discussed in Chapter 3.Chapter 4 represents a novel and efficient facial age group classification approach that combines holistic and local features extracted from facial images. These concatenated features are used to classify individuals into several age groups. Chapter 5 presents an optimal geometric ratios and wrinkle analysis for age range classification of facial images. The feature selection is performed considering PSO technique to find optimized subset of geometric features.Finally Chapter 6 presents the overall conclusions.

Chapter 2

HUMAN AGE ESTIMATIONMETHODS

2.1 Introduction

The existing age estimation approaches in the literature can be separated into two main steps: aging feature representation, and age estimation. The most common categories of image representation are anthropometric models [8, 9, 54-56], Active Appearance Models (AAMs)[24], AGingpattErn Subspace (AGES)[25], and appearance models [28]. The second stage for age estimation is to classify the age based on features extracted using classification algorithms.

2.1.1 Anthropometric Models

Anthropometric model utilizes cranio-facial development theory and was first proposed by Kwon and Lobo [8]. They used six biometric ratios of key features of the facial image to distinguish infants from adults. Their dataset consists of only 47 high resolution face images. The complete classification scheme was tested on 15 face images of babies, mid-age adults, and seniors. They used deformable templates and snakelets, which is computationally expensive and not appropriate for real time processing. In the experimental results they didn't present the overall accuracy rate on this small database, howeverpresented the results by using individual ratio.

Ramanathan and Chellappa in [54] proposed a craniofacial growth model that illustrates age development related shape differences perceived in subjects younger than 18 years of age. The problem tackled by the authors in this work is targeted at

aging developments, primarily shown in the form of shape variations from 0 to 18 years for given input images.

Ramanathan et al. [54] mentioned in their work that the craniofacial growth related studies are mainly depends on the assumption that any identifiable style of variation can be individually stated by some geometric invariants, forming the origin of perceptual data. The authors of [54] further assume that these geometric invariants can be detailed in the following three statements: (1) angular coordinates of each point on an object in a polar coordinate system is maintained.(2)bisideward symmetry about the vertical axis is obtained (3) continuity of object contours are maintained.

Using aforementioned hypothesis, Ramanathan et al. [54] used the following three aspects while modeling their craniofacial growth model based on a fluid filled spherical object model presented by authors in [38]. (1) pressure is directed radially outwards (2) pressure distribution is bisidewardly symmetric about the vertical axis (3) pressure distribution is continuous throughout the object.

The authors in [54] argued that when this model is applied on the face image of a kidbychanging the growth parameter, it was seen that the evolution of facial images closely look like the real facial growth. They assert that the observed age of every different face images improved with the increase in their growth parameter. The larger growth parameter value would mean a larger age transformation. It is seen in their work that, although age transformation was observable in the early few transformations, the feature ratio of face images became unnatural for larger age transformations.

In this research the authors also made use of face anthropometry. As their work dealt with only profile facial images, only the facial landmarks which could be consistently located using photogrammetry were considered. Hence from a total of 57 landmarks defined in [3] on the human face, they selected 24 landmarks in their work. The facial features such as the eyes, the nose and outer contour of the face were detected using eclipses of different scales and locations. Then using the ratio indices achieved using distance metrics between the obtained landmarks, estimates of the age group offacial images are performed. Ramanathan et al. further claimed that small errors in feature localizations did not affect the performance of their presented model much. They have provided the experimental results using an aging database containing the facial images of subjects less than 18 years of age. This database consists of a total of 233 images of 109 subjects and is maintained by FGNET aging database. [22]

Later, authors in [9, 55, and 56] used the distance ratios measured from facial landmarks for age categorization. Dehshibiand Bastanfard[55] used an anthropometric based model and added wrinkle features to their method to improve the age recognition task. They only classified IFDB frontal facial images into four age categories with accuracy of 86.64%.

The craniofacial growth related studies were mainly based on the assumption that any distinguishable change can be distinctively specified by some geometric invariants, forming the foundation of perceptual information. It is a well-known fact that age estimation techniques based on anthropometry model achieve high accuracy dealing with facial images where the shapes of faces develop largely.

2.1.2 Active Appearance Models(AAMs)

Utilizing the AAMs scheme [24], Lanitis et al. [3] presented a comparative performance evaluation of different classifiers in automatic age estimation task. The frontal images were represented by the AAMs approach and the greatest results were achieved using quadratic based and neural network classifiers. In their approach, the facial images ranging from 0 to 35 years were classified into three age categories. Therefore, the age estimation ability of “35 above” adult faces leaving in doubt.

2.1.3 AGingpattErn Subspace (AGES)

AGES method is another technique which firstly uses AAMs technique to encode each input facial images. However, the difference of this method from AAMs is to use face image of the same person at different ages. Authors in [25, 26] explored this idea, for estimation of the age automatically,forming the aging patterns using a representative subspace. They define the “aging pattern” as a set of individual facial images ordered along the time. AGES method is defined to describe the chronological facial images and as a resultobtain the differences generally seen in aging images.

Since obtaining a person’s face images for all the yearsis difficult, obtaining the complete aging pattern is a left as a problem. Geng et al. [25,26]presented the ‘aging pattern subspace’ using the methods that develop an eigenspace [27] using facial image with are not complete.Byutilizing the subspace, the aging pattern and the age of an unknown face image presented for the first time to test the system is identified by projecting it into the subspace that reconstructs the face in a best way.

They classify the facial images from FGNET [22] and MORPH [29] databases into four different age groups from 0 to 9 years, 10 to 19 years, 20 to 39 years and 40 to

69 years. This technique simulates the aging pattern which is defined as a sequence of the same person's facial image arrangedalong the time by making a representative subspace. Although, AAMs technique [24] considers both shape and texture in the task of classification, they did not facial creases for adults. This shortcoming is occurred as the AAM technique only considers the image intensities without considering any texture patterns at local regions. Additionally, due to global characteristics in AAM texture, this method is sensitive to partial occlusion and illumination variation. In order to overcome this problem the texture patterns at spatial neighborhood requires to be calculated.

2.1.4 Appearance Models

The effective LBP operator [32, 33] is one of the finest performing appearance models for texture description. It has been gaining interest lately in many areas such as image classification, face recognition [34, 35], age and gender classification [10]. It is highly discriminative and robust against pose or illumination variations. Authors in [10] have utilized Local Binary Patterns Histogram (LBPH) for classifying facial images using age, gender and ethnicity information. They classified facial images into three age categories as child, youth and oldness. Age classification is performed by using a binary tree structure that first classifies child from adult and then differentiating old age ranges from young. The experiments conducted with the use of three databases in order to categorize gender (as male or female), age (as three age periods) and ethnicity (as non-Asian or Asian) successfully considering LBPH features. In all the experiments the LBPH feature is compared with Haar like feature. Results proved that LBPH feature performs much better in age classification.

Later, authors in [10] used LBP analysis to identify the age from frontal face image

samples. The facial images were separated into several identical sized blocks and shown as the mixture of LBP histogram features from all blocks. They used minimum distance, nearest neighbor and k-nearest neighbor classifiers in the classification stage of their algorithm. Global and spatial LBP histograms were generated to classify the subjects into different age groups. They classify the Facial Recognition Technology (FERET) [11] images based on their ages with 10 year intervals with 80% accuracy using nearest neighbor classification. Guo et al. [2] presented the age estimation of human using manifold learning method for extracting facial aging features and model a locally adjusted robust regressor for determining the future face image age. For this purpose, two different databases were used, an internal age database and the FGNET database.

While, a number of stand-alone local matching approaches have been presented by researches, combined features [36, 37] are more effective and robust in age categorization. Authors in [37] used both local and holistic features extracted from face for classifying the images. However, they applied these extracted features that coarsely classify a face image into only two age groups from 0 to 20 as young or from 21-69 as old.

Chapter 3

FEATURE EXTRACTION METHODS

3.1 Introduction

A key issue in designing a successful facial age classification is to find efficient feature extraction methods concerning the age periods. There are many significant factors which impact the aging procedure during human life[41, 57]. Generally, facial aging can be separated into two main stages: remodeling, and adulthood aging [59]. The facial developments are very faster in early ages in comparison to adulthood. During childhood, facial shapes undergo significant variations and the feature extraction focuses on alternation on cranium's size and geometric ratios of facial features.

With the termination of growing procedure at approximately 20 years of age, facial shapes and the distance between the main features change with a limited amount. Furthermore, the skin experiences an array of changes in appearance and texture in the form of fine lines and creases. Accordingly, the separation of input images into different age groups with similar characteristics helps the extraction and development of more accurate age classification system.

Two different methods have been suggested to extract facial features: Subpattern-based and holistic methods. In the following sections, LBP is explained as a Subpattern-based method and thenPCA and subspace LDAare described

3.2 Subpattern-based Methods

Subpattern based methods mainly divide the facial images into equal size non-overlapped partitions.Since different features of face may contribute not in a similar mannerin recognition task, these partitions are individually experimented in order to obtainlocal features. The output of individual partition corresponding to the local projectionis combined into aoverall feature of the original facial image using Local Binary Patterns for further classification.Latera classifier technique is used for classifying of facial images. For this purpose, the global features related to the training and test image projections are compared using those classifiers.

The main idea of using partitioning-based approaches for facial image description is in spiredfrom the fact that human face can be viewed as a combination of small parts which can be described very well using LBP operator.

Subpattern based methods can be implemented with different number of partitions. An example of partitioned facial image with different number of subpatterns is shown in Fig.1

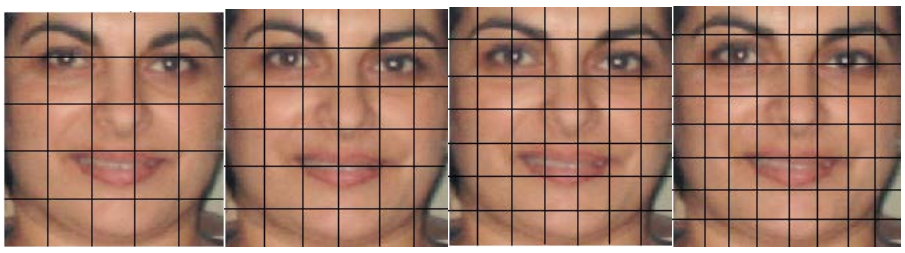


Figure 1. Image samples of partitioned images applying different number of subpartitions (5x5, 6x6, 7x7, 8x8)

3.2.1 Local Binary Patterns

LBP operator is one of the best performing texture operators firstintroduced by Ojala et al. [32]that divide the facial image into equal-width non-overlapped subpartitionThe LBP texture descriptors are calculatedfor each partitionindividually.

These extracted descriptors are then combined to form a global description of the facial image.

For creating local descriptors, LBP extracts the features from the image by giving a label to each pixel of an image by thresholding pixels in eight neighborhood of each pixel with the value of the center pixel. Afterwards, the histogram of the labels is used as a texture descriptors or features. Lastly, all histograms are concatenated to build a code image.

Later in [42],the LBP operator was extended to use an arbitrary number of interpolated pixels on a circle with arbitrary size as neighbor pixelsin order to deal with textures at different scales. In the cases that sampling point does not fall in the center of a pixel, bilinear interpolation is used [43].

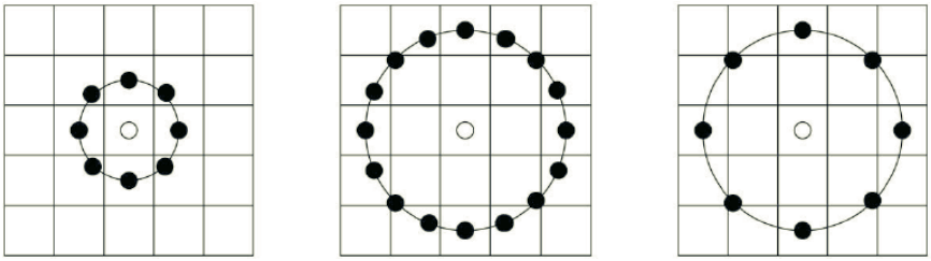
A circular neighborhood and bilinear interpolating values at non-integer pixel coordinates let any radius and number of pixels in the neighborhood. The notation (P, R) is used for pixel a neighborhood which means P sampling points on a circle of radius R.An example of different circular neighborhoods with different scales is illustratedin Fig.2.

3.2.2 LBP Algorithm

The LBP code construction for each pixel of an image consists of three main stages:

Stage 1: Firstly the input image is divided into local partitions and texture descriptors are obtainedfor everypartition individually. The size of each partition and the number of sub-partitions in a given image should be determined. The total number of partitions in an image withAxB pixels where “A” and “B” are the pixel values of rows and columns is computed by having the size of the region.

For instance, we assume an image is a matrix 40x35 pixels which contains of 40 pixels in rows and 35 pixels in columns. Fig.2 shows an instance of a facial image divided into sub-windowsof size 5x5 pixels. In this case, the image size is divided by region size (40x35/5x5) and as a result 8x7 subpartitions are obtained.It should be mentioned that the partitions do not require being rectangular and they can be of different size or shape and they do not need to cover the entire image. For instance, they could be circular partitions located at fiducial points as illustrated in Fig. 2 [44].



a) b) c)

Figure 2. Extended LBP operator, Example of a) (P, R) = (8, 1), b) (P, R) = (16, 2), c) (P, R) = (8, 2) circular neighborhoods.

Stage2: In this step block processing should be applied independently on each region extracted in the previous step. Basically the middle pixel, the number of sampling points on the circle of desired radius should be defined as illustrated in Fig 2.The LBP operator allocates an 8-bit binary code to every pixel of an input image in the corresponding region by comparing the center pixel value with 8 neighbor pixels. Whenever the middlepixel's value is greater than its neighbor, the label is 1; otherwise it is assigned0, as shown in Fig 3. This assigns an 8-digit binary number and can be converted to decimal formfor conveniences.

As demonstrated in Fig.4, the LBP value of the notable middle pixel in the 8 neighborhood is achieved by converting the binary code into a decimal value using

following equation:

$$LBP_{(P,R)} = \sum_{P=0}^{7} S(g_p - g_c)2^P$$

$$S(g_p - g_c) = \begin{cases} 1, & g_p > g_c \\ 0, & g_p < g_c \end{cases}$$

where g_c corresponds to the gray value of the middle pixel of a local neighborhood. g_p (p = 0,...,7) corresponds tothe gray values of 8 equally spaced pixels on a circle of radius R(R > 0) that form a circularly symmetric set ofneighbors.The notation (P, R) is used for pixel neighborhoods which mean P sampling points on a circle of radius R.

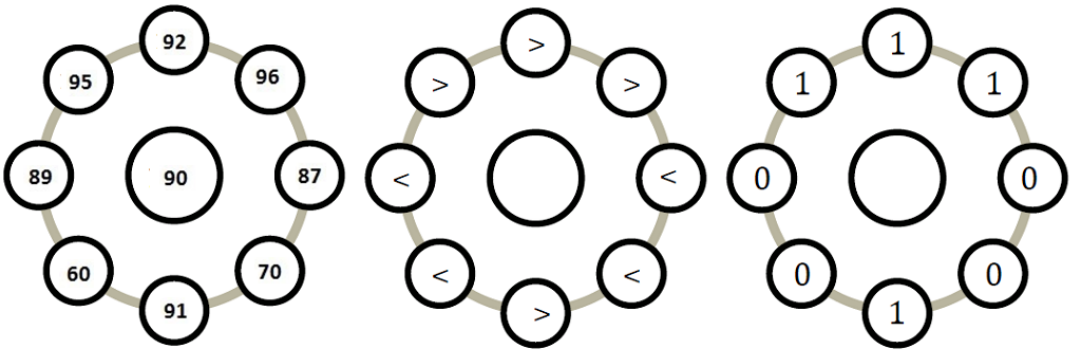


Figure 3. Assigning the 8-bit binary code

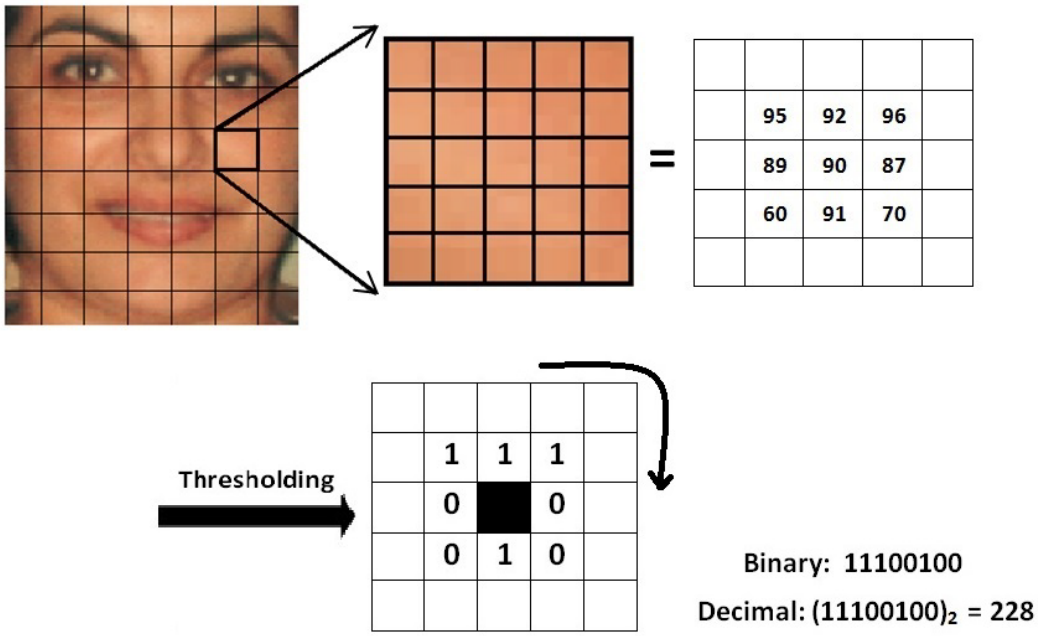


Figure 4. The basic LBP operator

Later, the basic LBP operator is extended to the so-called uniform LBP [42]. A local binary pattern is called uniform if there are at most two bitwise transitions from 0 to 1 or 1 to 0 in its binary code. Some patterns of uniform and non-uniform are given in Table 1.

Table 1. Basic LBP Operator

Pattern	Number of transitions	Type of pattern
11111111	0	Uniform
01111111	1	Uniform
01110000	2	Uniform
11001001	4	Non-uniform
01010011	5	Non-uniform

Uniform pattern is useful in reducing the length of feature vector.

Stage3: In this step, the occurrence histograms of theLBP labelsobtained in the previous step are used as texture features or descriptors. The spatially enhanced

histogram will be obtained from the basic histogram which is used as a face description and encodes the spatial relations of facial regions.

The LBP labels for the histogram hold information related to the patterns on a pixel-level. In order to obtain data on a local region, for each of the partitions a histogram is extracted individually. For this purpose, the labels are summed over these small regions.

Finally different local histograms are concatenated to form a global code image of the face. This concatenated form is the feature vector for the image. Its main power which is robustness against variations in pose and illumination makes it appropriate for image analysis and age classification approaches.

The original image and LBP code image are demonstrated in Fig. 5. In the code image, the original image and the combined descriptors to form the code image are illustrated.

Figure 5. Original image and LBP code image

We used a nearest neighbor classifier and Chi Square distance metrics for classifying adult face images into different age groups. The distance between two different

histograms X and Yusing Chi Square distance is obtained as follows:

$$\chi^2(\mathrm{X},\mathrm{Y}) = \sum_{\mathrm{i}=1}^{\mathrm{n}} \frac{(\mathrm{X_i} - Y_\mathrm{i})^2}{X_\mathrm{i} + Y_\mathrm{i}}$$

where n is the number of elements in the histogram. Chi square is a useful measurement of likelihood between a pair of histograms; therefore it is appropriate for nearest neighbor classification.

3.3 Holistic Approaches

Holistic matching or appearance-based approaches use the entire face area as the input to classification scheme. They attempt to classify faces using global representations and are extensively used in face classification applications. These methods obtain features from the entire facial image, decrease the size and then categorize them accordingly. In this method, after extracting the features, the pattern classifiers are used to classify the facial images. A key potential advantage of appearance-based methods compared to feature-based ones is to have any information about the facial images to distinguish them from others. However, there may be redundant data in the image which will affect the performance of the face classification approach.

While using holistic methods, an image of size AxB pixels is represented by a vector in an A.B-dimensional space. Practically, on the other hand, these matrices composed of every pixel of the image are too large in size to permit efficient and fast face classification. The most important problem in working on a space of high dimensionality is overfitting. Another possible problem could be the increase of computational complexity while working on large databases.

In order to solve this problem, appropriate statistical dimensionality reduction

techniques are useful in order to reduce the dimension of the studied domain. In this study, two of the most widely used holistic techniques namely PCA and subspace LDA are used.

These projection based techniques was required because of the limitations of the straightforward approaches based upon template matching. In order to avoid the curse of dimensionality, the face images are then projected and compared in a low-dimensional subspace.

PCA and Subspace LDA are proved to be suitable for many areas such as face classification. In the following subsections, these two approaches are discussed in detail.

3.3.1 Face Classification by PCA

Principal Component Analysis (PCA) is one of the best and general projection based techniques used in many applications such as face recognition [45], image compression [46], video coding and compression [47]. A key main advantage of using PCA is that it uses a simple linear algebra reducing a complex data set to a lower dimension space without much loss of information.

Prior to explaining PCA, it is required to explain eigenvector and eigenface terms. Eigenfaces is the name assigned to a set of eigenvectors. A non-zero vector C is called eigenvector of a square matrix A if and only if there exists a real or complex number λ such that the following equation holds:

$$AC = \lambda C$$

If such a number λ exists, it is called an eigenvalue of A. The vector C is the eigenvectorof A associated to the eigenvalue λ and it is not a zero vector.If we consider the squared matrix A as a transformation matrix,then multiplying it with

theeigenvector will notchange its direction. The reason is that if you scale a vector by some value, the vector is just getting longer without changing its direction.

Finally, all the eigenvectors of a matrix are orthogonal regardless of the dimension.It should be noted that eigenvectors can only be calculated for square matrices. In addition, not every square matrix has eigenvectors.For instance, there are m eigenvectors available for a given mxm image.

Turk and Pentland in [48] proposed the Eigenface Method based on the Karhunen-Loeve expansion, which they worked on the entire image.A set of eigenimages can be generated by applying PCA on a group of different facial images. For simplicity we can consider that eigenfaces are standard faces, resulting from statistical analysis of several facial images. Each facial image can be considered to be a mixture of theaformentioned standard faces. As an example, a person's face might be a combination of the average face and 12% of the firsteigenface, 45% of the secondeigenface, and even -5% of the thirdeigenface.

Unusually, many combinedeigenfaces are not required in order to obtain a reasonable approximation of a face. Moreover, since an individual’s face is recorded as just a list of values, lesser memory space is needed for keeping each person's face. As illustrated in Fig. 6the image of the eigenfacea bit resembles the original facial image.

Figure 6. Original images (the first row), and the corresponding Eigenfaces obtained (the second row)

Principal component analysis is a mathematical process which treats the facial images as two dimensional data, and categorizes the face images by projecting them to the eigenfacespace. This space is composed of eigenvectors obtained by the variance of theface images.It transforms the data into a new coordinate systemin a way that the greatest variance by any projection of thedata comes to lie on the first coordinate (called the firstprincipal component), the second greatest variance onthe second coordinate, and so on [49]. In other words, this transformation is defined such that the first principal component has as high a variance as possible (that is, accounts for as much of the variability in the data as possible), and each succeeding principal component in turn has the highest variance possible under the constraint that it is orthogonal to the preceding components.

The number of principal components is selected less than or equal to the number of images in the training set. PCA can be used fordimensionality reduction in a dataset while retaining thosecharacteristics of the dataset that contribute most to itsvariance, by keeping lower-order principal components andignoring higher-order ones.

3.3.2 Principal Component Analysis (PCA) Steps

The steps required to perform a principal component analysis on a set of data are as

follows:

Step1:Read images

Suppose that all the images in the dataset are arranged as a set of n data vectors $x_1,x_2,\ldots,x_n$ with each x_i defining a single grouped observation of the p variables. Take x_1 ,$x_2,\ldots,x_n$as row vectors, each of which has p columns. The resulting nxp dimension matrix is referred to as X.

Step2:Calculate the mean of images

Calculate the mean m along each dimension j= 1,…,p which is the sum of all the training images divided by the total number of training images N. The calculated mean values are represented into a vector m of size px1as

$$m[j] = \left(\sum_{i=1}^{n} X[i,j]\right) * 1/N$$

Step3:Mean subtraction

Mean subtraction is required to make sure the initial principal component defines the direction of maximum variance. Subtract the vector m from each row of the data matrix X and save the result in the n × p matrix B as

$$B = X - hm^T$$

where h is an nx1 column vector of consists of ones.

$$h[i] = 1 \qquad \text{for} i = 1, \ldots, n$$

Step4:Calculate the covariance matrix

Calculate the pxp covariance matrix C using following formula:

$$C = \frac{1}{n-1}\, B.B^T$$

where B^T is transpose of matrix B. Based on Bessel's correction [50] we are using n-1 instead of n in the formula.

Step 5: Calculate the eigenvectors and eigenvalues of the covariance matrix

The matrix V of eigenvectors contains useful information about our data which diagonalizes the covariance matrix C. As the covariance matrix C is square, we can easily find the eigenvectors and eigenvalues for this matrix.

$$D = V^{-1}CV$$

The matrix D is MxM diagonal matrix of eigenvalues of C. Matrix V is pxp dimension andconsists of p column vectors.Each column vector has p length, which shows the p eigenvectors of the covariance matrix C.

Step 6: Sort the eigenvectors and eigenvalues

While maintaining the correct pairings between the columns in each matrix,sort the columns of the eigenvector matrix V by eigenvalue matrix D,highest to lowest.This gives you the components in order of significance.

Step 7: Choose components and form the basis vectors

The eigenvector with the highest eigenvalue is the principle component of the data set. At this time, it is possible to ignore the components of lesser importance. This causes loss of some information, but if the eigenvalues are small, the loss is insignificant. Leaving some components out, cause the final data set to have less dimensions than the original.

A basis vector is constructed by taking the first L eigenvectors of V that are chosen from the list of eigenvectors, and forming a pxL matrix W with these eigenvectors in the columns.

$W[k,l] = V[k,l]$for $k = 1, \dots, p$ and $l = 1, \dots, L$

where $1 \leq L \leq p$.

Step 8: projection

By considering the similarity degree, the projection matrix of every training image is

compared with the test image's projection matrix. The output will be the image with the maximum similarity to the test image.

3.3.3 Subspace Linear Discriminant Analysis (Subspace LDA)

Linear Discriminant Analysis (LDA) [51] is a commontechnique used in many areas such as statistics, pattern recognition and machine learning for classification of data. The principle of LDA is to construct a subspace and find a linear combination of features which separates two or more classes of images. The facial images are projected and compared in a low-dimensional subspace. The resulting combination may be used as a linear classifier, or more generally, for dimensionality reduction before later classification.

LDA is closely related to PCA as both look for linear combinations of variables which best explain the data [52]. PCA is based on the sample covariance which characterizes the scatter of the whole dataset; regardless of class-membership (i.e. it does not take into account any difference in class). As such, the projection axes chosen by it might not lead to sufficient discrimination power. On the other hand, LDA explicitly attempts to model the difference between the classes of data. It tries to find directions along which the classes are best separated. It does so by preserving the maximal class discriminatory information and decision region between the given classes by maximizing the ratio of between-class scatter to the within-class scatter in any particular data set. In order to obtain the variance between the items in the same class within-class scatter matrix is used. On the other hand, to obtain the amount of variance between classes the matrix of between-classes scatter is calculated [53]. Fig.7 illustrates an example of three different classes of facial images classified using LDA. Each row shows different age groups.

The first row belongs to child subjects, the second row are young and the third row adulthood subjects. Large variance between classes and small variance within classes are obvious.

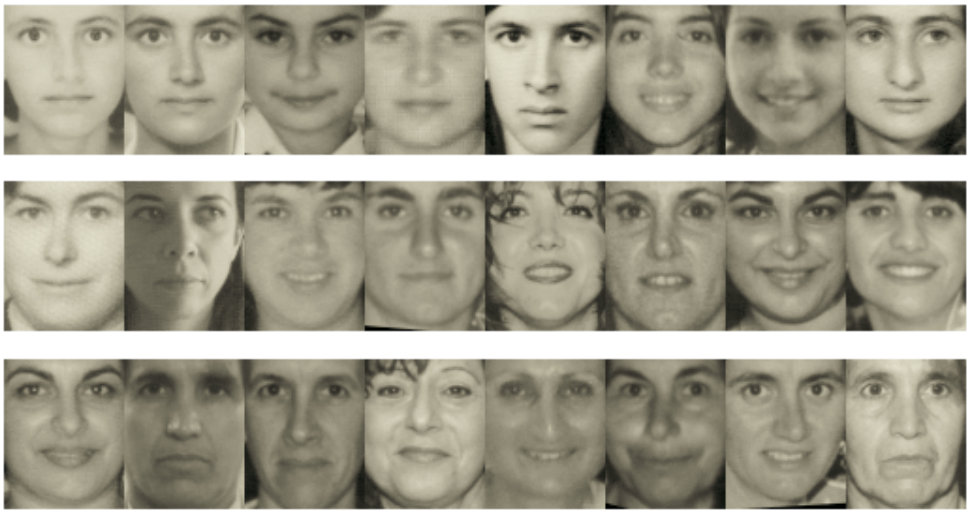


Figure 7. Three age classes, first row belongs to child subjects, second and third row belongs to young and adulthood subjects respectively.

LDA classifies the unseen classes in test images based on the seen classes in training images. This method helps to better understand the distribution of the feature data and to find the most discriminant projection direction.

Subspace LDA method consists of two steps. The first step is to project the facial images from original vector space to a low-dimensional facesubspace using PCA. The second step is to use LDA to find a linear classifier in the subspace.

3.3.4 Subspace Linear Discriminant Analysis (Subspace LDA) steps

The steps required to perform a principal component analysis on a set of data are as follows:

Step1: Read images

Suppose that all the images in the dataset are arranged as a set of n data vectors x_1 ,x_2,…,x_n with each x_i defining a single grouped observation of the p variables. Take

x_1 ,x_2,…,x_n as row vectors, each of which has p columns. The resulting nxp dimension matrix is referred to as X.

Step2: calculate the mean

For PCA to work precisely you have to calculate the mean m along each dimension j= 1,…,p which is sum of all the training images divided by the total number of training images N. Write the calculated mean values into a vector m of size px1.

$$m[j] = \left(\sum_{i=1}^{n} X[i,j]\right) * 1/N$$

Step3:Mean subtraction

Mean subtraction is required to make sure the initial principal component defines the direction of maximum variance. Subtract the vector m from each row of the data matrix X and save the result in the n × p matrix B as

$$B = X - hm^T$$

where h is an nx1 column vector of consists of ones.

$$h[i] = 1 \qquad \text{for} i = 1, \dots, n$$

Step4:Calculate the covariance matrix

Calculate the pxp covariance matrix C using following formula:

$$C = \frac{1}{n-1}\, B.B^T$$

where B^T is transpose of matrix B. Based on Bessel's correction [50] we are using n-1 instead of n in the formula.

Step 5:Calculate the eigenvectors and eigenvalues of the covariance matrix

The matrix V of eigenvectors contains useful information about our data which diagonalizes the covariance matrix C. As the covariance matrix C is square, we can easily find the eigenvectors and eigenvalues for this matrix.

$$\mathrm{D} = \mathrm{V}^{-1}\mathrm{CV}$$

The matrix D is MxM diagonal matrix of eigenvalues of C. Matrix V is pxp dimension and consists of p column vectors. Each column vector has p length, which shows the p eigenvectors of the covariance matrix C.

Step 6: Sort the eigenvectors and eigenvalues

While maintaining the correct pairings between the columns in each matrix,sort the columns of the eigenvector matrix V by eigenvalue matrix D,highest to lowest.This gives you the components in order of significance.

Step 7: Choose components and form the basis vectors

The eigenvector with the highest eigenvalue is the principle component of the data set. At this time, it is possible to ignore the components of lesser importance. This causes loss of some information, but if the eigenvalues are small, the loss is insignificant. Leaving some components out, cause the final data set to have less dimensions than the original.

A basis vector is constructed by taking the first L eigenvectors of V that are chosen from the list of eigenvectors, and forming a pxL matrix W with these eigenvectors in the columns.

$W[k,l] = V[k,l]$ for $k = 1, \dots, p$ and $l = 1, \dots, L$

where $1 \leq L \leq p$.

Step 8: projection

Use the obtained projection from PCA and supply it to LDA as input data.

Step 9: Find within-class scatter matrix

The within-class scatter matrix S_w of X is calculated using the following formula:

$$S_w = \sum_{j=1}^{J} p_j \times (cov_j)$$

where cov_j is the covariance of class j and is calculated as follows:

$$cov_j = (x_i - m_j)(x_i - m_j)^T$$

Step 10: Find between-class scatter matrix

The between-class scatter matrix$S_b$can be thought of as the covariance of data set whose members are the mean vectors of each class which is obtained using the following equation:

$$S_b = \sum_{j=1}^{J} p_j (m_j - m)(m_j - m)^T$$

where j is the number of classes,$p_j$is the fraction of data belonging to class j, m_j is the mean vector of class j, and m is the mean of all vectors.

Step 11: Calculate the eigenvectors of the projection matrix

An eigenvector of the projection matrix represents a 1 dimensional invariant subspace of the vector space in a way the projection is applied. The eigenvectors of the projection matrix is computed using the following equation:

$$W = eig(S_w^{-1} S_b)$$

Step 12: By considering the similarity degree, the projection matrix of every training image is compared with the test image's projection matrix. The output will be the image with the maximum similarity to the test image..

Generally the algorithms based on LDA are superior to those based on PCA. The reason is that LDA deals directly with discrimination between classes, whereas PCA deals with the data in its entirety for the principal components analysis without paying any particular attention to the underlying class structure. It should be noted that in the cases that the training data set is small, PCA can outperform LDA. Additionally, PCA is less sensitive to different extracted components (features) that well describe the pattern.

In this study, the similarity measure used for PCA and subspace LDA is Manhattan

Distance. The Manhattan Distance between two points is the sum of the differences of their corresponding components.

The Manhattan distance between the point P1 with coordinates (x1; y1) and the point P2 at (x2; y2) is:

$$\mathrm{D}_{\mathrm{Manhattan}} = |\mathrm{x}_1 - \mathrm{x}_2| + |\mathrm{y}_1 - \mathrm{y}_2|$$

and Manhattan distance [19]between two vectors X, Y is:

$$D(x, y) = \sum_{i=1}^{n} |x_i - y_i|$$

where x and y are vector with length n.

Chapter 4

FFFECT OF LBP ON FACIAL AGE CLASSIFICATION

4.1 Introduction

Facial age classification determines an individual's age range according to features extracted from input facial images.A number of approaches have already been used for a computer-based solution for the problem of age estimation from facial images [41], but complexities added by circumstances like increasing number of classes, inconsistency of acquisition conditions, and lack of suitable database which cover enough age range, have made the task quite challenging.

In this chapter we propose a novel and efficient age group classification approach that combines holistic and local features extracted from facial images. These combined features are used to classify subjects into several age groups in two key stages. First, geometric features of each face are extracted to construct a global facial feature. SVC is utilized to classify the facial images into several age rangeswith the use ofextractedgeometric ratios extracted from the face images. Then, local facial features are extracted utilizing subpattern-based Local Binary Patterns to classify adults. It appears that both the geometric based features and Subpattern-based approaches have only been studied individually, but no research combining them has yet been conducted in the previous works.The combined features proposed in this study allow us to extract more accurate features prior to classification.

Moreover, the geometric feature extraction was performed with the understanding of the effect of age on facial anthropometry. Accordingly, in order to increase the performance, a new set of geometric features are extracted and proposed.In the classification stage we tried to choose a fast and efficient classifier for each stage of the proposed method. The proposed combined features are used to classify subjects into six major age groups.The superiority of subpattern-based LBP over principal component analysis and subspace linear discriminant analysis techniques is presented.

The efficiency and accuracy of combining geometric and local facial features are demonstrated using two standard and publicly available databases for age estimation namely FGNET and IFDB. Furthermore, some insight and comparison, into which type of features – holistic or local – is better for age estimation, are also provided.

4.2 Proposed Method

Age group classification process highly depends on finding a set of reliable features as the bases for age classification. For this purpose, some novel features that are obtained by extracting geometric ratios based on face anthropometry combined with LBP operator are used in the proposed method. The block diagram of the complete system is shown in Fig.8, which classifies input facial images into six well-ordered age intervals. The process of the method is briefly outlined in this section, consisting of four key stages: preprocessing, geometric feature extraction, local feature extraction and age classification. Preprocessing is the first stage to provide a standard face image prior to feature extraction. The preprocessing steps required are detailed in the subsequent section.

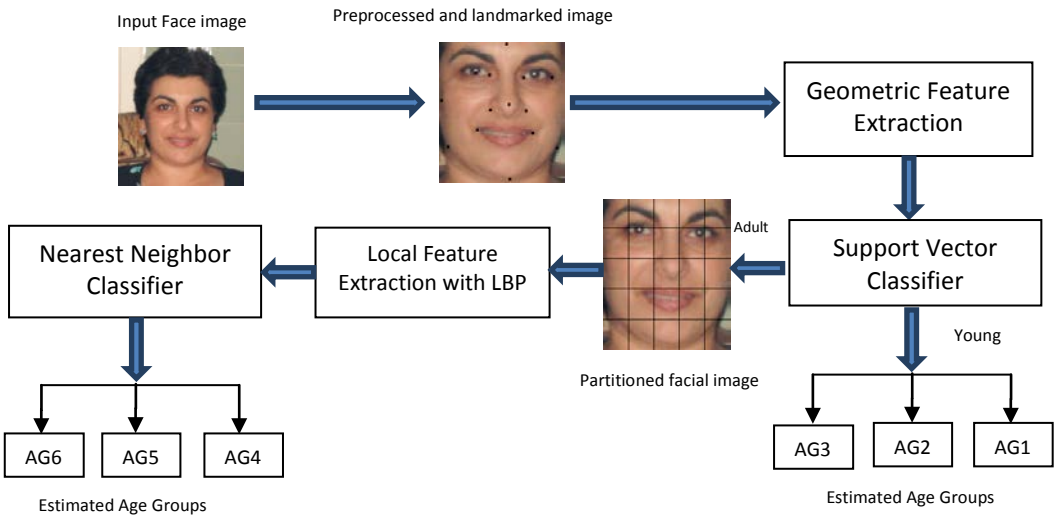


Figure 8. Block diagram of the proposed method

In the second step, using anthropometric properties of the human face, some geometric ratios are extracted to categorize young faces “below 20” from adult faces. Since anthropometric based models achieve high accuracy dealing with young ages, they are appropriate for this step of our algorithm. The ratio features extracted are applied to Support Vector Classifiers (SVC) to categorize images into four age intervals: AG1, AG2, AG3 and adults. Details of the geometric ratio extraction and age classification are given in Section 4.2.1.

The adult facial images need to go through local feature extraction presented in Subsection 4.2.2 to be further classified into appropriate ages. For this purpose, the local features extracted with LBP from “above 20” facial images are applied to nearest neighbor classifier to classify the facial instances into one of the three adult age groups from AG4 to AG6.

4.2.1 Preprocessing

Facial images captured by camera are usually out of phase with each other due to variation in background, head pose, contrast of the images. These deficienciescannot

be solved automatically by most of the programs. Preprocessing stage produces standard facial images prior to feature extraction. The preprocessing stage, illustrated in Fig. 9,is applied to all imagesin the training and test samples. First, color facial images are converted into grayscale. Then, it is necessary to align the eyes for all images and crop the facial images afterwards. For this purpose, the original image is cropped by first aligning the left and right eye centers on the same line and then after cropping the image, the output image is resized to a fixed size. These steps are done manually to remove the requirement of mug-shot conditions.

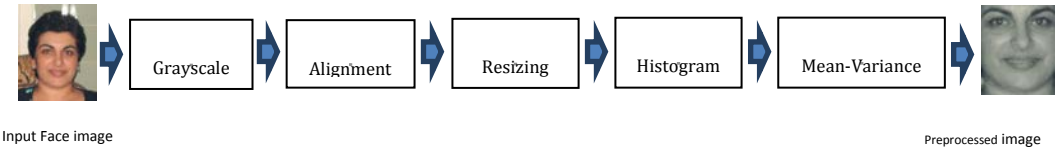


Figure 9. Preprocessing steps

Although the literature review shows that there are a number of sophisticated automatic face detection and localization methods, they are not completely reliable [58].Considering the significance of face localization in the proposed approach, we apply it manually in order to avoid any possible error. Afterwards, resizing is performed to unify size of all images. Finally, image enhancement step is applied. Histogram Equalization (HE) andMean-VarianceNormalization technique (MVN)areused to extracted face region to enhance the robustness of recognition features and improve the classification performance.

4.2.2 Feature Extraction Method

A key issue in designing a successful facial age classification is to find efficient feature extraction methods concerning the age periods. There are many factors affecting the age progression of a person during life [41, 57]. The facial transformation is faster in early years in comparison to adulthood. During childhood,

facial shapes undergo significant variations and the feature extraction focuses on changes on head size and ratios among facial features.

With the termination of growth at mid years of age, facial shapes and the distance between the main features changes with a limited amount. Furthermore, skin experiences an array of changes in appearance and texture in the form of fine lines and creases.

Accordingly, the separation of input images into different age groups with similar characteristics helps the extraction and development of more accurate age classification system. In the following sections, considering the numerous of changes that happen during the aging ofhuman we present two different feature extraction approaches for separating young and adult ages.

4.2.2.1 Geometric Feature Extraction

Geometric feature extraction uses ratio distancescalculated the main landmarks on face. In order to calculate the ratios on human face, key facial landmarks are necessary to calculate. Farkas [57] defined 57 landmarks and provide the overview of face anthropometry.

The extracted 17 landmarks and 10 distances are used in the formulation of the 6 proposed proportions are shown in Fig. 10.

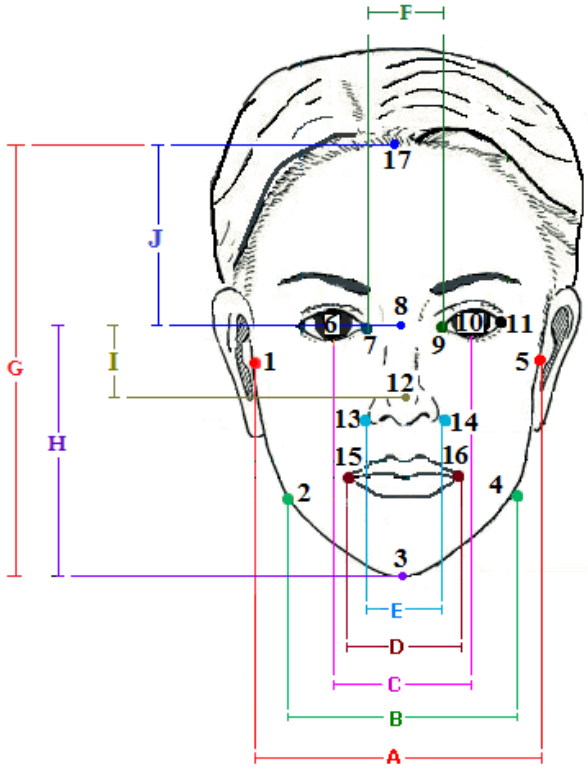


Figure 10. Seventeen landmarks and ten facial measurements used in the proposed method.

Facial landmarks and the corresponding abbreviations used in this study are shown in Table 2. The proposed method calculates the Euclidean distances between those landmarks after their formulation.

Table 3 demonstrates 10 facial distances shown in Fig.10 to formulate ratios. In Table 3, D(A,B) is assumed as Euclidean distance between point “A” and point “B” where “A” and “B” are arbitrarily facial feature points in two-dimensional space with coordinates (x_1, y_1), (x_2, y_2) respectively.

Table 2. Facial landmarks and their abbreviations

#	*Description*	*Abbr.*
1	Most sideward point of the cheek bone(left)	P_L
2	Most sideward point at the angle of the lower jaw (left)	M_L
3	Lowest point of chinchin	C
4	Most sideward point at the angle of the lower jaw (right)	M_R
5	Most sideward point of the cheek bone(right)	P_R
6	Middle point of left eye	E_L
7	Medial hinge of the eyelid (left)	E_{ML}
8	Mid-point between two eyes	E_M
9	Medial hinge of the eyelid (right)	E_{MR}
10	Middle point of right eye	E_R
11	Sideward hinge of the eyelid (right)	E_{LR}
12	Tip point of nose	N
13	Most sideward point on the wing of the nose (left)	N_L
14	Most sideward point on the wing of the nose (right)	N_R
15	Most sideward point of the lips (left)	L_L
16	Most sideward point of the lips (right)	L_R
17	Top point of the head	T

Table 3. Distances between landmarks shown in Fig.10.

Label	***Distance***	***Label***	***Distance***
A	$D(P_L, P_R)$	**F**	$D(E_{ML},E_{MR})$
B	$D(M_L, M_R)$	**G**	D(C ,T)
C	$D(E_L, E_R)$	**H**	$D(C, E_{LR})$
D	$D(L_L, L_R)$	**I**	$D(E_M, N)$
E	$D(N_L, N_R)$	**J**	$D(E_M, T)$

Finally, the six ratios in the proposed method that are based on face anthropometry are as follows:

$$R1 = \frac{D(P_L,P_R)}{D(C,T)} \quad , \quad R2 = \frac{D(E_M,T)}{D(C,T)}, \; R3 = \frac{D(E_{ML},E_{MR})}{D(E_M,N)}$$

$$R4 = \frac{D(L_L,L_R)}{D(M_L,M_R)}, R5 = \frac{D(C,E_{LR})}{D(C,T)} \quad , \quad R6 = \frac{D(E_L,E_R)}{D(E_M,N)}$$

4.2.2.2 Local Feature Extraction with LBP

A key issue in adult facial age classification is finding efficient descriptors for representing aged people. LBP operator is one of the best performing texture operators that partitions the facial image into equal-width non-overlapped regions. In our system the LBP texture descriptors are calculated for each partitionindividually. The descriptors are then concatenated to form a global description of the face. In order to build local descriptions, LBP extract the features from the image by assigning a label to every pixel of an image and thresholding pixels in eight neighborhood of each pixel with the middle value as explained in detail in Section 3.2.1.

The experiments are performed for age classification of adult facial images into more specific age intervals usingLBP technique. We used (8, 2) pixel neighborhood which means 8 sampling points in the circle of radius 2 and 8x8 subpatterns for age classification within LBP method because it gave the best results comparing to other ones.

4.2.3 Age Classification Method

The classification phase is the key decision making stage of face classification system and uses the features extracted in the previous stages to identify the correct age groups. After we extract geometric features, a Support Vector classifier (SVC) is applied to classify the first four categories of young age groups: (0-2), (3-7), (8-19), and (20+) years old. In this work, we use the SVC classifier for determining the age of facial images. The support vector classifiers are based on support vector machines. In support vector machine based classification, each data point in the dataset is represented by a k-dimensional vector. Assuming that each data point belongs to

only one of two classes, the support vector classifier separate the dataset with a k − 1 dimensional hyperplane with maximum separation between the two classes. In other words, the nearest distance between a data point in one hyperplane and a data point in the other hyperplane is maximized [69].

We used a nearest neighbor classifier and Chi Square distance metrics for classifying adult face images into one of the three adult age groups from AG4 to AG6. AG4 contains faces of age (20-40), AG5 is (40-60), and finally (60+) are placed in AG6.

Chi Square distance between two different histograms X and Y is

$$\chi^2(\mathrm{X},\mathrm{Y}) = \sum_{\mathrm{i}=1}^{\mathrm{n}} \frac{(\mathrm{X}_\mathrm{i} - Y_\mathrm{i})^2}{X_\mathrm{i} + Y_\mathrm{i}}$$

Where n is the number of elements in the histogram. Chi square is an effective measurement of likelihood between a pair of histograms; hence it is suitable for nearest neighbor classification.

4.3 Experimental set up and results

Experiments are carried out on two publically available databases to prove the effectiveness and accuracy of the presented feature extraction. Currently there are many facial image databases [11,22, and 29]. These databases, generally, suffer from at least two points. The first is absence of age relateddataof the samples in the database, whereas second problem is lack of variety of ages among individual images.

Among the existing databases, FGNET [22] has different facial images of an individual in different years hence, suitable for age related studies. The FGNET aging database is publicly available. The age range is from 0 to 69 years with

chronological aging images available for each individual. Each image was annotated with 68 landmark points which are essential for locating facial feature and calculation of biometric ratios.

IFDB [21] is another large face database collected recently, containing over 3600 digital color face images of subjects aged from 1 to 85 years. The images, with variations of pose, expression, have 640 x 480 resolutions and focused on age information.

In all experiments performed in this study, a subset of 1104 images from the publicly available FGNET and IFDB database is used. The dataset contains 679 frontal images from FGNET and 425 images from IFDB database. We used only frontal images without glasses or beard.

The original dataset is split into training and test sets illustrated in Table 4. The training set is used for generating the classification model and the test set is used to test the classification performance of the classifier. To show the effectiveness and precision of the proposed feature extraction a support vector classifier is applied. For this purpose, we used PRTools which is a Matlab based toolbox for pattern recognition[60].

Table 4. Numbers of subjects in each age group

Age Group	Sample size	Train	Test
(0-2)	83	35	48
(3-7)	252	147	105
(8-19)	425	280	145
(0-19)	760	462	298
(20-40)	211	108	103
(41-60)	84	34	50
(60+)	49	20	29
(20+)	344	162	182
total	1104	624	480

As a classification problem, the best way to reflect the performance in classification problem is using the classification success rate. Success rate is expressed as a percentage using the following formula:

Si = Ti / Ni x 100 % where Si is the success rate in an age group I, Ti is total number of test images that correctly labeled in age group i, and Ni is the sample size in age group i.

For showing the effect of the proposed method in age estimation, the geometric features proposed in this work were compared with the ones proposed by the authors in [8, 54, and 55] and are listed in Table 5. The table shows the precision of success rate (%) in separating the subjects into young age groups named above. It shows that the proposed set of 6 ratios outperforms the other methods in separating the subjects in these four age groups.

Table 5. Comparing proposed method with the work of other researchers, separating three age groups

Methods	Dehshibi&Bastanfard [55]	Ramanathan [54]	Kwon & Lobo [8]	Proposed Method
(0-2)	95.91	93.87	93.87	97.91
(3-7)	93.33	94.28	91.42	94.56
(8-19)	84.82	86.89	83.44	88.54
(20+)	95.45	94.88	93.75	96.59

Comparison between the results of proposed geometric ratio analysis for separating young facial images from adults with the work of other researchers [8, 54, and 55] is illustrated in Fig.11. It can be clearly seen from the graphs in Fig. 11 that the method proposed in this section outperforms other state-of-the-art methods.

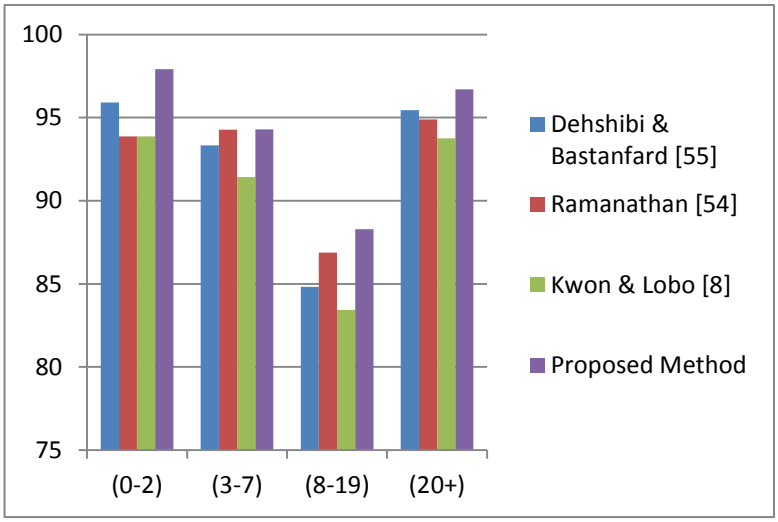


Figure 11. Comparison between the accuracy rate of proposed method with the work of other researchers

The second set of experiments is performed for age classification of adult facial images into more specific age intervals with Local Binary Patterns. We used (8, 2) neighborhood and 8x8 subpatterns for age classification within LBP method because it gave the best results comparing to other ones. Also, PCA and subspace LDA methods are used on the image samples using different number of non-zero

eigenvectors. The maximum number of eigenvectors which is the number of image samples minus one for PCA and the number of classes minus one for subspace LDA are used [74]. Thenumber of eigenvectors has been chosen experimentally in which the best results are obtained and demonstrated in the experiments.

The classification accuracies using LBP, PCA and subspaceLDA are compared in Table 6. In this table, the overall age classification performances for different age groups of facial images using three different methods are given.The experimental results demonstrate that LBP gives moreaccurate results as compared to holistic matching methods such as PCA and subspaceLDA on facial images of FGNET and IFDB datasets in categorizing adult facial images.

Table 6. Comparing LBP method with PCA, subspace LDA approaches for “20 above” adult age classification

Methods	LBP	Subspace LDA	PCA
(20-40)	90.29	89.32	84.47
(41-60)	84	82	74
(60+)	86.21	82.76	79.31

Table 7 demonstrates the rate of success achieved by the proposed algorithm in classifying the facial images in the different age groups. This table clearly displays that the proposed system, as a whole, yields very high success rate in classifying the images in their respective age groups, correctly. It can be seen that the minimum success rate is 84%, whereas the maximum success rate is as high as 97.92%. Our proposed system works exceptionally well in identifying the age of facial images of children aged in 0-2. It should also be noted that the overall system success rate is 91.09%.

Table 7. Success rate of the complete proposed algorithm

Age Groups / Percentage	*(0-2)*	*(3-7)*	*(8-19)*	*(20-40)*	*(41-60)*	*(60+)*
Proposed Method	97.92	94.29	88.28	90.29	84	86.21

Table 8 details the overall system accuracy rate using different methods. The first row shows the overall performance of the proposed system for classifying the input facial images into six different age groups, using Geometric ratio features and LBP operator. The second row is the overall performance of the system using Geometric ratio features and subspace LDA method. Finally the third row gives the results using Geometric ratio features and PCA method.

Table 8. Comparison of age classification results on FGNET and IFDB databases using Geometric ratios (R) with different methods

Methods	Accuracy rate (%)
R+LBP	91.09
R+subspace LDA	90.18
R+PCA	87.85

Chapter 5

AGE CLASSIFITION WITH OPTIMAL GEOMETRIC RATIOS AND WRINKLE ANALYSIS

5.1 Introduction

This chapter describes a novel age group classification method [78] using frontal face images taken from two large and publicly available databases namely IFDB [21] and FGNET[22]. Geometric ratios and wrinkles as featuresin the face images are used for age group classification. Geometric ratios are calculated based on the distance and the size of main facial features to distinguish young faces from adults. Using ratio indices instead of distances between the facial landmarkswill help in eliminating the image scale dependency. Particle swarm optimization is a technique that is used to find the optimal set of geometric ratios by creating particles that search a space of various ratios of the input facial image while attempting to produce the best accuracy rate on the data.The adult face images go through wrinkle analysis to be further classified into precise age groups. The age differentiation capability of the individual features (geometric and wrinkle feature) and various combinations of the features are evaluated using SVC [23]. The experimental results demonstrate that the developed feature selection method is able to improve the classification performance.Our results indicate that precise feature selection can help classifiers to categorize the images with high accuracy with a classification rate of above 90 percent.

5.2 Age Group Classification Algorithm

Age group classification algorithm is mainly based on extracting features that help classify the age group of an input image. The subsequent subsection describes feature extraction, leading to explanation of feature classification method for age group classification. Among the different design issues involved in building an age group classification system, the most consequential one is the selection of the type and set of features. For this purpose, some ratios based on face geometry using Particle Swarm Optimization and wrinkle features are extracted. The proposed method is detailed in section 5.3 using a block diagram of the complete algorithm.

5.2.1 Geometric Feature Extraction using PSO

The selection of a reliable and representative set of features is crucial for any classification system. Geometric feature-based approaches use anthropometric distances extracted from different facial regions in estimating an individual's age. Selection of ratios based on face anthropometryis applied as the means of achieving this goal. Different authors proposed different set of ratios in the literature [8, 9, 28, 55, and 56]. In order to choose an optimized subset of features based on a certain objective function from all feature sets by eliminating the redundant and inappropriate information feature selection approaches are used [69]. In this study, PSO algorithm is used, which was first presented in 1995 by authors in [70].PSO has been successfully applied to many applications such as biometric system[69], autonomous mobile robots and biomedical engineering [71]. In this work, we use the Particle Swarm Optimization technique to find the optimized subset of geometric features because of its good performance in several applications.

The solutions that are randomly initialized in Particle Swarm Optimization are referred to as "particles". A fitness function is then used to evaluate them. The

representation of a specific particle is as Xi = (xi1, xi2, . . . ,xin). PSO has an advantage that it can memorize previous positions. The best positions, namely pbest and gbestfrom the previous iterations are stored and used to update the particles. The best fitness value is considered to find pbestvalue with a size of population P as Pi = (pi1, pi2, . . . , pin). The best particle in the overall population is stored using its index value named gbest.

The velocity for ith particle is represented as $Vi = (v_{i1}, v_{i2}, \ldots, v_{in})$. The particles are updated according to the following equations [23]:

$$v_i = wv_i + c1 \times V + c2 \times Z,$$

$$x_i = \begin{cases} 1, & if \dfrac{1}{1+e^{-v_i}} > rand_3() \\ 0, & otherwise \end{cases}$$

$$V = rand_1() \times (p_i - x_i)$$

$$Z = rand_2() \times (p_{gbest} - x_i)$$

$$V \sim V[0,1]$$

where *w* is the inertia weight, $c_1$and $c_2$are acceleration constants and $rand_1()$, $rand_2()$, $rand_3()$ are independent random numbers.$c_1$and $c_2$are cognitive parameter and social parameter respectively.

In the presented study, we set the value of *w*to 1,and constant 2 is selected for both$c_1$and $c_2$constants. Regarding the particle'sprior velocity and comparing the distances between the current position and the group's best one, the velocityis updated to a new value.Measuring how well the model is performing is byconsidering the fitness function.

Moreover, the fitness function is the recognition rate and the selection of features is based on a bit string of length *M*, where *M* is the number of geometric ratios applied on facial images. In other words, every bit here represents one geometric ratio; value '1' means that the geometric ratio is selected and '0' means that it is not selected. In the presented study, the size of the population and iterations are assigned to 10 and 50 respectively. The PSO method is implemented as it is explained in [69].

Facial landmarks location (FLL), extracts the biometric ratios on human face. Facial anthropometry is to study scientifically the measurements and proportions of the individual faces. Face anthropometric models deals with a quantitativedescription of the cranio-facial complex with the help ofthe measurements taken over the main landmarks on facial images across age [57].

Farkas [57] 57 landmarks on human faces and provides a comprehensive overview of face anthropometry.As shown in Fig. 12, 21 landmarks and 16 distances are used to formulate ratios in our method. These ratios form the geometric features consumed by the classifiers for appropriate decision making of facial age. Facial landmarks and the corresponding abbreviations are shown in Table 9. The Euclidean distances between those landmarks are then calculated.

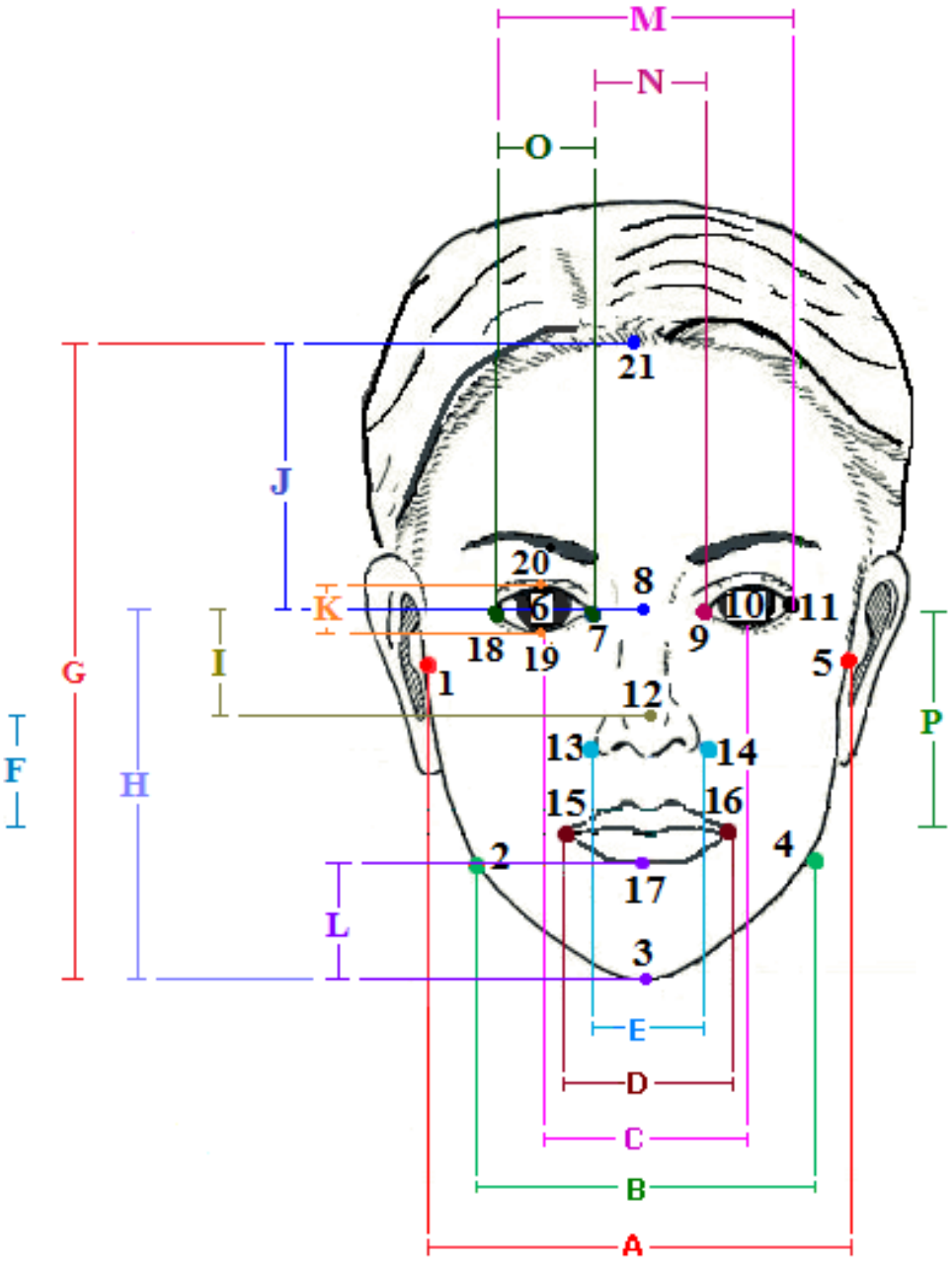


Figure 12. Facial landmarks and facial measurements.

Table 9. Facial landmarks and the abbreviations

#	*Description*	*Abbr.*
1	Most sideward point of the cheek bone(left)	P_L
2	Most sideward point at the angle of the lower jaw (left)	M_L
3	Lowest point of chinchin	C
4	Most sideward point at the angle of the lower jaw (right)	M_R
5	Most sideward point of the cheek bone(right)	P_R
6	Middle point of left eye	E_L
7	Medial hinge of the eyelid (left)	E_{ML}
8	Mid-point between two eyes	E_M
9	Medial hinge of the eyelid (right)	E_{MR}
10	Middle point of right eye	E_R
11	Sideward hinge of the eyelid (right)	E_{LR}
12	Tip point of nose	N
13	Most sideward point on the wing of the nose (left)	N_L
14	Most sideward point on the wing of the nose (right)	N_R
15	Most sideward point of the lips (left)	L_L
16	Most sideward point of the lips (right)	L_R
17	Middle point of the lower margin of the lower lip	L_{LM}
18	Sideward hinge of the eyelid (left)	E_{LL}
19	Lowest point of the margin of the lower eyelid (left)	E_{LML}
20	Highest point of the eyelid (left)	E_{HL}
21	Top point of the head	T
22	Midline point where the lips	L

Table 10 demonstrates ten facial distances shown in Fig. 12 to formulate ratios. The proposed two sets of ratios based on face anthropometry using PSO for each stage of the algorithm are shown below where *D(A, B)* is the Euclidean distance between two given points "*A*" and "*B*".

Table 10. Distances between landmarks shown in Fig.12.

Lable	Distance	Abbreviation
A	Distance between zygion	$D(P_L, P_R)$
B	Distance between sides of face	$D(M_L, M_R)$
C	Distance between left and right pupils	$D(E_L, E_R)$
D	Lip length	$D(L_L, L_R)$
E	Distance between sides of nose	$D(N_L, N_R)$
F	Distance between tip of nose and middle of lip	D(N, L)
G	Distance between Chin and Top of head	D(C ,T)
H	Distance between middle of eyes and chin	$D(E_M, C)$
I	Distance between middle of eyes and tip of nose	$D(E_M, N)$
J	Distance between middle of eyes and top of head	$D(E_M, T)$
K	Left eye height	$D(E_{HL},E_{LML})$
L	Distance between Chin and lower lip	$D(C, L_{LM})$
M	Distance between outer corners of eyes	$D(E_{LL}, E_{LR})$
N	Distance between inner corners of eye	$D(E_{ML},E_{MR})$
O	Left eye length	$D(E_{LL}, E_{ML})$
P	Distance between middle of eyes and lip	$D(E_M, L)$

Biometric ratios set $R1_x$:

$R1_1 = D(E_L, E_R)/D(E_M,L)$, $R1_2 = D(L_L, L_R)/D(P_L,P_R)$, $R1_3 = D(C, L_{LM})/D(M_L,M_R)$,

$R1_4 = D(E_{ML}, E_{MR})/D(E_{LL}, E_{LR})$, $R1_5 = D(E_M, N)/ D(E_M, L)$, $R1_6 = D(E_{HL}, E_{LML})/D(E_{LL}, E_{ML})$,

$R1_7 = D(N_L, N_R)/D(E_M, N)$ $R1_8 = D(C, T)/D(E_M, C)$ $R1_9 = D(E_{ML}, E_{MR})/D(E_M,N)$

Biometric ratios set $R2_x$:

$R2_1 = D(E_L, E_R)/D(E_M,N)$, $R2_2 = D(E_M, N)/ D(N, L)$, $R2_3 = D(P_L, P_R)/D(C,T)$,

$R2_4 = D(E_M, T)/D(C,T)$,

$R2_5 = D(E_L, E_R)/ D(M_L,M_R)$, $R2_6 = D(L_L, L_R)/ D(M_L,M_R)$ $R2_7 = D(E_M,C)/D(P_L,P_R)$

The key aspect of the method proposed in this chapter is that we use different groups

of ratios for separating different age groups. This increases the percentage of successful age classification of facial image. Experimental and theoretical evidence shows the effectiveness and accuracy of the proposed feature selection.

In the proposed method, $R1_1$ to $R1_9$is used as 'biometric ratios set1' that classifies images in groups of under or above age 20. The ratios $R2_1$ to $R2_7$help choose 'biometric ratios set2' that classifies images in groups of 'AG1', 'AG2' and 'AG3'. AG1 contains images of faces aged 0-2, AG2 lists faces classified from age 3-7, AG3 is a set of images that pertains to faces categorized to be 8-19 years old.

The facial transformation is very much different in young ages in comparison to the later years. During childhood, the development of cranium's shape is very rapid and the feature extraction focuses on changes on head size and ratios among facial features. In young adults, [68] there is little change in cranium's size, furthermore, the upper and lower part of the face grows independently from one another.

In adults, sides of the face changea little and the distance between the main features changes vaguely. As human beings grow and become old, skin artifacts such as lines, wrinkles and eyelid bags appear on their face [72, 73]. Clearly, distances between some features in face can help distinguish young instants from others.

5.2.2 Wrinkle Feature Analysis

Following geometric ratio extraction, extracting a set of reliable wrinkle features is crucial for adult faces prior to classification task. It is important to extract informative sets of wrinkle features from which classification might be performed more efficiently and accurately. As people age,obvious wrinkles appear around the eyes, cheeks and forehead. These wrinkle areas have different characteristics that can

give clues about the age of a person.

The adult color image is first prepared in a preprocessing stage to provide enhanced images before applying feature extraction. During preprocessing, once face image is converted into grayscale, the wrinkle area is extracted and cropped. In order to enhance the further detection of the wrinkles on the grayscale face, noise reduction is applied to the cropped grayscale image.

Noise reduction reduces the effects of noise factors that can be mistakenly identified as an edge or wrinkle. Once the noise is reduced, Histogram Equalization is applied to increase the precision of the further wrinkle detection on the faces. The detection of wrinkle regions is done based on the fact that adults usually have prominent wrinkles on certain parts of the face. As people age, normally fine wrinkles can be observed clearly on some parts of the face such as the forehead, corners of the eyes, under cheeks, and above mouth as shown in Fig. 13.

According to the human facesymmetry, wrinklefeatures extracted from one side of the facial imageis sufficient in order to consider in the wrinkle analysis which includes the forehead, left eye tail and canthus as illustrated in Fig. 13a.

a) b) c) d) e) f) g)

Figure 13. a) original image; wrinkle densities of b) Forehead c) Left eye corner d) Left canthus; edge detection after applying filtering on e) Forehead f) Left eye corner g) Left Canthus

Forehead area includes horizontal lines as shown inFig. 13b. It is placedon the upper third part of the face between the eyebrows and the normal hairline. Forehead area is determined as a rectangleusing the distance between left and right eyes as:

$$Width = (4/3) \times D(E_L, E_R) \quad , \quad Height = (1/3) \times D(E_L, E_R)$$

The cropped forehead area shown in Fig. 13a is exactly $(0.75) \times D(E_L, E_R)$ upperfrom left and right pupils and located at the center region of the forehead.

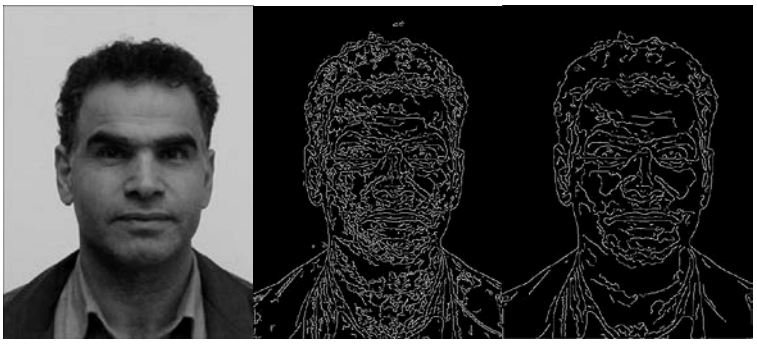
Eye tails corners have obviouscreases that are illustrated in Fig. 13a. Thelocation of this region is at the lateral sideward of the left eye under the eyebrow. The distance from left upper side of this rectangleto the left pupil is about $(5/12) \times D(E_L, E_R)$. The eye tail region is a rectanglethat is defined as follows:

$$Width = (1/6) \times D(E_L, E_R) \ , \quad Height = (1/4) \times D(E_L, E_R)$$

Left eye corner is shown in Fig. 13d and the wrinkles extracted from left eye tail after applying filtering is demonstrated in Fig. 13g. Canthus region is located on either side of the face below the eyes. This area is located $(1/8) \times D(E_L, E_R)$beneath the left and right pupils and above the mouth, as shown in the Fig. 14a. Left canthus region is a square that is defined as follows:

$$Width = (1/2) \times D(E_L, E_R) \ , \quad Height = (4/5) \times D(E_L, E_R)$$

Left canthus is shown in Fig. 13c and the wrinkles extracted from left eye tail after applying filteringon the forehead, left eye corner and left canthus are demonstrated in Fig. 13e, f, g respectively. On the other hand, Fig. 14a shows an enhanced image while Fig. 14b and c illustrate the wrinkles with and without applying filtering respectively.



a) b) c)

Figure 14. a) Enhanced face; b) canny edge detection; c) edge detection after applying filtering

The wrinkles on faces become more prominent when human beings become older. After preprocessing the facial images Wrinkle Feature Analysis can be applied on facial images. In this stage, Canny Edge Detection is used on the extracted areas to determine the density of wrinkles.

Once the edge detection method is applied on the images, the grayscale image is converted into its binary equivalent to calculate the wrinkles density. These densities are applied to aged people to approximate their age range. Since the gray levels variances in an image, wrinkles have clearvariation in intensity and even some of them can be observed as fine lines. As a consequence, the pixel belonging to an edge is considered as a wrinkle pixel. If the function distinguishes a wrinkle pixel in the given image it returns 1, and in other cases returns 0.

Wrinkle density in region ‘‘*R*’’ can be defined as follows:

$$Wrinkle\ density = (Edge_R / TotalPixel_R) x100.$$

In order to show the difference in the skin rhytids of a child and old person, the wrinkle features related to two sample images are demonstrated in Fig. 15.Enhanced image of a child and an adult is shown in Fig. 15a and c respectively.Moreover, the wrinkle densities extracted from corresponding sample images are demonstrated in Fig. 15b and d respectively. On the other hand, Fig. 15e and f illustrate the wrinkles extracted from their forehead region separately.

Wrinkle areas like other facial feature perimeters are very relevant to the anthropometric properties of the face, especially distances between the eyes. For instance, it is possible that some parts of the hair take place in wrinkle areas where the wrinkle density increases. Furthermore, abnormalities, severe skin diseases, scar tissues, and subjective facial expressions can be considered as a mistake for wrinkles by the system. In our facial database, in order to reduce the margin of error, the images with facial hair on the three selected regions, particularly in forehead region, are not included.

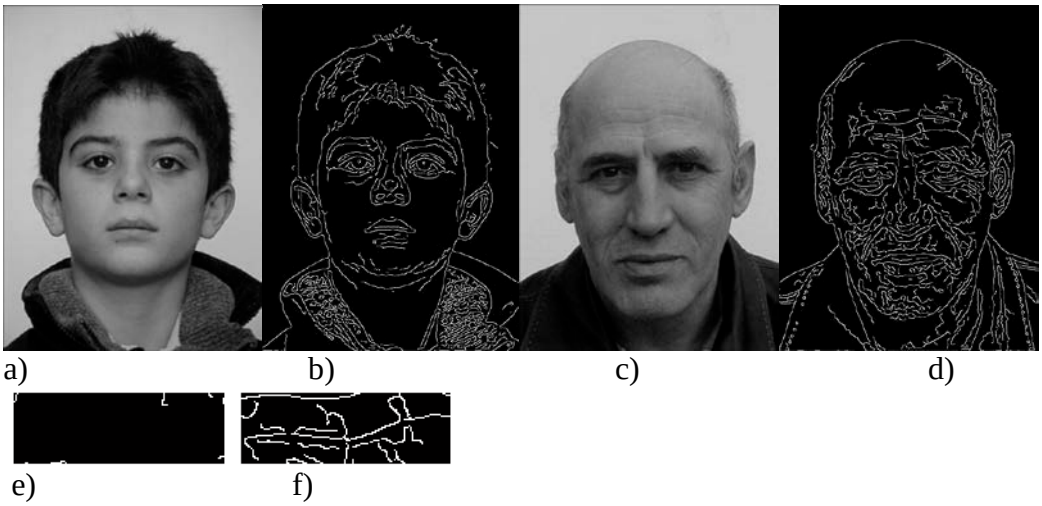


Figure 15. a) Enhanced image of a young face b)wrinkle density of a young face; c) enhanced image of the old face d)wrinkle density of the old face; e) young forehead; and f) old forehead.

5.3 Age Group Classifications

The classification stage is the main decision making stage of face classification system and uses the features extracted in the previous stages to identify the age groups. After extracting the features, we designed an algorithm to classify the face images into appropriate age groups.

The classifiers employ different features extracted in 'biometric ratio set1' to distinguish whether a facial image is "20 above" or not. If it is not, the 'biometric ratio set2' is applied to classifiers which, in turn, classify the first three categories of age groups: AG1, AG2, and AG3. AG1 contains images of faces aged 0-2, AG2 lists faces classified from age 3-7, AG3 is a set of images that pertains to faces categorized to be 8-19 years old. The adult facial images, needs to go through Wrinkle Feature Analysis (WFA) to be further classified into appropriate ages. For this purpose, the wrinkle features extracted from "20 above" facial images are applied to SVM classifier to classify the images into one of the four adults Age Groups from AG4 to AG7, where AG4 contains faces of age 20-29, AG5 is 30-45, AG6 is 46-60 and finally 60+ are placed in AG7.

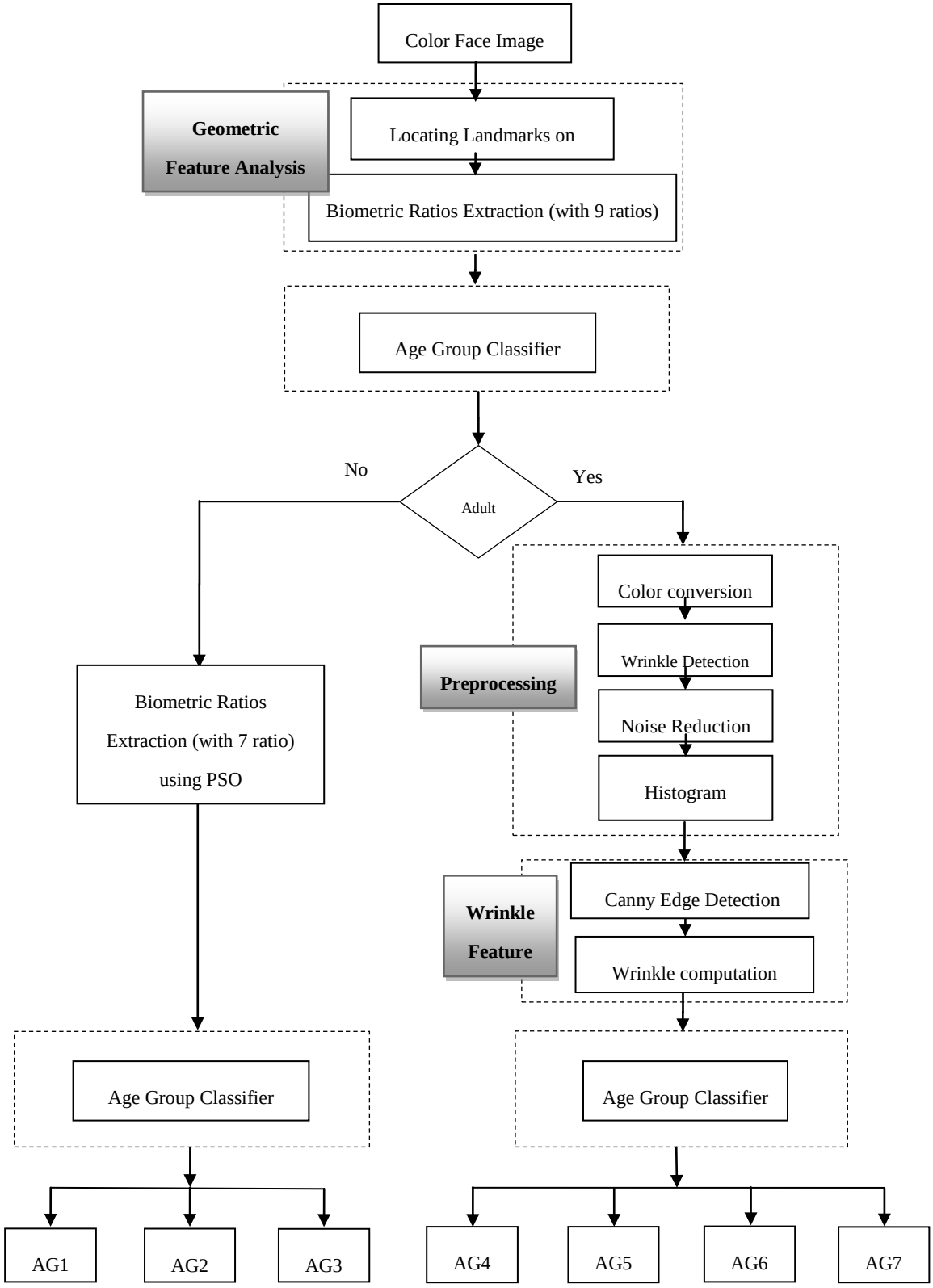


Figure 16. Age group determination algorithm block diagram.

5.4 Proposed Age Group Classification Method

The proposed Age Group Classification (AGC) method [78] illustrated in Fig. 16classifies the facial images into one of the well-ordered age intervals. Initially geometric feature information of the color facial image is extracted using Geometric Feature Analysis (GFA). GFA locates certain facial landmarks, depicted in Table 9, and using these landmarks calculates two sets of biometric ratios namely 'biometric ratios set1' and 'biometric ratios set2'. In the next stage of the algorithm, the first set of ratios extracted using GFA are applied to SVC classifier.

'Biometric ratios set1' is passed to classification phase. During classification, the features generated in 'Biometric ratio set1' are applied to the SVC. The results gathered from the classifier are then used to distill the images into two categories: images below 20 years of age and those above 20 years, labeling them as 'young face' and 'adult face' respectively. In the subsequent steps, these two categories resulting from this classification process are used to predict specific age groups. The 'biometric ratio set2' resulting from biometric ratio extraction is applied to classifier, which, predicts the first three categories of age groups: AG1, AG2, and AG3.

The color image, classified as 'adult face', needs to go through Wrinkle Feature Analysis (WFA) so that images are classified into appropriate ages. For this purpose, the image is first processed and prepared, for the usage of WFA in a preprocessing stage. The facial image is ready after preprocessing step to be analyzed using WFA. The first task performed by WFA is that of Canny Edge Detection where wrinkles are located and marked on the image. The system improves the wrinkle detection with preprocessing the image and then with Canny Edge Detection, the system asserts that there are minimal missed wrinkles. Once the wrinkles are determined on the images, Wrinkle Computation performs the task of converting the grayscale image into its binary equivalent. This wrinkle information is applied to classifiers

that classify the images into appropriate Age Groups from AG4 to AG7.

5.5 Experimental Results and Evaluation

The experiments demonstrated in this section were conducted to evaluate the performances of the proposed facial age classification approaches presented in this chapter. These experiments aim to evaluate the performance of the proposed classification approach.

Two sets of experiments were conducted. Sub-section 5.5.1 provides details concerning the parameter settings and database used for two set of experiments.The first set of experiments are performed to show the performance of the proposed algorithm and to compare the results with the work of other researchesand these results are presented in subsection5.5.2.

The second set of experiments, presented in Subsection 5.5.3, is conducted on the same database to evaluate the classifiers performance using hold-out method for 10 times.

5.5.1 Experimental set up

In order to show the effectiveness and accuracy of the proposed method; experiments are carried out on two publicly available databases. Currently there are many facial image databases in the literature [11, 21, 22 and 29].

These databases, generally, suffer from at least two points. The first is absence of age related data of the samples in the database, whereas second problem is lack of variety of ages among individual images. Among the existing databases, FGNET [22] has different facial images of an individual in different years hence, suitable for age related studies. The FGNET aging database is publicly available. The age range is

from 0 to 69 years with chronological aging images available for each individual. Each image was annotated with 68 landmark points which are essential for locating facial feature and calculation of biometric ratios.

IFDB [21], focused on age information is the other database used in experiments presented in this work. IFDB contains color facial imagery of a large number of Iranian subjects. IFDB contains digital images of subjects aged 1 to 85 years. The images, with variations of pose, expression, without glasses, have 640 x 480 resolutions.

In all experiments conducted in this study, a subset of 1104 images from the publicly available FGNET and IFDB databases are used. The dataset contains 425 frontal images from FGNET and 679 images from IFDB database. We used only frontal images of subjects, without glasses or beard. All samples are taken from different ages. Sample facial images from seven different age intervals are shown in Fig. 17. For experimental purposes the original dataset is split into two training and test sets. In the training phase, 429 images and in the test phase 675 images are used. The train and test samples are divided into two further subsets. Here, 217 images are used to test the performance of the PSO, whereas 458 images are used for testing the performance of the complete system. Table l1tabulates the number of subjects in each age group.

Figure 17. Facial images from seven different age intervals (the first row images are taken from IFDB, and the second row belongs to FGNET dataset)

SVC isused to display the effectiveness and precision of the proposed method. The simulation has been performed on Intel Core 2 Duo CPU 2.0 Ghz with 8 GB Ram. The columns in Table 11, starting from the left-hand side, should be interpreted asfollows:

(i) “Age group” indicates the age group sample image belongs, (ii) “Sample size” shows the total number of samples available in the corresponding age range , (iii) “Train A” is the number of samples used in thetraining stage for PSO, (iv) “Test A” is the number of samples in each age range totest the performance of PSO and(v)”Train B” is the number of samples used in the training the whole system and (vi)“Test B” is the number of samples in each age range to test the performance of the complete system.

Table 11. Number of samples in each age groups used in experiments

Age group	Sample size	Train A	Test A	Train B	Test B
(0-2)	83	10	11	35	48
(3-7)	252	40	50	147	105
(8-19)	425	80	95	280	145
(20-29)	128	24	20	77	51
(30-45)	112	18	22	60	52
(46-60)	55	8	10	28	27
(60+)	49	5	9	20	29
total	1104	185	217	647	457

5.5.1.1 Simulation Parameters for Feature Selection using PSO

PSO is an optimization algorithm which is used as a feature selection method in order to find the optimized set of geometric features.In the presented study, we set the value of *w*to 1,and constant 2 is selected for both$c_1$and $c_2$constants. Regarding the particle'sprior velocity and comparing the distances between the current position and the group's best one, the velocityis updated to a new value.Measuring how well the model is performing is byconsidering the fitness function.

Moreover, the fitness function is the recognition rate and the selection of features is based on a bit string of length *M*, where *M* is the number of geometric ratios applied on facial images. In other words, every bit here represents one geometric ratio; value '1' means that the geometric ratio is selected and '0' means that it is not selected. In the presented study, the size of the population and iterations are assigned to 10 and 50 respectively [75]. The PSO method is implemented as it is explained in [70].

5.5.1.2 SVC parameter setting

In this work, we use the SVC classifier for determining the age of facial images. The support vector classifiers are based on support vector machines.Support vectors machinesare a set of supervised learning methods used for classification problem.

SVC is a class effective in high dimensional spaces and capable of performing multi-class classification on a dataset.SVC implements the one-against-one approach [69] for multi- class classification.

For this purpose we used a Matlab based toolbox for pattern recognition (PRTools) [60]. This toolbox offers many routines which is useful for researchers studying or applying pattern recognition systems. It covers the main steps: representation, classification and evaluation.

The power of PRTools is based on the carefully designed operations between different variables of specific programming classes. We use the SVC routine which optimizes a support vector classifier for the dataset D by quadratic programming.
The parameters of the classifier are as follows. The KERNEL is radial basis function (RBF). The regularization parameter C is assumed to be NaN. In this case the regularisation parameter is optimised by another routine REGOPTC provided by PRTools. This routine is used inside classifiers and mappings to optimize a regularisation or complexity parameter [33].

5.5.2 Experiments

Geometric features proposed in this work were compared with the ones proposed by the authors in [8, 54, and 55] and the results are listed in Table 12.The table shows the precision of estimation rate (%) in separating the subjects into two age groups i.e. age groups of (0-19) and (20+).

Table 12. Proposed method compared with the work of other researchers, separating young (0-19) from adults (20+)

Methods	(0-19) (20+)
Dehshibi&Bastanfard [55]	96.99
Ramanathan[54]	97.99
Kwon & Lobo[8]	91.30
Proposed Method	**98.99**

It shows that the proposed set of 9 ratios outperforms the other methods in separating the subjects into young and adults.Table 13 compares different classifiers experimented with the work of other researchers, including the proposed method. The table shows the precision of different geometric features estimation rate (%) in separating the subjects into 3 age groups i.e. age group of (0-2), (3-7) and (8-19). It shows that the proposed 7 ratios selected by PSO outperform the other methods.

Table 13. Proposed method compared with the work of other researchers, separating three age groups

Methods	*AG1(0-2)*	*AG2(3-7)*	*AG3(8-19)*
Dehshibi&Bastanfard [55]	95.91	93.33	84.82
Ramanathan[54]	93.87	94.28	86.89
Kwon & Lobo[8]	93.87	91.42	83.44
Proposed Method	**97.95**	**96.19**	**89.65**

Finally, Table 14 demonstrates the rate of success of the complete proposed algorithm achieved in classifying the facial images in the seven age groups.

Table 14. Test phase for separating all seven age groups using proposed method

Age Groups / Percentage	*AG1 (0-2)*	*AG2 (3-7)*	*AG3 (8-19)*	*AG4 (20-29)*	*AG5 (30-45)*	*AG6 (46-60)*	*AG7 (60+)*	*Overall success*
All ratios without using PSO	95	95.48	88.33	91.54	86.48	83.78	89.47	90.01
Selected ratios without using PSO [22]	96.66	96.12	89.16	92.95	87.83	86.48	92.10	91.61
Proposed Method using PSO	97.95	96.19	89.65	94.11	88.46	88.88	93.10	92.62

The second row shows the result of applying complete system by using all geometric ratios of the face without using PSO. In addition, the results using geometric features proposed by the authors in [56] are demonstrated in the third row of Table 14.

It shows that the proposed method using PSO outperforms the other methods in separating the subjects into seven age groups. This table clearly shows that the proposed system, as a whole, yields very high success rate in classifying the images in their respective age groups, correctly. It can be seen that the minimum success rate using PSO is 88.46%, whereas the maximum success rate is as high as 97.95%. Our proposed system works exceptionally well in identifying the age of facial images of children aged in 0-2. It should also be noted that the overall system success rate is 92.62%.

5.5.3 Experimental estimation of classification accuracy

In a classification problem, a model is usually given a dataset of known data on which training is run (training dataset), and a dataset of unknown data against which the model is tested (testing dataset) [23]. It is important that the test data is not used

in any way to create the classifier. One way to overcome this problem is to not use the entire data samples when training the system. Cross-validation, [76, 77] is a method for validating how well the results of a classification approach will generalize to independent input facial images. The purpose that we use this technique is to select a set of images to test the system in the training step to eliminate problems such as overfitting. It gives an understanding on how well the model will classify the testing facial images has not already seen.

The holdout method is the simplest kind of cross validation. In hold-out technique the original dataset (facial images) is split into complementary training and test sets randomly. The training set is used for generating the classification model and the test set is used to evaluate the classification performance of the classifier.

As such, we repeated the experiments using hold-out method. In order to obtain more reliable results for hold-out training technique, the process is repeated 10 times with different subsamples. In each round, certain partition of samples is randomly selected for training and the rest is used for validating the analysis on the other subset called the testing set. The error rates (classification accuracies) on the different rounds are averaged to yield an overall performance of the classifier. In addition standard deviation is used for comparing learning algorithms.

In this section, all the experiments reported in the previous section are repeated to demonstrate the performance of the algorithm in the classification of the images into different age groups using hold out approach.Geometric features proposed in this work were compared with the ones proposed by the authors in [8, 54, and 55]. The experiments are repeated 10 times using hold-out method to demonstrate the

performance of the algorithm. The classification accuracies on each round are averaged as shown in Table 15.

In each column, two values are recorded, each of which correspond to the average and standard deviation (in brackets) of the results generated using 10 round of hold-out technique. The highest value obtained for each evaluation metric is indicated in bold font. All the results are produced using IFDB and FGNET facial images illustrated in Table 11.

The precision of estimation rate (%) in separating the subjects into two age groups i.e. age groups of (0-19) and (20+) are shown in Table 15. Observation of the results obtained using the two class dataset presented in Table 15 indicates that the proposed set of 9 ratios outperforms the other methods in separating the subjects into young and adults. The results generated across the different rounds indicated that consistent results were produced, with standard deviations of less than 2% for accuracy in all methods presented.

Table 15. Average classification results obtained for proposed method compared with the work of other researchers, separating young (0-19) from adults (20+)

Methods	Accuracy
Dehshibi&Bastanfard [55]	95.85(1.02)
Ramanathan[54]	96.45(1.90)
Kwon & Lobo[8]	91.81(0.91)
Proposed Method	**98.03(0.81)**

Table 16 compares the proposed method with the work of other researchers. The table shows the precision of different geometric features estimation rate (%) in separating the subjects into 3 age groups i.e. age group of (0-2), (3-7) and (8-19). The

results presented in this table indicate that the proposed 7 ratios selected by PSO, produced better results than other methods. The presented results show that, all the accuracies for the proposed method are greater than 91% and with the best classification accuracy of 96.12% for the first age range.Inspection of the standard deviations demonstrated that similar accuracy was produced across the different runs, with a standard deviation of less than 3.5%.

Table 16. Average classification results obtained for proposed method compared with the work of other researchers, separating three age groups

Methods	*Dehshibi&bastanfard [55]*	*Ramanathan[54]*	*Kwon & Lobo[8]*	*Proposed Method*
AG1(0-2)	94.28 (3.3)	93.47(2.32)	91.63(2.8)	**96.12(1.51)**
AG2(3-7)	92.19(2.76)	93.14(1.67)	89.33(1.95)	**95.24(1.27)**
AG3(8-19)	85.79(1.57)	86.21(2.13)	84(1.83)	**91.38(2.69)**

Finally, Table 17 demonstrates the rate of success of the complete proposed algorithm achieved in classifying the facial images in the seven age groups.The first column labels in Table 17, starting from the upper side, should be interpreted asfollows: “All ratios” indicatesthe result of applying complete system by using all geometric ratios of the face without using PSO, “Ave1” and “Std1” are the corresponding average and standard deviation of the results generated using 10 different runs of hold-out method respectively, “Selected ratios” is the results using geometric features proposed by the authors in [56], and “Ave2” and “Std2” are the corresponding average and standard deviation of the results generated using 10 different runs and “Proposed Method” indicates the proposed Method using PSO and the “Ave3” and “Std3” are the corresponding average and standard deviation of the results.

Inspection of Table 17 shows that the proposed method using PSO performed better than the other methods in separating the subjects into seven age groups. It can be seen that the minimum success rate for proposed method using PSO is 88.46%, whereas the maximum success rate is as high as 97.95%, while the minimum average is 85.56% and the maximum success rate is 96.12%.

Table 17. Test phase for separating all seven age groups using proposed method

Age Groups / Percentage	*AG1 (0-2)*	*AG2 (3-7)*	*AG3 (8-19)*	*AG4 (20-29)*	*AG5 (30-45)*	*AG6 (46-60)*	*AG7 (60+)*	*Overall success*
All ratios	95	95.48	88.33	91.54	86.48	83.78	89.47	90.01
Ave1	93.67	**95.33**	90.76	90	87.31	84.07	88.28	89.92
Std1	3.11	1.23	2.7	2.84	2.07	3.51	2.91	-
Selected ratios[22]	96.66	96.12	89.16	92.95	87.83	86.48	92.10	91.61
Average2	96.12	95.43	90.48	90.98	86.15	84.81	**90**	90.57
Std2	1.51	1.48	2.7	2.3	1.77	3.68	2.54	
Proposed Method	97.95	96.19	89.65	94.11	88.46	88.88	93.10	92.62
Average3	**96.12**	95.24	**90.9**	**92.35**	**87.88**	**85.56**	88.62	**90.95**
Std3	1.51	1.27	2.87	1.72	2.23	3.24	3.27	

Our proposed system works exceptionally well in identifying the age of facial images of children aged in 0-2. As shown in the table, the overall success rate for the proposed system is 90.95% which is greater than the other ones. Consistent results were produced across the different sets of runs, where the standard deviation calculated with respect to accuracy, were less than 4%. It should also be noted that the overall system success rate is 92.62% whereas the overall success rate of average values is 90.95%. The performance is relatively stable, as the recorded standard

deviation is small and less than 4%.

5.5.4 Comparative Study

In this thesis, we have presented two different approaches to age group classification of facial images. Methods presented are mainly a classification approach of facial images after extraction of the features. In Chapter 4, we have proposed a novel age classification method that combines geometric feature models and LBP. These combined features are used to classify individuals into six major age groups from AG1 to AG6 aforementioned in Section 4.2.

The secondproposed method, presented in Section 5.3,uses geometric features and wrinkle feature in order to classify facial images into seven major age groups from AG1 to AG7.

For further investigation and in order to compare both proposed methods we used the seven age ranges applied in the second proposed method for the first one as well. As such, the algorithm of the first proposed method remains the same while the adult age ranges becomes from AG4 to AG7. For this purpose, the LBP features extracted from “20 above” facial images are applied to SVM classifier to classify the images into one of the four adults Age Groups, where AG4 contains faces of age 20-29, AG5 is 30-45, AG6 is 46-60 and finally 60+ are placed in AG7.

Using the same training and testing sets shown in Table 11, we show the recognition performance of the first proposed method in Table 18.

Table 18: Accuracy rate using geometric ratio and LBP features

Age Groups	(0-2)	(3-7)	(8-19)	(20-29)	(30-45)	(46-60)	(60+)	overall success
First proposed Method	97.92	94.29	88.28	88.23	84.61	77.77	86.2	88.18

The result in Fig. 18 illustrates that the second proposed method using geometric ratios and wrinkle analysis out performs in all seven age ranges than the first one which uses geometric ratios and LBP.

This provides evidence that discriminating among features, plays an important role in answering the question which features representation is optimal for facial classification. LBP features might not be as efficient enough as wrinkle features to utilize the discriminatory information available at old age ranges.

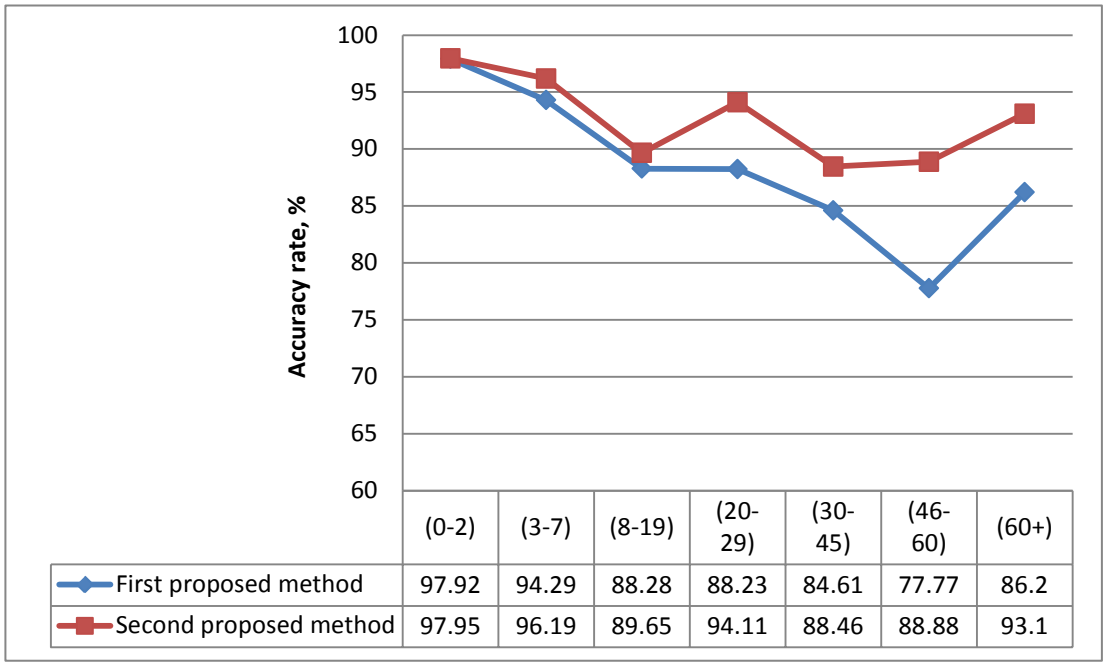


Figure 18.Comparison of both proposed methods

Chapter 6

CONCLUSION

In this thesis, we have presented two different approaches to age group classification of facial images. In the first part of our study, we have proposed an advanced and novel age classification method based on the characteristics of human face development that combines Geometric feature models and Local Binary Patterns, to substantially improve the accuracy rate of age classification over the current state-of-the-art methods. Achieving efficiency, in both, feature extraction and classifier training without decreasing the performance of the system is challenging in the age classification of facial images. We address this challenge using two publicly available databases namely FGNET and IFDB. In this technique, the geometric features extracted from input facial images are used to discriminate young faces from adults. The faces classified as adults are then passed to the adult age-estimation function. These combined features are used to classify subjects into six major age groups, namely AG1 (0-2), AG2 (3-7), AG3 (8-19), AG4 (20-40), AG5 (41-60), and AG6 (60+). The experimental results clearly show the superiority of Local Binary Patterns over Principal Component Analysis (PCA) and subspace Linear Discriminant Analysis (subspaceLDA) techniques for classifying adult facial images. The experimental results for all age groups demonstrate that LBP technique used in this research is a suitable approach in classifying aged facial images into correct age groups.

In the second part of our study, a novel age classification model is presented that

classifies input facial images in seven age groups, namely AG1 (0-2), AG2 (3-7), AG3 (8-19), AG4 (20-29), AG5 (30-45), AG6 (46-60) and AG7 (60+). The proposed method uses two sets of geometric feature based ratios and three wrinkle features on facial images. The proposed method is substantially fast and efficient as it does not use complex computations and relies on straightforward ratios and wrinkles for estimating the facial ages of the images. The system has been tested using SVC that categorizes the images based on their facial features. The performance of the age estimation using facial images has been improved using thewrinkle analysis. In order to achieve high accuracy rate on age group classification, it is demonstrated that the selection of the appropriate features and feature extraction method is essential. Experimental results verify the effectiveness of the proposed age classification method. It is demonstrated that the proposed method is significantly better than the other methods. The proposed system gave a minimum success rate of 88.46%. This success rate was in the most challenging age group (middle aged), where the face does not change very dramatically. The maximum success rate is as high as 97.95% showing that the proposed system works exceptionally well in identifying the age of facial images of children aged 0-2. It should also be noted that the overall system success rate is 92.62%. Additionally, the proposed method attains very high success rate in classifying the infants (0-2) among various age groups.

REFERENCES

[1] C.M. Hill, C.J. Solomon, and S.J. Gibson, “Aging the Human Face-A Statistically Rigorous Approach”, Proc. IEE Int’l Symp. Imaging for Crime Detection and Prevention, pp. 89-94, 2005.

[2] G. Guo, Y. Fu, C. Dyer, and T.S. Huang, “Image-Based Human Age Estimation by Manifold Learning and Locally Adjusted Robust Regression,” IEEE Trans. Image Processing, vol. 17, no. 7, pp. 1178-1188, July 2008.

[3] A. Lanitis, C. Draganova, and C. Christodoulou, “Comparing Different Classifiers for Automatic Age Estimation”, IEEE Trans. Systems, Man, and Cybernetics Part B, vol. 34, no. 1, pp. 621-628, Feb. 2004.

[4] N. Ramanathan and R. Chellappa, “Face Verification across Age Progression”, IEEE Trans. Image Processing, vol. 15, no. 11, pp. 3349-3361, Nov. 2006.

[5] M. Das and A.C. Loui, “Automatic Face-Based Image Grouping for Albuming”, Proc. IEEE Conf. Systems, Man, and Cybernetics, vol. 4, pp. 3726-3731, 2003.

[6] V. Blanz and T. Vetter, “A Morphable Model for the Synthesis of 3D Faces,” Proc. ACM SIGGRAPH, pp. 187-194, 1999.

[7] R.M. Koch, M.H. Gross, F.R. Carls, D.F. Von Buren, F. Fankhauser, and Y.I.H. Parish, “Simulating Facial Surgery Using Finite Element Models”, Proc. ACM SIGGRAPH, pp. 421-C428, 1996.

[8] Y. H. Kwon and N. da Vitoria Lobo. “Age classification from facial images”, Computer Vision and Image Understanding Journal, vol. 74, no. 1, pp. 1-21, 1999.

[9] W. B. Horng, C. P. Lee, C. W. Chen. “Classification of age groups based on facial features”, Tam kang Journal of Science and Engineering ,vol. 4, pp. 183–

192, 2001.

[10] Z. Yang, H. Ai.” Demographic classification with local binary patterns”, LNCSspriger ICB, vol. 4642, pp. 464–473, 2007.

[11] P. J. Phillips, P. J. Rauss, and S. Z. Der, “FERET (Face Recognition Technology) Recognition Algorithm Development and Test Results”, Army Research Lab Technical Report 995, October 1996.

[12] J. Hayashi, M. Yasumoto, H. Ito, Y. Niwa, H. Koshimizu, “Age and gender estimation from facial image processing”, in: Proceedings of the 41st SICE Annual conference vol.1, pp.13-18, 2001.

[13] A. Lanitis. “Evaluating the performance of face-aging algorithms”, in Int. Conf. Face & Gesture Recognition (FG), Amsterdam, The Netherlands, pp. 17-19, 2008.

[14] S. Mancusi, Image Modification, “Aging Progression and Age Regression”, http://www.forartist.com/forensic/modification/modificationpg.htm, 2010.

[15] K. Scherbaum, M. Sunkel, H.-P. Seidel, and V. Blanz, “Prediction of Individual Non-Linear Aging Trajectories of Faces”, Proc. Ann. Conf. European Assoc. Computer Graphics, vol. 26, no. 3, pp. 285-294, 2007.

[16] http://www.scholarpedia.org/article/Facial_Age_Estimation

[17]Electronic Customer Relationship Management (ECRM), http://en.wikipedia.org/wiki/ECRM, 2010.

[18]Haibin Ling, Stefano Soatto, Narayanan Ramanathan, David W. Jacobs, "A Study of Face Recognition as People Age”, Computer Vision, IEEE International, ICCV07(1-8), 2008.

[19] K. Ricanek and E. Boone, “The Effect of Normal Adult Aging on Standard PCA Face Recognition Accuracy Rates”, Proc. Int’l Joint Conf. Neural Networks, pp.

2018-2023, 2005.

[20] E. Patterson, K. Ricanek, M. Albert, and E. Boone, “Automatic Representation of Adult Aging in Facial Images”, Proc. IASTED Int’l Conf. Visualization, Imaging, and Image Processing, pp. 171-176, 2006.

[21] A. Bastanfard, M.A. Nik, M.M. Dehshibi, “Iranian Face Database with Age, Pose and Expression”, Proc. Int’l Conf. Machine Vision, pp. 50-55, 2007.

[22]The FGNET Aging Database, http://www.fgnet.rsunit.com/, http://www-prima.inrialpes.fr/FGnet/ 2012.

[23] R.O. Duda, P. E. Hart, D. G. Stork, “Pattern Classification”, New York: Wiley-Interscience, 2000.

[24] T. Cootes, G. Edwards, and C. Taylor, “Active Appearance Models”, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 23, no. 6, pp. 681-685, 2001.

[25] X. Geng, Z.-H. Zhou, and K. Smith-Miles, “Automatic Age Estimation Based on Facial Aging Patterns”, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 29, no. 12, pp. 2234-2240, 2007.

[26] X. Geng, Z.H. Zhang, G. Li, H. Dai, “Learning from facial aging patterns for automatic age estimation”, in: Proceedings of the 14th ACM International Conference on Multimedia, pp. 307–316, 2006.

[27]C.M. Bishop, “Pattern Recognition and Machine Learning”, Springer, Berlin, 2006.

[28]Y. Fu, G. Guo, and T. S. Huang, “Age Synthesis and Estimation via Faces: A Survey,” IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI), 2010.

[29]K. Ricanek Jr. and T. Tesafaye, “MORPH: a longitudinal imagedatabase of

normal adult age-progression," in Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition (FGR '06), pp. 341–345, Southampton, UK, 2006.

[30] J. Zou, Q. Ji and G. Nagy, "A Comparative Study of Local Matching Approach for Face Recognition", IEEE Transactions on Image Processing, vol.16, no.10, 2007.

[31] A. Lanitis, "Comparative Evaluation of Automatic Age-Progression Methodologies", EURASIP Journal on Advances in Signal Processing, vol.8, no. 2, 2008.

[32]T. Ojala, M. Pietika¨inen, and D. Harwood, "A Comparative Study of Texture Measures with Classification Based on Feature Distributions", Pattern Recognition, vol. 29, no. 1, pp. 51-59, 1996.

[33] L. Nanni, A. Lumini, S. Brahnam, "Local binary patterns variants as texture descriptors for medical image analysis", Journal of Artificial Intelligence in Medicine, vol. 49, no. 2, pp. 117-125, 2010.

[34] T. Ahonen, A. Hadid, M. Pietikainen, "Face recognition with local binary patterns", Proceedings of ECCV 2004, LNCS, vol. 3021, pp. 469–481, 2004.

[35] X. Tan and B. Triggs, "Enhanced Local Texture Feature Sets for Face Recognition under Difficult Lighting Conditions", IEEE Transactions on Image Processing, 2010.

[36]J.-G. Wang, W.-Y.Yau, and H. L. Wang, "Age categorization via ECOC with fused gabor and LBP features", in Proc. IEEE Workshop Applic.Comput.Vis., pp. 1–6, 2009.

[37] K. Luu, T.D. Bui, C.Y. Suen, K. Ricanek, "Combined local and holistic facial features for age-determination", ICARCV, pp.900-904, 2010.

[38]Mark, L. S., Todd, J. T., & Shaw, R. E.“Perception of growth: A geometric analysis of how different styles of change are distinguished”, Journal of Experimental Psychology: Human Perception and Performance, 7, 855 868, 1981.

[39] R. Singh, M. Vatsa, A. Noore, and S.K. Singh, “Age Transformation for Improving Face Recognition Performance”, Proc.Second International Conference on Pattern Recognition and Machine Intelligence, LNCS Vol. 4815, pp. 576-583, 2007.

[40] N. Ramanathan and R. Chellappa, “Face verifcation across age progression”, in Proc. IEEE Conf. Computer Vision and Pattern Recognition, San Diego, CA, vol. 2, pp. 462469, 2005.

[41] Narayanan Ramanathan, Rama Chellappa, Soma Biswas: “Computational methods for modeling facial aging: A survey”, .J. Vis. Lang. Comput. 20(3): 131-144, 2009.

[42] T. Ojala, M. Pietikainen, and M. Maenpaa,“Multiresolution gray-scale and rotation invariant texture clasification with local binary patterns”, IEEE Trans. PAMI, 24:971-987, 2002.

[43] R. Gonzalez and R. Woods, “Digital image processing”, Prentice Hall, 2nd edition, 2002.

[44] Ahonen, T., Hadid, A. and Pietikäinen, M., “Face Description with Local Binary Patterns: Application to Face Recognition”, IEEE Trans. Pattern Analysis and Machine Intelligence vol. 28(12):2037-2041, 2006.

[45] S. J. Lee, S. B. Yung, J. W. Kwon, and S. H. Hong, “Face Detection and Recognition Using PCA”, pp. 84-87, IEEE TENCON, 1999.

[46] B. Moghaddam, and A. Pentland, “An Automatic System for Model-Based

Coding of Faces", pp. 362-370, IEEE, 1995.

[47] J. L. Crowley, and K. Schwerdt, "Robust Tracking and Compression for Video Communication", pp. 2-9, IEEE, 1999.

[48] M. A. Turk, and A. P. Pentland, "Face Recognition using Eigenfaces", pp. 586-591, IEEE, 1991.

[49] W. Zhao, "Subspace Methods in Object/Face Recognition", pp.

3260-3264, IEEE, 1999.

[50]Upton,G. and Cook, I., "Oxford Dictionary of Statistics", 2nd ed. Oxford:OUP. p79, 2008..

[51] P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection", IEEE Trans. Pattern Analysis and Machine Intelligence, 19(7), 1997, 711-720.

[52]Aleix M. MartõÂnez, Avinash C. Kak, "PCA versus LDA", IEEE Trans. Pattern Anal.Mach. Intell., vol. 23, no. 2, pp. 228-233, 2001.

[53] McLachlan, G. J.,"Discriminant Analysis and Statistical Pattern Recognition", Wiley Interscience, 2004.

[54] N. Ramanathan, R. Chellappa, "Modeling Age Progression in Young Faces", IEEE Computer Vision and Pattern Recognition (CVPR), 1, pp.387-394 , 2006.

[55] M.M. Dehshibi, A. Bastanfard, "A new algorithm for age recognition from facial images, Signal Processing", 90 (8), pp. 2431-2444, 2010.

[56]Izadpanahi S, Toygar O, "Geometric feature based age classification using facial images", IET Image Processing, London, 2012.

[57] L. G. Farkas, "Anthropometry of the Head and Face", Raven Press, New York, 1994.

[58]M.-H. Yang, J.K. David, N. Ahuja, "Detecting faces in images: a survey" , IEEE

Transactions on Pattern Analysis and Machine Intelligence 24 (1), pp. 34–58, 2002.

[59] M. G. Rhodes. “Age estimation of faces: a review”, Applied Cognitive Psychology, vol. 23, no. 1, pp. 1–12, 2009.

[60]PR-Tools http://www.prtools.org/

[61] W. L. Chao, J. Z. Liu, J. J. Ding, “Facial age estimation based on label-sensitivelearning and age-oriented regression”, Pattern Recognition, vol. 46 (3), pp. 628-641, 2013.

[62] J. Lu, Y. P. Tan, “Ordinary Preserving Manifold Analysis for Human Age and Head Pose Estimation”, IEEE Transactions on Human-Machine Systems, vol. 43 (2), 2013.

[63]M. P. Beham and S. M. M. Roomi, “A Review of Face Recognition methods”,International Journal of Pattern Recognition and Artiffcial Intelligence, vol. 27 (04), 2013.

[64] H. Demirel, G. Anbarjafari, “Pose Invariant Face Recognition Using Probability Distribution Functions in Different Color Channels”, IEEE Signal Processing Letters, vol. 15, pp. 537-540, 2008.

[65] A. Lanitis, N. Tsapatsoulis, “Quantitative evaluation of the effects of aging”, IET Computer Vision, vol. 5 (6), pp. 338-347, 2011.

[66] J.B. Pittenger , R.E. Shaw, “Aging faces as viscal-elasticevents : Implications for a theory of nonrigid shape perception”, Journal of Experimental Psychology : Human Perceptionand Performance”, vol. 1(4), pp. 374-382, 1975.

[67] L.S. Mark, J.T. Todd, “The perception of growth in three dimensions”, Journal of Perception and Psychophysics, vol. 33 (2), pp. 193-196, 1983.

[68]F. Gao, H. Ai, “Face Age Classification on Consumer Images with Gabor

Feature and Fuzzy LDA Method", The 3rd IAPR International Conference on Biometrics, University of Sassari, Italy, pp. 2-5, 2009.

[69] H.F. Liau,D. Isa, "Feature selection for support vector machine-based face-iris multimodal biometric system", Expert System With Applications, vol. 38, pp. 11105-11111, 2011.

[70] J. Kennedy,R.C. Eberhart,"Particle swarm optimization", IEEE International Conference on Neural Networks, Perth, Australia, pp. 1942-1948, 1995.

[71] W. Yiheng, L. Liang, and G. Haibing, "Frontier-based multi-robot map exploration using particle swarm optimization", IEEE Symposium on Swarm Intelligence, pp. 1-6, 2011.

[72] A.M. Albert, K. Ricanek,"Implications of Adult Facial Aging Factors onBiometrics". In Biometrics, Vienna, Austria, pp. 89-106, 2010.

[73] A. Lanitis, "A survey of the effects of aging on biometric identity verification", J. Biometrics, vol. 2 (1), pp. 34-52, 2010.

[74] Maryam Eskandari, Önsen Toygar, Hasan Demirel: A New Approach for Face-Iris Multimodal Biometric Recognition using Score Fusion. IJPRAI 27(3) (2013)

[75] Liau HF, Isa D. Feature selection for support vector machine-based face-iris multimodal biometric system. *Expert Systems with Applications*. 2011;38(9):11105–11111.

[76]Fukunaga K. Introduction to Statistical Pattern Recognition, New York:Academic Press, 1990.

[77] B.D. Ripley, Pattern Recognition and Neural Networks, Cambridge:Cambridge University Press,1996.

[78] Izadpanahi S, Toygar O, Human age classification with optimal geometric ratios and wrinkle analysis, International Journal of Pattern Recognition and Artificial

Intelligence (IJPRAI).