

MOHAMED KHIDER UNIVERSITY - BISKRA
FACULTY OF SCIENCE AND TECHNOLOGY
DEPARTMENT OF ELECTRICAL ENGINEERING
DIVISION OF ELECTRONICS
Ref:



جامعة محمد خيضر - بسكرة
كلية العلوم والتكنولوجيا
قسم الهندسة الكهربائية
شعبة الإلكترونيك
المرجع:

Facial Soft Biometrics:

Extracting demographic traits

By

SALAH EDDINE BEKHOUCHE

Thesis submitted to the department of electrical engineering in candidacy for the
Degree of **Doctorate (3rd Cycle)** in **Signals and Communications**.

Members of the jury:

President:	Pr. Zine-eddine Baarir	Prof	University of Biskra
Supervisor:	Dr. Abdelkrim OUAFI	MCA	University of Biskra
Examiner:	Pr. Abdelmalik Taleb-Ahmed	Prof	University of Valenciennes
Examiner:	Pr. Mahmoud Taibi	Prof	University of Annaba
Examiner:	Dr. Salim Sbaa	MCA	University of Biskra
Examiner:	Dr. Athman Zitouni	MCA	University of Biskra

JULY 2017

DEDICATION

I dedicate this modest work to:

My beloved parents: Abdelhamid and Abdallah Fatiha, for their encouragement, affection, advice and sacrifice.

I hope you will find in this work my deep appreciation and respect for you.

My brothers and sisters: Yasmina, Khaled, Nawel, Ali, Hamida, Saloua et Chems eddine.

For their precious assistant and encouragement, when I needed moral support.

My grandmothers: Yamina and Zohra, for your prayers and your loves.

May **Allah** grant you health, happiness and long life and make sure that I never disappoint you.

My fiancée and her family.

All my friends and colleagues.

All my teachers of the Department of Electrical Engineering.

All my English teachers.

Finally, I dedicate it to all those I love and appreciate.

ACKNOWLEDGEMENTS

I gratefully acknowledge my gratitude and indebtedness to my supervisor, **Dr. Abdelkrim Ouafi** for his support and guidance throughout this work.

Special thanks also go to other co-supervisors Pr. **Abdelmalik Taleb-Ahmed**, Pr. **Abdenour Hadid** and Pr. **Fadi Dornaika**.

My thanks go also to the members of the jury who honor me by evaluating this work.

I also extend my sincere thanks and appreciation to all the teachers who have supported and guide me in my work with a critical and wise eye.

ABSTRACT

Abstract

Soft biometrics topic attracted a lot of attention recently due to its ability to improve biometrics systems. It has a lot of traits which can be used in biometrics. Some of these traits is most popular among the other traits. These traits are called demographic traits (ie. age, gender, and ethnicity). It belongs to facial soft biometrics traits. Recently, several applications that exploit demographic attributes have emerged. These applications include : access control, reidentification in surveillance videos, integrity of face images in social media, intelligent advertising, human-computer interaction, and law enforcement.

In this dissertation, facial demographic estimation through facial images is studied. Starting with the existing techniques like Deep Learning-based approaches, Image-Based approaches, and Anthropometrics-based approaches. Also, the databases used for age estimation, gender classification or ethnicity classification are exploited. Moreover, the different evaluation terms are mentioned. Ending with the proposed approach and the results on different databases.

The proposed approach consists of the following three main stages: 1) face alignment and preprocessing; 2) feature extraction and selection; 3) demographic estimation.

The purpose of face alignment is to localize faces in images, rectify the 2D or 3D pose of each face and crop the region of interest. This preprocessing stage is important since the subsequent stages depend on it and since it can affect the final performance of the system. The processing stage can be challenging since it should overcome many variations that may appear in the face image. Feature extraction and selection stage extract the face features. These features are extracted either by a holistic method or by a local method. The extracted features are then selected using a supervised feature selection method in order to omit possible irrelevant features. In the last stage, we propose to feed the obtained features to a hierarchical estimator having three layers where we firstly classify the ethnicity and the gender then we estimate the age.

Finally, the obtained results using different databases was stable and good compared with the state of the art methods. The proposed approach is also suited for real-time applications.

Keywords: Soft Biometrics, Demographic Estimation, Age estimation, Gender Classification, Ethnicity Classification, Facial analysis.

الملخص

القياسات الحيوية الناعمة موضوع جذب الكثير من الاهتمام مؤخرا نظرا لقدرتها على تحسين نظم القياسات الحيوية، للقياسات الحيوية الناعمة الكثير من الصفات التي يمكن استخدامها في القياسات الحيوية. مجموعة من هاته الصفات هي الأكثر شعبية بين الصفات الأخرى، وهذه الصفات تسمى الصفات الديموغرافية (أي العمر والجنس والعرق)، وهاته الأخيرة تنتمي إلى الصفات الحيوية الناعمة الوجهية. وقد ظهرت مؤخرا عدة تطبيقات تستغل الخصائص الديموغرافية. وتشمل هذه التطبيقات التحكم في الوصول، وإعادة التحقق في أشرطة الفيديو المراقبة، وسلامة صور الوجه في وسائل الاعلام الاجتماعية، والإعلان الذكي، التفاعل بين الانسان والحاسوب، والقانون.

في هذه الأطروحة، يتم دراسة التقدير الديموغرافي للوجه من خلال صور الوجه. بدءا من التقنيات والنهج الحالية مثل النهج القائم على التعلم العميق، والنهج القائمة على الصورة، والنهج القائمة على القياسات البشرية. كما يتم دراسة قواعد البيانات المستخدمة لتقدير العمر أو تصنيف الجنس أو التصنيف العرقي. وعلاوة على ذلك، يتم ذكر شروط التقييم المختلفة. وفي الأخير نذكر النهج المقترحة ونتائجها على مختلف القواعد البيانية المعروفة.

ويتألف النهج المقترح من المراحل الرئيسية الثلاث التالية: (1) محاذاة ومعالجة الوجه؛ (2) بناء الهرم التمثيلي متعدد المستويات للوجه الذي يتم من خلاله استخراج الميزات المحلية؛ (3) تغذية الميزات التي تم الحصول عليها من التسلسل الهرمي الى مقدر من ثلاث طبقات.

الغرض من محاذاة الوجه هو تحديد مكان الوجه في الصور، وتصحيح الوضع ثنائي او ثلاثي الابعاد لكل وجه للحصول على المنطقة ذات الاهتمام وقصها. مرحلة المعالجة المسبقة مهمة لاعتماد المراحل التالية عليها لأنها تؤثر على أداء النظام بصورة كبيرة. مرحلة التجهيز يمكن أن تكون تحديا لأنه ينبغي التغلب على العديد من الاختلافات التي قد تظهر في صورة الوجه. مرحلة استخراج واختيار خصائص الوجه، هاته الخصائص يمكن استخراجها اما بطريقة كلية او محلية بعدها تنفج لتبقى الميزات المهمة. المرحلة الأخيرة هي تصنيف الديموغرافي حيث نبدأ بتصنيف العرق وبناء على ذلك نصحف الجنس وبناء على العرق والجنس المصنفين نقدر السن.

وأخيرا، نذكر النتائج التي تم الحصول عليها باستخدام قواعد بيانات مختلفة، النتائج كانت مستقرة وجيدة مقارنة مع النهج المعروفة، ايضا النهج المقترح مناسب للتطبيقات في الوقت الفعلي.

الكلمات الدلالية: القياسات الحيوية المرنة، التقدير الديموغرافي، التقدير العمري، التصنيف النوعي، التصنيف العرقي، تحليلات الوجه.

SCIENTIFIC PRODUCTIONS

Publications in journals

- **SE. Bekhouche**, A. Ouafi, F. Dornaika, A. Taleb-Ahmed, and A. Hadid, “Pyramid multi-level features for facial demographic estimation,” *Expert Systems with Applications*, vol. 80, pp. 297–310, Sep. 2017.

Publications in international conferences

- **SE. Bekhouche**, A. Ouafi, A. Taleb-Ahmed, A. Hadid, and A. Benlamoudi, “Facial age estimation using bsif and lbp,” *First International Conference on Electrical Engineering ICEEB’14*, December 2014.
- A. Benlamoudi, A. Samai, A. Ouafi, A. Taleb-Ahmed, **SE. Bekhouche**, and A. Hadid, “Face spoofing detection from single images using active shape models with Stasm and LBP,” *Troisième Conférence internationale sur la Vision Artificielle CVA’ 2015*, April 2015.
- **SE. Bekhouche**, A. Ouafi, A. Benlamoudi, A. Taleb-Ahmed, and A. Hadid, “Facial age estimation and gender classification using multi level local phase quantization,” *2015 3rd International Conference on Control, Engineering & Information Technology (CEIT)*, May 2015.
- A. Benlamoudi, D. Samai, A. Ouafi, **SE. Bekhouche**, A. Taleb-Ahmed, and A. Hadid, “Face spoofing detection using local binary patterns and Fisher Score,” *2015 3rd International Conference on Control, Engineering & Information Technology (CEIT)*, May 2015.
- **SE. Bekhouche**, A. Ouafi, A. Benlamoudi, A. Taleb-Ahmed, and A. Hadid, “Automatic age estimation and gender classification in the wild,” *International Conference on Automatic control, Telecommunications and Signals (ICATS15)*, November 2015.

- A. Benlamoudi, D. Samai, A. Ouafi, **SE. Bekhouche**, A. Taleb-Ahmed, and A. Hadid, “Face spoofing detection using multi-level local phase quantization (ML-LPQ),” International Conference on Automatic control, Telecommunications and Signals (ICATS15), November 2015.

Publications in national conferences

- ME. Zighem, A. Ouafi, **SE. Bekhouche**, A. Benlamoudi, and A. Taleb-Ahmed, “Age Estimation Based on Color Facial Texture,” 10^{ème} Conférence sur le Génie Electrique, April 2017.
- A. Benlamoudi, F. Bougourzi, ME. Zighem, **SE. Bekhouche**, A. Ouafi, and A. Taleb-Ahmed, “Face Anti-Spoofing Combining MLLBP and MLBSIF,” 10^{ème} Conférence sur le Génie Electrique, April 2017.
- F. Bougourzi, **SE. Bekhouche**, ME. Zighem, A. Benlamoudi, A. Ouafi, and A. Taleb-Ahmed, “A Comparative Study On Textures Descriptors In Facial Gender Classification,” 10^{ème} Conférence sur le Génie Electrique, April 2017.

TABLE OF CONTENTS

	Page
List of Tables	x
List of Figures	xii
List of Acronyms	xvi
1 INTRODUCTION	1
1.1 Overview	2
1.2 Problematic	4
1.3 Contributions	5
1.4 Thesis structure	5
2 LITERATURE REVIEW	6
2.1 Introduction	7
2.2 Databases	7
2.2.1 FG-NET	7
2.2.2 MORPH	8
2.2.3 FERET	11
2.2.4 LFW	12
2.2.5 PAL	13
2.2.6 IoG	14
2.2.7 Adience	15
2.2.8 IMDB-WIKI	16
2.3 Age estimation	17
2.3.1 Evaluation metrics	19
2.3.2 Existing works	20
2.4 Gender classification	28

2.4.1	Evaluation metrics	28
2.4.2	Existing works	28
2.5	Ethnicity classification	33
2.6	Demographic estimation	33
2.7	Conclusion	36
3	EXTRACTING DEMOGRAPHIC TRAITS FROM FACE IMAGES	37
3.1	Introduction	38
3.2	Face preprocessing	38
3.2.1	Facial parts detection	38
3.2.2	Pose correction	39
3.2.3	Face region selection	40
3.3	Feature extraction using PML	41
3.3.1	Texture descriptors	41
3.3.2	PML face representations	47
3.3.3	Dimensionality reduction	48
3.4	Hierarchical demographic estimation	49
3.5	Conclusion	50
4	RESULTS AND DISCUSSION	51
4.1	Introduction	52
4.2	Settings	52
4.3	Effect of the face preprocessing	53
4.4	Effect of the face representation on ethnicity classification	56
4.5	Effect of features ratio on the CPU time	62
4.6	Effect of demographic attributes on each other	63
4.7	Single database evaluation	65
4.7.1	MORPH II	65
4.7.2	PAL	68
4.7.3	IoG	70
4.7.4	LFW	71
4.7.5	FERET	72
4.8	Challenge database evaluation	73
4.9	Cross-database evaluation	76
4.9.1	Age estimation	77
4.9.2	Gender classification	77

TABLE OF CONTENTS

ix

4.9.3 Ethnicity classification	78
4.10 Real-time application	78
4.11 Conclusion	79
5 CONCLUSIONS AND FUTURE WORKS	80
5.1 Conclusions	81
5.2 Future works	82
Bibliography	83

LIST OF TABLES

TABLE	Page
1.1 Table of soft biometric traits [1].	3
2.1 Number of Facial Images by Gender and Ethnicity in MORPH I database.	9
2.2 Number of Facial Images by Gender and Age in MORPH I database.	9
2.3 Number of Facial Images by Gender and Ethnicity in MORPH II database.	10
2.4 Number of Facial Images by Gender and Ethnicity in FERET database.	12
2.5 Number of Facial Images by Gender in LFW database.	12
2.6 Number of Facial Images by Gender and Ethnicity in PAL database.	14
2.7 Number of Facial Images by Age and Gender in IoG database.	14
2.8 Distribution of gender in IMDB-WIKI database.	16
2.9 A comparison of demographic estimation approaches.	36
4.1 Results of demographic estimation on four databases with three scenarios: 2D face alignment, 3D face alignment without symmetry, and 3D face alignment with face symmetry.	53
4.2 Accuracy of demographic estimation as a function of different features ratios. The last row depicts the Central Processing Unit (CPU) time (in seconds) of the training phase as a function of different features ratios.	62
4.3 CPU time associated with the testing phase as a function of different features ratios.	63
4.4 Age estimation for different ethnicities and gender on MORPH II database. The ground-truth ethnicity and gender are used by the age estimator.	64
4.5 Gender classification using predicted and ground-truth ethnicities.	65
4.6 A comparison of the proposed approach with other demographic estimation approaches on MORPH II database.	66
4.7 A comparison of the proposed approach with other demographic estimation approaches on PAL database.	69
4.8 Confusion matrix of age classification on IoG database.	70

4.9	A comparison of the proposed approach with other demographic estimation approaches on IoG database.	71
4.10	A comparison of the proposed approach with other gender classification approaches on LFW database.	72
4.11	A comparison of the proposed approach with other demographic estimation approaches on FERET database.	72
4.12	Comparison of the results between the participating teams approaches and our approach in the apparent age-estimation competition.	74
4.13	The images distribution based on facial expression and gender in the dataset.	75
4.14	Comparison of the results between the participating teams approaches and our approach in the gender/smiles classification competition.	75
4.15	The performance (MEA in years) of the cross-database evaluation in the case of age estimation.	77
4.16	Cross-database evaluation in the case of gender classification.	78
4.17	Cross-database evaluation in the case of ethnicity classification.	78
4.18	CPU time (ms) of the proposed approach.	79

LIST OF FIGURES

FIGURE	Page
1.1 Examples of soft biometric traits.	2
1.2 Examples of facial demographic estimation.	4
2.1 Samples of images from FG-NET database.	8
2.2 Number of male and female face images per age ranges in FG-NET database. . .	8
2.3 Samples of images from MORPH I database.	9
2.4 Samples of images from MORPH II database.	10
2.5 Distribution of age in MORPH II database.	11
2.6 Samples of images from FERET database.	11
2.7 Samples of images from LFW database.	12
2.8 Samples of images from PAL database.	13
2.9 Distribution of age in PAL database.	13
2.10 Samples of images from IoG database.	14
2.11 Samples of images from Adience database.	15
2.12 Number of male and female face images per age ranges in Adience database. . .	15
2.13 Distribution of age in IMDB-WIKI database.	16
2.14 Samples of images from IMDB-WIKI database.	17
2.15 Changes observed of the face appearance during the age progression.	17
2.16 The six ratios used by Kwon and Lobo [2].	21
2.17 Anthropometric landmarks on the face used by Gunay and Nabiyevev [3].	22
2.18 Euclidean distance and Geodesic distance between pronasal and right exocanthion.	22
2.19 Ylioinas et al. [4] proposed Local Binary Patterns (LBP) magnitude histogram. .	23
2.20 Han <i>et al.</i> [5] hierarchical age estimator.	24
2.21 Bekhouche <i>et al.</i> [6] Multi-Level (ML)-Local Phase Quantization (LPQ) approach.	24
2.22 Yan <i>et al.</i> [7] facial age classification framework structure.	25
2.23 Levi and Hassner [8] facial age classification framework structure.	25
2.24 Wang <i>et al.</i> [9] facial age classification framework structure.	26

2.25	Pipeline of DEX method for apparent age estimation.	26
2.26	Vectorization of the aging pattern [10].	27
2.27	Fiducial points and distances [?].	29
2.28	Local Block Difference Pattern (LBDP) facial features extraction framework [11].	30
2.29	The progressive Convolutional Neural Network (CNN) training with attention. [12].	31
2.30	Tian <i>et al.</i> [13] demonstration of pruning on filter level for their CNN.	32
2.31	Hadid <i>et al.</i> age classification scheme	34
2.32	The proposed hierarchical demographic estimator of Guo and Mu	34
2.33	Han <i>et al.</i> demograpic estimation scheme	35
2.34	The structure of Yi <i>et al.</i> network which can do age estimation, gender classification and ethnicity classification well simultaneously.	35
3.1	General structure of the proposed approach [14].	38
3.2	Facial parts detection step.	39
3.3	The 68 facial landmarks used in Kazemi algorithm.	39
3.4	Facial pose correction step.	40
3.5	Face region selection step.	41
3.6	Different texture primitives detected by the LBP.	42
3.7	The basic operator of LBP.	42
3.8	Kirsch compass kernel.	43
3.9	13 natural images used for filters learning of Binarized Statistical Image Features (BSIF) descriptor [15].	46
3.10	Multi-Block (MB) face representation (level 3).	47
3.11	ML face representation (level 3).	47
3.12	Pyramid Multi-Level (PML) face representation (level 3).	48
3.13	Pyramid Multi-Level Local Phase Quantization example of level 4.	48
3.14	The proposed hierarchical demographic estimator.	50
4.1	Aligned face images from PAL database. The upper part represent the 2D face align- ment (a 2D similarity transform). The middle part represent the 3D frontalization without symetry .The lower part represent the out-plane rotation.	54
4.2	Aligned face images from MORPH II database. The upper part represent the 2D face alignment (a 2D similarity transform). The middle part represent the 3D frontalization without symetry .The lower part represent the out-plane rotation. . .	54

4.3	Aligned face images from IoG database. The upper part represent the 2D face alignment (a 2D similarity transform). The middle part represent the 3D frontalization without symmetry .The lower part represent the out-plane rotation.	55
4.4	Aligned face images from Chalearn database. The upper part represent the 2D face alignment (a 2D similarity transform). The middle part represent the 3D frontalization without symmetry .The lower part represent the out-plane rotation. . .	55
4.5	Ethnicity classification results using LBP descriptor and MB face representation.	56
4.6	Ethnicity classification results using Local Directional Pattern (LDP) descriptor and MB face representation.	57
4.7	Ethnicity classification results using LPQ descriptor and MB face representation.	57
4.8	Ethnicity classification results using BSIF descriptor and MB face representation.	58
4.9	Ethnicity classification results using LBP descriptor and ML face representation.	58
4.10	Ethnicity classification results using LDP descriptor and ML face representation.	59
4.11	Ethnicity classification results using LPQ descriptor and ML face representation.	59
4.12	Ethnicity classification results using BSIF descriptor and ML face representation.	60
4.13	Ethnicity classification results using LBP descriptor and PML face representation.	60
4.14	Ethnicity classification results using LDP descriptor and PML face representation.	61
4.15	Ethnicity classification results using LPQ descriptor and PML face representation.	61
4.16	Ethnicity classification results using BSIF descriptor and PML face representation.	62
4.17	Absolute age error of misclassified ethnicity images and ground-truth ethnicity images.	64
4.18	Comparison of different face representations using 7 levels in gender classification on MORPH II database.	65
4.19	Comparison of different face representations in age estimation on MORPH II database.	66
4.20	Correlations between True ages and Predicted ages by the proposed approach on MORPH II database.	67
4.21	Cumulative scores obtained by the proposed approach on MORPH II database. . .	67
4.22	Examples of good and poor demographic estimation on MORPH II database. . . .	68
4.23	Correlations between True ages and Predicted ages by the proposed approach on PAL database.	68
4.24	Age estimation performance by the proposed approach on PAL database.	69
4.25	Examples of good and poor demographic estimation on PAL database.	70
4.26	Examples of good and poor demographic estimation on IoG database.	71
4.27	Examples of good and poor demographic estimation on LFW database.	72

4.28	Examples of good and poor demographic estimation on FERET database.	73
4.29	Samples from Chalearn2016 dataset.	73
4.30	Poor face preprocessing on Chalearn2016 dataset (age).	76
4.31	Poor face preprocessing on Chalearn2016 dataset (gender and smile).	76

LIST OF ACRONYMS

AAM	Active Appearance Model
AEM	Accuracy of an Exact Match
AEO	Allowing Errors of One age category
AGES	AGing pattErn Subspace
BIF	Biologically Inspired Features
BSIF	Binarized Statistical Image Features
CAM	Contourlet Appearance Model
CCA	Canonical Correlation Analysis
CH	Color Histogram
CNN	Convolutional Neural Network
CPU	Central Processing Unit
CS	Cumulative Score
DFT	Discrete Fourier Transform
DIF	Demographic Informative Features
FPLBP	Four Patch LBP
GHPF	Gaussian High Pass Filter
GLD	Gray Level Difference
GMM	Gaussian Mixture Models

GWT	Gabor Wavelets Transformation
HOG	Histogram of Oriented Gradients
ICA	Independent Component Analysis
ID	IDentity
KCCA	Kernel Canonical Correlation Analysis
KPLS	Kernel Partial Least Squares
LBDP	Local Block Difference Pattern
LBP	Local Binary Patterns
LDA	Linear Discriminant Analysis
LDP	Local Directional Pattern
LDPs	Local Derivative Patterns
LOSIB	Local Oriented Statistics Information Booster
LPQ	Local Phase Quantization
LSP	Local Salient Patterns
LTP	Local Ternary Patterns
MAE	Mean Absolute Error
MB	Multi-Block
ML	Multi-Level
MLBP	Multi-Quantized Local Binary Patterns
NILBP	Intensity based Local Binary Patterns
OHRank	Ordinal Hyperplanes Ranker
CPNN	Conditional Probability Neural Network
PCA	Principal Component Analysis

PLO	Preserving Locality
PLS	Partial Least Squares
PML	Pyramid Multi-Level
PSF	Point Spread Function
rCCA	regularized Canonical Correlation Analysis
rKCCA	regularized Kernel Canonical Correlation Analysis
ROC	Receiver Operating Characteristic
ROI	Region Of Interest
SIFT	Scale-Invariant Feature Transform
SRC	Sparse Representation Classification
STD	STandard Deviation
STFT	Short Term Fourier Transform
SVM	Support Vector Machine
SVR	Support Vector Regression
WLD	Weber Local Descriptor

INTRODUCTION

Contents

	Page
1.1 Overview	2
1.2 Problematic	4
1.3 Contributions	5
1.4 Thesis structure	5

1.1 Overview

Traditional biometric systems use physiological or behavioral characteristics of a person to identify him. These systems operate in two phases: enrollment and verification. The verification phase becomes very slow in terms of processing time. For this, recent biometrics systems include soft biometrics to assist the main system. In the contrary of biometrics, soft biometrics do not require enrollment which means it doesn't require updating the models each time.

The human face is crucial for the identity of persons because it contains much information about personal characteristics. It provides lots of useful information, including the person's identity, gender, ethnicity, age, emotional expression, etc. Thus, the face image is important for most biometrics systems.

Soft Biometrics traits (see Figure 1.1) are physical, behavioral or adhered human characteristics, classifiable in predefined human compliant categories. The aim of these categories is to distinguish between different individuals [1]. Alphonse Bertillon who was a French police officer and biometrics researcher applied the anthropometry to law enforcement creating an identification system based on physical measurements. Before that time, criminals could only be identified by name or photograph. Because of that, Alphonse Bertillon can be considered as the father of Soft Biometrics [16].



Figure 1.1: Examples of soft biometric traits [17].

Lately, In [18] they redefined Soft Biometrics as a set of traits providing information about an individual, though these are not able to individually authenticate the subject because they

lack distinctiveness and permanence. Another redefinition of Soft Biometrics was proposed in [19] as any characteristic which can be naturally described by humans.

Dantcheva *et al.* [1] divided the soft biometrics traits into face, body or accessory categories. In Table 1.1, they presented a list of soft biometrics traits where the category, the nature, the permanence, and distinctiveness of each trait have been included. The nature of the traits means the type of variable related to this trait, for example the age is continuous because of the human aging. In the other hand, gender considered as binary value because any person either is male or female. For the other term which is the distinctiveness, it is the strength amount of the trait to distinguish between individuals. When it comes to this term, some traits like gender considered very low due to fact that human can be male or female. The permanence expresses how much the trait remains the through the aging process. For instance, the age is considered low/medium in term of permanence. However, gender and ethnicity considered high in term of permanence because these traits are unchangeable.

Table 1.1: Table of soft biometric traits [1].

Traits	Category	Nature	Permanence	Distinctiveness
Skin color	Face	Continuous	Medium	Low
Hair color	Face	Continuous	Medium	Medium
Eye color	Face	Continuous	High	Medium
Beard	Face	Binary	Low/Medium	Low
Moustache	Face	Binary	Low/Medium	Low
Facial measurements	Face	Continuous	High	Medium
Facial shapes	Face	Discrete	High	High
Facial feature measurements	Face	Continuous	High	High
Facial feature shapes	Face	Discrete	High	High
Make-up	Face	Discrete	Low	Low
Ethnicity	Face	Discrete	High	Medium
Marks	Face/Body	Discrete	High	Medium/High
Gender	Face/Body	Binary	High	Low
Age	Face/Body	Continuous	Low/Medium	Medium
Height	Body	Continuous	Medium/High	Medium
Weight	Body	Continuous	Low/Medium	Medium
Gait	Body	Continuous	Medium	Medium
Body measurements	Body	Continuous	Medium/High	Medium/High
Body shapes	Body	Discrete	Medium	Medium
Clothes color	Accessory	Discrete	Low	Medium
Glasses	Accessory	Binary	Low/Medium	Low

Facial demographic estimation (also known as facial demographic classification) is to

estimate demographic attributes (age, gender, and ethnicity) jointly or separately of a person using his/her face image. As a fact, all the demographic can be extracted from the face. Thus, facial demographic estimation considered as part of facial soft biometrics. The age traits is low in permanence and medium in distinctiveness. For the gender, it is low in distinctiveness and high in permanence which is logical. The ethnicity is medium in distinctiveness and high in permanence. The existence of these traits together give is it greater value.

Recently, several applications that exploit demographic attributes have emerged. These applications include access control [18], re-identification in surveillance videos [20], integrity of face images in social media [21], intelligent advertising, and human-computer interaction, law enforcement [22].



Figure 1.2: Examples of facial demographic estimation.

1.2 Problematic

Generally, Surveillance camera networks provide us with images with insufficient resolution for facial recognition. Therefore, the need of additional tools to deal with low-quality images is inevitable, and here the role of soft biometrics comes. As is known, most of the soft biometrics traits can be extracted from low-quality images. These traits provide information about the individual but lack uniqueness and permanence to differentiate sufficiently between two individuals, unlike traditional biometrics which provides information about the identity of the person.

The main aim of this thesis is to improve the accuracy and the computational time of soft biometrics systems. This latter, can be very assistant to biometrics systems in terms of accuracy and identification time. Most of existing works focused only on the demographic traits which can be extracted from the face. For satisfactory results, it is better to improve the three main stages of traditional soft biometrics systems which are image preprocessing, feature extraction, and traits estimation.

1.3 Contributions

The aim of this thesis is the investigation and development of facial demographic estimation system which can support biometrics systems and it can be used in real-time applications. Our main contributions include:

- The use of specific hierarchical order for demographic estimation which leads to good performances as a hierarchical estimator.
- The proposition of a Pyramid Multi-Level face representation which gives most of distinctive features.
- The use of Fisher Score in two different problems (classification and regression).
- The implementation of a real-time application.

1.4 Thesis structure

The rest of this thesis was organized as follows:

In Chapter 2 we offer a literature review to facial demographic estimation, firstly we introduce the demographic attributes and their importance to biometrics. Then we summarize the most used databases on literature. Also, we give all the performance terms for facial age estimation, gender classification and age estimation. Moreover, we give a brief for some existing works. Finally, a comparison between the latest works is given.

Chapter 3 is the description of every method and step used in our approach. Starting with the face preprocessing approach. Then, the feature extraction and selection where different texture descriptor was tested. Also, the proposed hierarchical demographic estimator is illustrated.

Chapter 4 gives all results and discussions of our work. A comparison with the state of the art is also given.

The last chapter is a general conclusion about our work and an envision for some future works.

LITERATURE REVIEW

Contents

	Page
2.1 Introduction	7
2.2 Databases	7
2.2.1 FG-NET	7
2.2.2 MORPH	8
2.2.3 FERET	11
2.2.4 LFW	12
2.2.5 PAL	13
2.2.6 IoG	14
2.2.7 Adience	15
2.2.8 IMDB-WIKI	16
2.3 Age estimation	17
2.3.1 Evaluation metrics	19
2.3.2 Existing works	20
2.4 Gender classification	28
2.4.1 Evaluation metrics	28
2.4.2 Existing works	28
2.5 Ethnicity classification	33
2.6 Demographic estimation	33
2.7 Conclusion	36

2.1 Introduction

Facial recognition has been studied extensively. It has been implemented in many applications. These applications or systems collect other information besides the IDentity (ID) such as age, gender, hair color, eyes color, and ethnicity during the enrollment phase to improve the classification rates and the computation time. These ancillary information or soft biometrics traits like the demographic attributes can be used for other things besides improving the recognition systems.

In this chapter, we present most used databases on demographic estimation including the characteristics and statistics of each database. Also, we summarize the studies about facial demographic estimation. Starting with facial age estimation then moving to facial gender and ethnicity classification. Finally, we present the demographic estimation approaches which estimates at least two attributes. Moreover, we mentioned the evaluation terms for each demographic attribute and the most common application of facial demographic estimation systems.

2.2 Databases

Given the increase of research in the facial demographic estimation, it was expected that there are a lot of databases for benchmarking different algorithms. We summarize most of the existing databases.

2.2.1 FG-NET

The FG-NET¹ [23] aging database was released in 2004 in an attempt to support research activities related to facial aging. Since then a number of researchers used the database for carrying out research in various disciplines related to facial aging (see Figure 2.1). The FG-NET database cited in different research area such as age estimation, face recognition, age progression, gender classification, and face modeling.

¹<http://www.prima.inrialpes.fr/FGnet/>



Figure 2.1: Samples of images from FG-NET database.

The FG-NET database contains 1002 images from 82 different subjects with ages ranging between newborns to 69 years old subjects. However, ages between 0 to 20 years are the most populated in the database. Figure 2.2 shows the statistics of the database.

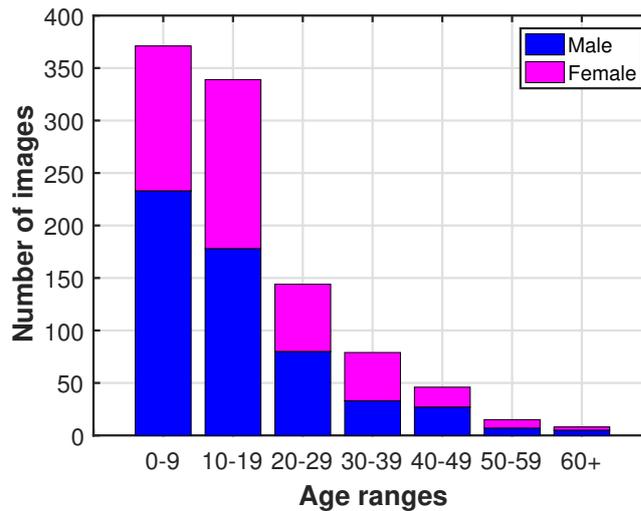


Figure 2.2: Number of male and female face images per age ranges on FG-NET database.

2.2.2 MORPH

MORPH² (Craniofacial Longitudinal Morphological Face Database) [24] is one of the largest publicly available longitudinal face database. The MORPH database corpus embraces thousands of facial images of individuals across time, collected in real-world conditions (uncontrolled conditions). Moreover, these images are available to the public for continued research. MORPH is comprised of two datasets, or “albums,” Album1 and Album2.

²<http://www.faceaginggroup.com/morph/>

2.2.2.1 Album1

Album 1 contains 1690 images of 631 subjects (see Table 2.1 and Table 2.2 for distribution of the images), these images has been taken between October 26, 1962 and April 7, 1998. In each image, they labeled the ID, age, gender, race, weight, height, facial hair, and glasses. This makes Album 1 useful for both facial recognition and facial soft biometrics, but the low number of images make the evaluation insufficient for the performance of facial analysis systems. Some samples from the Album 1 are given in Figure 2.3.



Figure 2.3: Samples of images from MORPH I database.

Table 2.1: Number of Facial Images by Gender and Ethnicity in MORPH I database.

	African	European	Other	Total
Male	1,037	365	3	1,405
Female	216	69	0	285
Total	1,253	434	3	1,690

Table 2.2: Number of Facial Images by Gender and Age in MORPH I database.

	<18	18-29	30-39	40-49	50+	Total
Male	142	803	345	93	22	1,405
Female	15	182	70	18	0	285
Total	157	985	415	111	22	1,690

2.2.2.2 Album2

Album 2 have two versions, the commercial version contains a large set of images, it contains 156,313 unique images of 35,871 subjects. The second version (academic) which is a

subset of the commercial version is available for academic researchers and contains 55,134 images of 13,000 individuals collected over four years labeled age, gender, and ethnicity. In Figure 2.4 we show some samples of this subset.



Figure 2.4: Samples of images from MORPH II database.

MORPH II (MORPH Album 2) is highly recommended to be used in the evaluation of demographic estimation systems due to the high number of images in it and the distribution of the demographic attributes especially age and gender, this can be seen in Table 2.3 for gender and Figure 2.5 for age.

Table 2.3: Number of Facial Images by Gender and Ethnicity in MORPH II database.

Gender	Asian	African	Caucasian	Hispanic	Indian	Other	Total
Male	146	36804	7999	1651	43	3	46646
Female	13	5758	2601	100	14	2	8488
Total	159	42562	10600	1751	57	5	55134

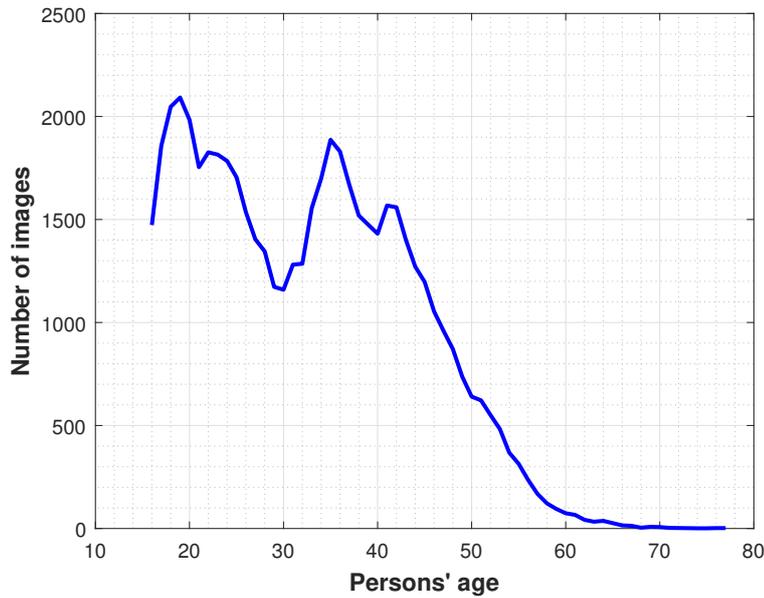


Figure 2.5: Distribution of age in MORPH II database.

2.2.3 FERET

The Color FERET³ database contains a total of 11338 facial images. They were collected by photographing 994 subjects at various angles, over the course of 15 sessions between 1993 and 1996, it has 13 different poses for each subject. The database was created to develop, test, and evaluate face recognition algorithms, also, it is labeled with age gender and ethnicity [25]. Samples from FERET database are shown in Figure 2.6.



Figure 2.6: Samples of images from FERET database.

Distribution of gender and ethnicity is given in Table 2.4.

³<http://www.nist.gov/humanid/colorferet>

Table 2.4: Number of Facial Images by Gender and Ethnicity in FERET database.

Gender	Asian	African	Caucasian	Hispanic	Other	Total
Male	1458	415	4580	277	593	7323
Female	577	473	2499	293	173	4015
Total	2035	888	7079	570	766	11338

2.2.4 LFW

The Labeled Faces in the Wild (LFW⁴) [26] is a database of faces which contains 13,000 images of 1680 celebrities labeled with gender. It has many variations in pose, lighting, focus, resolution, facial expression, age, gender, race, accessories, makeup, occlusions, background, and photographic quality (see Figure 2.7).



Figure 2.7: Samples of images from LFW database.

The distribution of gender in LFW database is shown in Table 2.5.

Table 2.5: Number of Facial Images by Gender in LFW database.

Gender	Images
Male	10256
Female	2977
Total	13233

⁴<http://vis-www.cs.umass.edu/lfw/>

2.2.5 PAL

The Productive Aging Lab Face (PAL⁵) database from the University of Texas at Dallas contains totally 1,046 frontal face images from different subjects (430 males and 616 females) in the age range from 18 to 93 years old (see Figure 2.8 for some samples) [27].



Figure 2.8: Samples of images from PAL database.

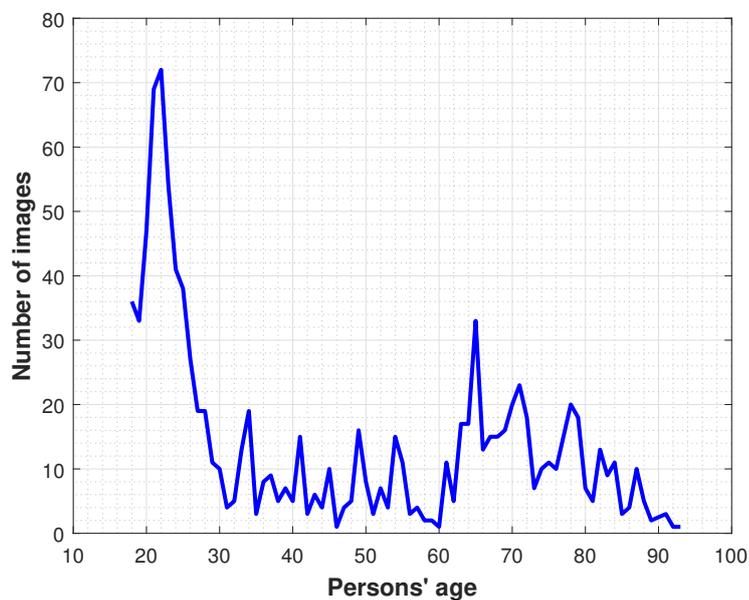


Figure 2.9: Distribution of age in PAL database.

⁵<http://agingmind.utdallas.edu/download-stimuli/face-database/>

Table 2.6: Number of Facial Images by Gender and Ethnicity in PAL database.

Gender	Asian	African	Caucasian	Hispanic	Indian	Total
Male	5	84	263	7	71	430
Female	9	124	469	3	11	616
Toral	14	208	732	10	82	1046

2.2.6 IoG

The Images of Groups (IoG⁶) database contains 5080 images with 28231 faces labeled with age and gender [28]. There are seven age categories as follows: 0-2, 3-7, 8-12, 13-19, 20-36, 37-65, and 66+ years, roughly corresponding to different life stages. In some images, people are sitting, laying, or standing on elevated surfaces. People often have dark glasses, face occlusions, or unusual facial expressions (see Figure 2.10).



Figure 2.10: Samples of images from IoG database.

The IoG database is only labeled with age groups and gender, it has an equitable distribution for gender, but for age, most images are between 20 and 36 years old, this can be observed from Table 2.7.

Table 2.7: Number of Facial Images by Age and Gender in IoG database.

Age groups	0-2	3-7	8-12	13-19	20-36	37-65	+66	Total
Male	439	771	378	956	7767	3604	644	14559
Female	515	824	494	736	7281	3213	609	13672
Total	954	1595	872	1692	15048	6817	1253	28231

⁶<http://agingmind.utdallas.edu/download-stimuli/face-database/>

2.2.7 Adience

Adience⁷ database contains 26,580 images from 2,284 subjects, it is labeled by age and gender, the images of this database have been collected from the web, these images acquired by smartphones and other mobile devices (see Figure 2.11 for some samples), and uploaded without manual filtering to online image repositories [29].



Figure 2.11: Samples of images from Adience database.

The Adience database has 8 age groups which are 0-2, 4-6, 8-13, 15-20, 25-32, 38-43, 48-53, and 60+. The distribution of age and gender is shown in Table 2.12.

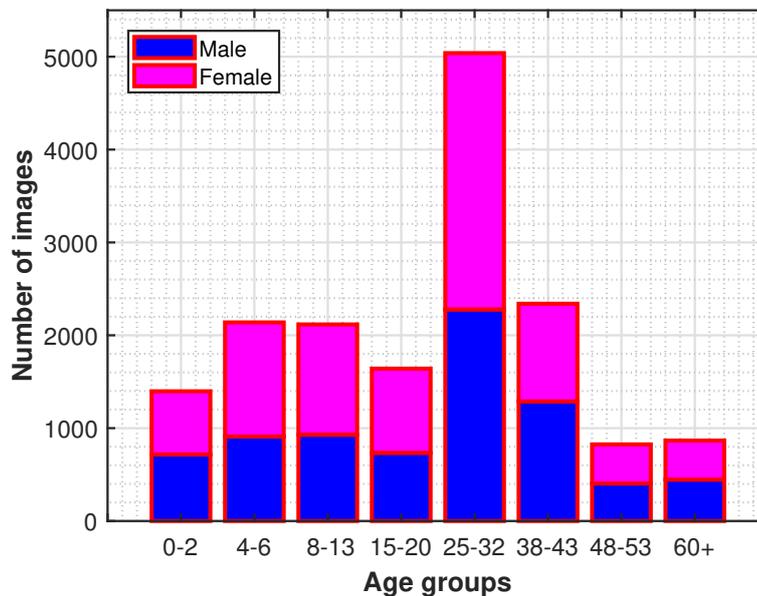


Figure 2.12: Number of male and female face images per age ranges on Adience database.

⁷<http://www.openu.ac.il/home/hassner/Adience/>

2.2.8 IMDB-WIKI

IMDB-WIKI⁸ is the largest publicly available dataset of face images labeled with age and gender, it is a collection of images from Wikipedia⁹ and IMDB¹⁰, it has more than 500,000 images [30]. The distribution of age is given in Figure 2.13 which shows that most of the persons in the database are between 20 and 40 years old.

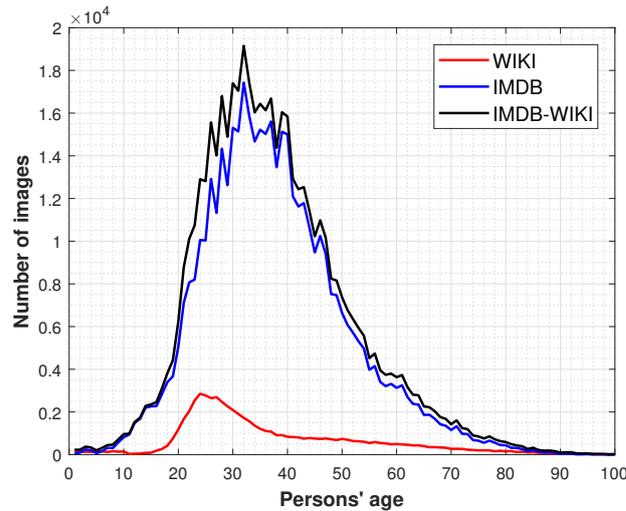


Figure 2.13: Distribution of age in IMDB-WIKI database.

In IMDB-WIKI database, there are some images are mislabeled, to have the correct distribution of this database, every mislabeled image has been ignored, the number of mislabeled images is 11105 images. Table 2.8 shows the distribution of the correct labeled images with gender in IMDB-WIKI database.

Table 2.8: Distribution of gender in IMDB-WIKI database.

	WIKI	IMDB	IMDB-WIKI
Male	47063	263214	310277
Female	12622	189047	201669
Total	59685	452261	511946

Some Samples from IMDB-WIKI database are given in Figure 2.14.

⁸<https://data.vision.ee.ethz.ch/cvl/rrothe/imdb-wiki/>

⁹<https://www.wikipedia.org/>

¹⁰<http://www.imdb.com/>



Figure 2.14: Samples of images from IMDB-WIKI database.

2.3 Age estimation

Humans live a certain period of time. With the progress of time, the human appearance shows some remarkable changes due to the age progression. Gonzalez-Ulloa and Flores [31] presented the observed changes in the external face appearance (see Figure 2.15), the sagging of the skin appears at 30 years of age and at 40 years, the frontal wrinkles start appearing. Starting from 50 years of age, the frontal wrinkles and the Glabellar lines become permanent, then at 60 years, all the wrinkles become deeper and longer. At 70 and 80 years of age, the descent of nasal tip is obvious

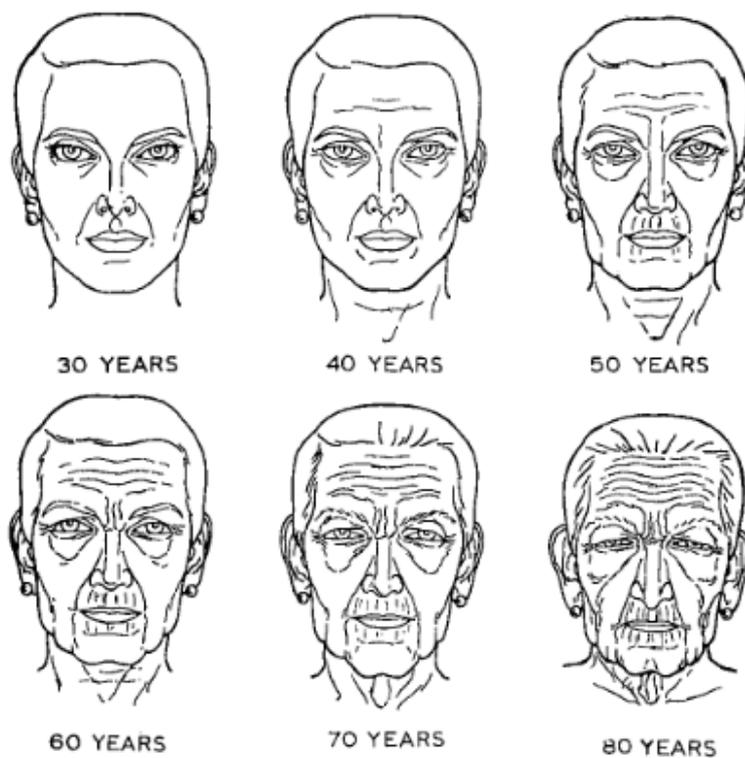


Figure 2.15: Changes observed of the face appearance during the age progression [31].

According to Berry [32], we can recognize some differences between adults and children such as:

1. Infants and young children have larger pupils;
2. Children's lips are redder and proportionately larger than lips of adults;
3. Baby's nose is typically small, wide, and concave.

Predicting human age or age estimation has become an active research topic in computer vision and pattern recognition field, due to the potential applications in human-computer interaction and multimedia communication. Geng *et al.* [33] divided the age into four types:

1. Chronological age: the real or actual age;
2. Appearance age: age of person based on his face appearance;
3. Perceived age: the defined age by others based on the person's appearance;
4. Estimated age: the predicted age by the computer.

Facial age estimation is an important task in facial image classification. It is defined as the age of a person based on his or hers face features. The predicted age can be an exact age (years) or age group (year range) [34].

Predicting the age is a difficult task for humans and it is more difficult for computers, although, the accurate age estimation is very important for some applications [33, 35]. Recently, several applications that exploit the exact age or the age group have emerged. The person's age information can lead to higher accuracy in establishing the user ID for the traditional biometric identifiers which can be used in access control applications. Jain *et al.* [36] proposed a framework for integrating the soft biometric information like age. This framework provided very promising results.

Park and Jain [20] proposed to use the demographic information and facial marks for improving face image matching and retrieval performance, this work have the ability to be used in re-identification in surveillance videos applications.

There are other applications use the person's age information such as integrity of face images in social media, cigarette and alcohol vending machines, law enforcement, and internet safety for minors [14, 33].

2.3.1 Evaluation metrics

There are two types of age estimation evaluation metrics. The first one deal with age groups and the other for exact age. When dealing with age groups there are two popular evaluation metrics, the accuracy of an exact match (AEM) and the accuracy of allowing an error of one age category (AEO). For the exact age, there are many evaluation metrics, the Mean Absolute Error (MAE), the Cumulative Score (CS), the Standard Deviation (STD) of the age error, and Epsilon (EPS).

2.3.1.1 Mean absolute error

The Mean Absolute Error (MAE) is the average of the absolute errors between the ground-truth ages and the predicted ones. The MAE equation is given by [37]:

$$MAE = \frac{1}{N} \sum_{i=1}^N |p_i - g_i| \quad (2.1)$$

where N , p_i , and g_i are the total number of samples, the predicted age, and the ground-truth age respectively.

2.3.1.2 Cumulative score

The Cumulative Score (CS) reflects the percentage of tested cases where the age estimation error is less than a threshold. The CS equation is given by [38]:

$$CS(T) = \frac{N_{e \leq T}}{N} \% \quad (2.2)$$

where T , N and $N_{e \leq T}$ are threshold (years), the total number of samples and the number of samples on which the age estimation makes an absolute error no higher than the threshold.

2.3.1.3 Standard deviation

The Standard Deviation (STD) is a measure that is used to quantify the amount of variation or dispersion of a set of data values [39]. The STD equation is given by:

$$STD = \sqrt{\frac{1}{N} \sum_{i=1}^N (e_i - MAE)^2} \quad (2.3)$$

where N , e_i , and MAE are the total number of samples, the absolute age error, and the mean absolute error respectively.

2.3.1.4 Epsilon

In [40, 41], a new evaluation term for performance called ϵ or *EPS* has been introduced. The EPS equation is given by:

$$EPS = \frac{1}{N} \sum_{i=1}^N 1 - e^{-\frac{(p_i - \mu_i)^2}{2\sigma_i^2}} \quad (2.4)$$

where N and p_i are the total number of samples and the predicted age, μ_i and σ_i are the mean and standard deviation of the human labels.

2.3.1.5 Classification rates

The most popular evaluation term for age group classification is the Accuracy of an Exact Match (AEM), the AEM equation is given by:

$$AEM = \frac{N_m}{N} \% \quad (2.5)$$

where N and N_m are the total number of samples and the number of an exact match with N samples respectively.

Gallagher and Chen [28] introduced a new evaluation term for age group classification called Allowing Errors of One age category (AEO), the AEO equation is given by:

$$AOE = \frac{N_o}{N} \% \quad (2.6)$$

where N and N_o are the total number of samples and the number of correct prediction when allows an error of one age category respectively.

2.3.2 Existing works

From a general view, the age estimation approaches can be categorized based on the image representation and the estimation algorithms. Based on that, we divided the approaches into four types anthropometric-based, image-based, deep-learning-based, and other approaches.

2.3.2.1 Anthropometry-based approaches

The anthropometry-based approaches mainly depend on measurements and distances of different facial landmarks. Kwon and Lobo [2, 42] proposed an age classification method

which classify input images into one of three age groups: babies, young adults, and senior adults. Their method is based on craniofacial development theory and skin wrinkle analysis. The main theory in the area of craniofacial research is that the appropriate mathematical model to describe the growth of a person’s head from infancy to adulthood is the revised cardioid strain transformation, written in polar form as: $\theta' = \theta, R' = R(1 + k(1 - \cos(\theta)))$, where θ is the angle formed from the Y-axis, R is the radius of the circle, k is a parameter that increases over time, and (R', θ') are the successive growths of the circle over time [43].

In the implementation of Kwon and Lobo [2, 42], primary features of the face are found first, followed by secondary feature analysis. The primary features are the eyes, nose, mouth, chin, virtual-top of the head and the sides of the face. From these features, ratios that distinguish babies from young adults and seniors are computed. In secondary feature analysis, a wrinkle geography map is used to guide the detection and measurement of wrinkles. The computed wrinkle index is sufficient to distinguish seniors from young adults and babies.

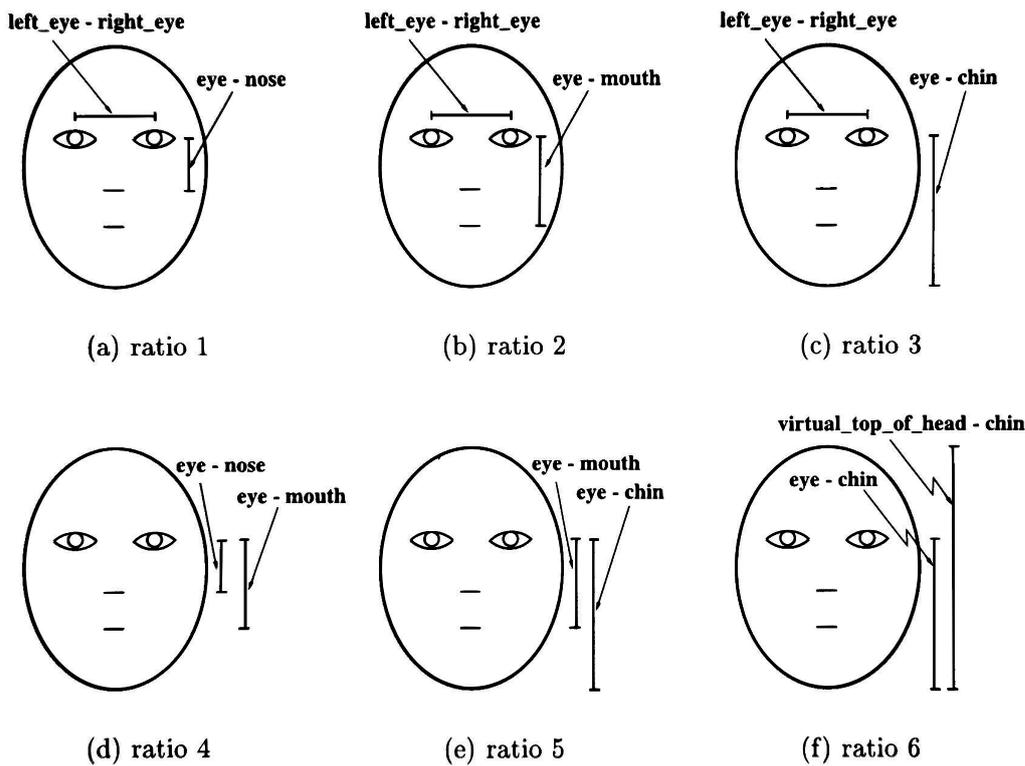


Figure 2.16: The six ratios used by Kwon and Lobo [2].

Gunay and Nabiyevev [3] proposed an approach to detect anthropometric features for age estimation, to locate the facial features they calculate the vertical and horizontal projections and

search them for minimums and maximums. Figure 2.17 shows some of the landmarks which can be used to describe the face.

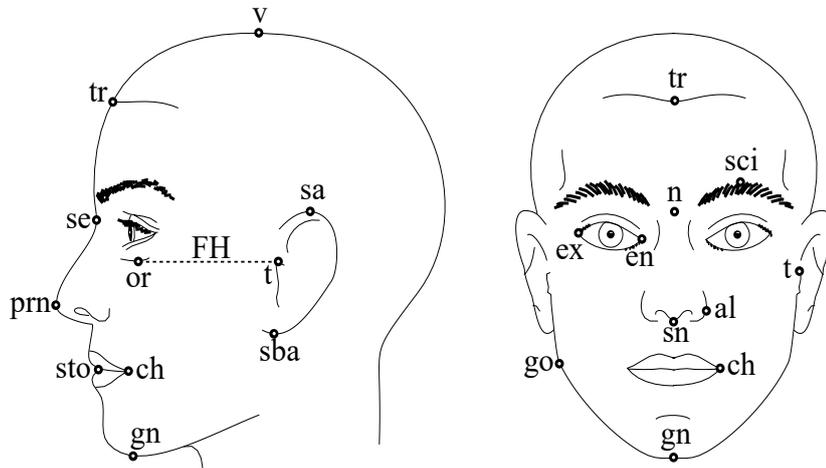


Figure 2.17: Anthropometric landmarks on the face used by Gunay and Nabiyevev [3].

Vezzetti and Marcolin [44] give a survey about the morphometric measures and geometrical features which used to describe faces. Figure 2.18 shows the anthropometric soft-tissue landmarks and two types of distances between the landmarks.

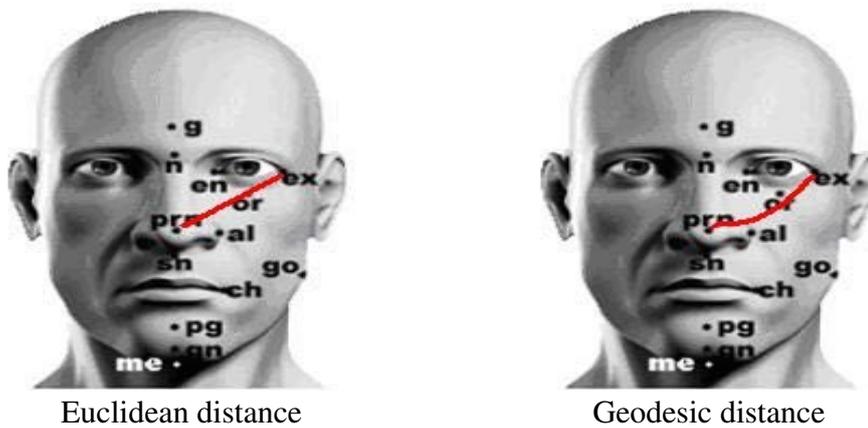


Figure 2.18: Euclidean distance and Geodesic distance between pronasal and right exocanthion.

There are few anthropometry-based approaches due to their weakness when estimating the age of adults or old people. However, these approaches are very useful when it comes to estimating the age of young people due to the growth of the craniofacial.

2.3.2.2 Image-based approaches

Image-based approach is one of the most popular approaches for facial age estimation since a face image can be viewed as texture pattern. Many texture features have been used like LBP, Histogram of Oriented Gradients (HOG), Biologically Inspired Features (BIF), BSIF, and LPQ in age estimation works.

LBP and its variants were also used by many works like in [4, 29, 45–47]. Gunay and Nabiyev [45] used LBP feature as an efficient face descriptor. They divided the faces into small regions from which the LBP histograms are extracted and concatenated into a feature vector. Three different classifiers have been used which are the minimum distance, the nearest neighbor, and the k-nearest neighbor classifiers. They got 80% age classification rates in FERET database. Shan [46] exploited LBP and Gabor features to investigate age estimation on real-life faces acquired in unconstrained conditions. He used Support Vector Machine (SVM) to classify the face image into age groups. He got 50.3% rate on IoG database.

Ylioinas *et al.* [4] proposed a novel method for age classification in unconstrained conditions using a combination of LBP variants as facial features (see Figure 2.19), these features were fed to SVM. They conducted their experiments on IoG database which gives them 51.7% rate. In [48], they proposed another novel method where LBP statistics constructed for facial representation in human age estimation.

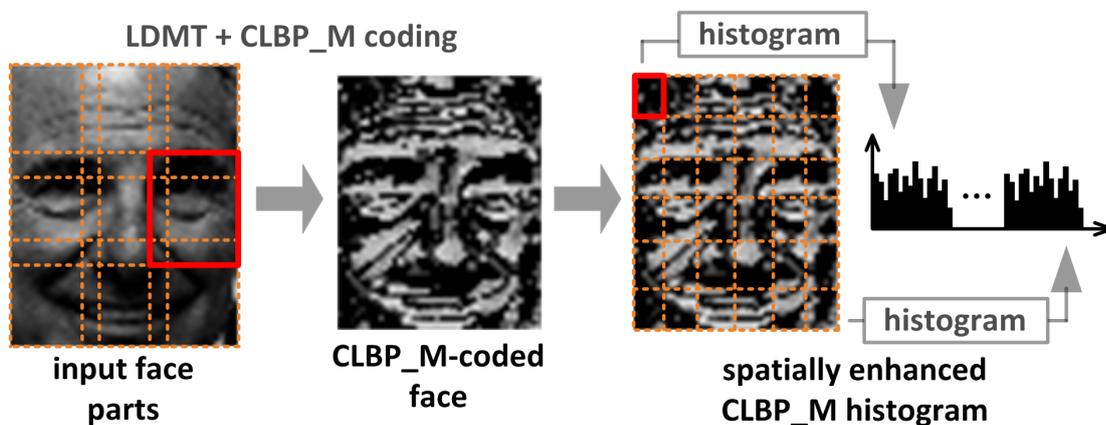


Figure 2.19: Ylioinas *et al.* [4] proposed LBP magnitude histogram.

BIF and its variants are widely used in age estimation works such as [49–52]. Guo *et al.* [49] investigated the BIF features for facial age estimation. They found that the pre-learned prototypes for the S_2 layer and then progressing to C_2 cannot work well for age estimation. Also, the use of Gabor filters with smaller sizes have been proposed. Based on their work in [53], they chose SVM as a classifier and Support Vector Regression (SVR) as a regressor. They

evaluate their approach on YGA and FG-NET databases. They got MAE of 4.77 years on FG-NET, 3.47 years on YGA male, and 3.91 years on YGA female.

Han *et al.* [5] proposed an approach which uses BIF features and a hierarchical age estimator. The BIF features are extracted holistically from different components of the face (see Figure 2.20). Their hierarchical age estimator using SVM on the first layer to classify the age group then based on this group the second classifier estimate the age using SVR. They conducted their approach on three different databases (FG-NET, MORPH II, and PCSO) and they got MAE of 4.6, 4.2, and 5.1 years respectively.

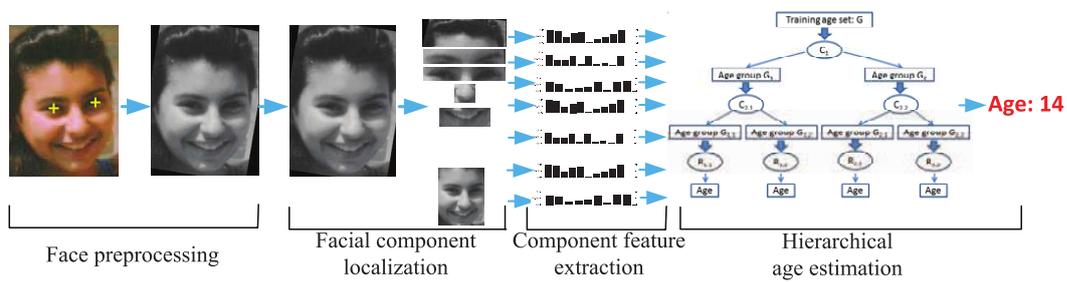


Figure 2.20: Han *et al.* [5] hierarchical age estimator.

In [6], the authors proposed the use of LPQ features which are extracted from ML face representation (see Figure 2.21). They conducted their approach on IoG database, and got AEM of 56% and AEO of 88.8% in age classification.

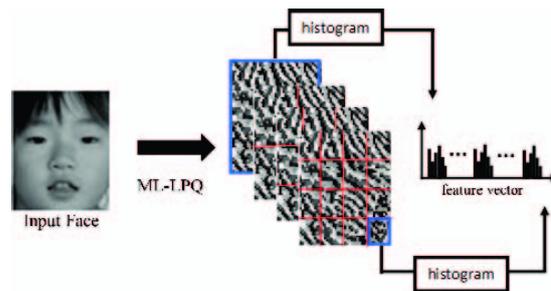


Figure 2.21: Bekhouche *et al.* [6] ML-LPQ approach.

Some researchers used multi-modal features. For instance, we proposed an approach which uses LBP and BSIF features extracted from MB face representation [54]. Fazl-Ersi *et al.* [47] proposed a framework for gender and age classification, where they combined LBP, Scale-Invariant Feature Transform (SIFT), and Color Histogram (CH) features. As well, Eidingner *et al.* [29] mixed LBP with one of its variants which is Four Patch LBP (FPLBP) to do a facial estimation of two attributes (age and gender).

2.3.2.3 Deep-learning-based approaches

Deep learning approaches mainly use the CNN which is a type of feed-forward artificial neural networks in which the connectivity pattern between its neurons is inspired by the organization of the animal visual cortex [55].

Yan *et al.* [7] proposed a novel approach which uses the CNN to extract the facial features instead of using texture descriptors. Their proposed network has 7 layers and it gives 4096 features. For the classification part, they use SVM to classify the face into one of thirteen age groups. figure 2.22 shows their facial age classification framework structure.

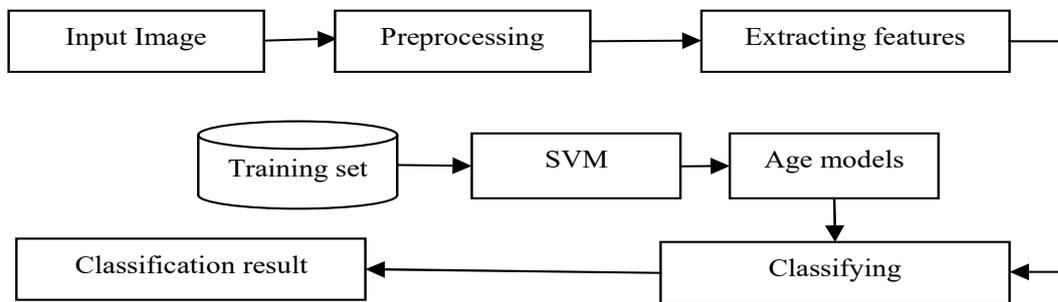


Figure 2.22: Yan *et al.* [7] facial age classification framework structure.

Levi and Hassner [8] proposed a network architecture for both age and gender classification. The first Convolutional Layer contains 96 filters of $7 \cdot 7$ pixels, the second Convolutional Layer contains 256 filters of $5 \cdot 5$ pixels, the third and final Convolutional Layer contains 384 filters of $3 \cdot 3$ pixels. Finally, two fully-connected layers are added, each containing 512 neurons. Figure 2.23 illustrates their CNN architecture. In their experiments, they used Adience database, a significant increase in performance obtained by their architecture.

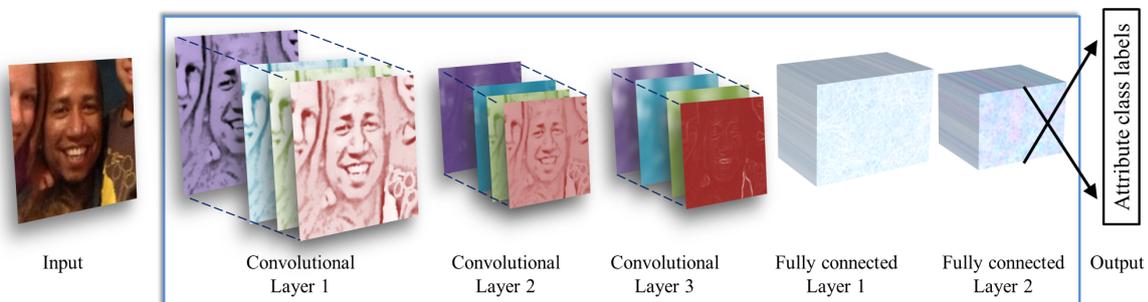


Figure 2.23: Levi and Hassner [8] facial age classification framework structure.

Wang *et al.* [9] investigated deep learning techniques for age estimation based on the CNN, they extract the features based on their built deep learning model. Also, a manifold learning algorithm was incorporated in their proposed scheme. Figure 2.24 shows their CNN structure for extracting deep learned age features.

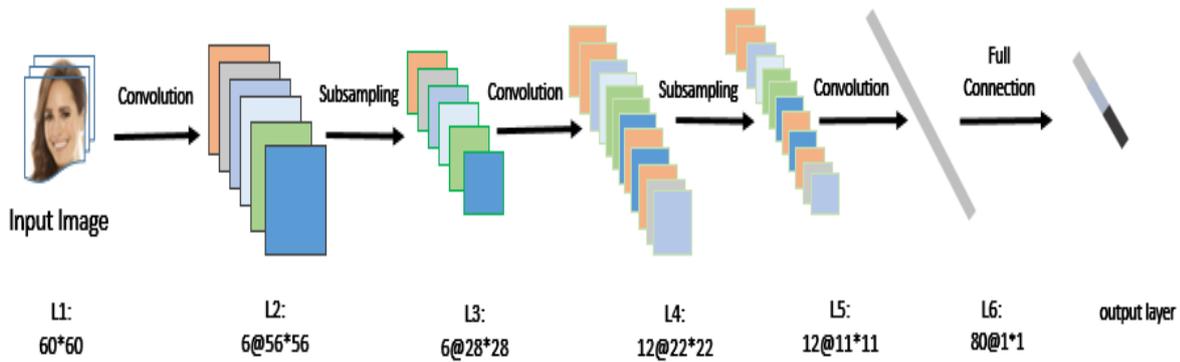


Figure 2.24: Wang *et al.* [9] facial age classification framework structure.

Rothe *et al.* [56] won the ChaLearn LAP 2015 challenge on apparent age estimation [40], their proposed CNN uses the VGG-16 architecture [57]. In their proposed approach which called DEX (Deep EXpectation), firstly it detects the face from the input image and then extracts the CNN predictions from an ensemble of 20 networks on the cropped face (see Figure 2.25).

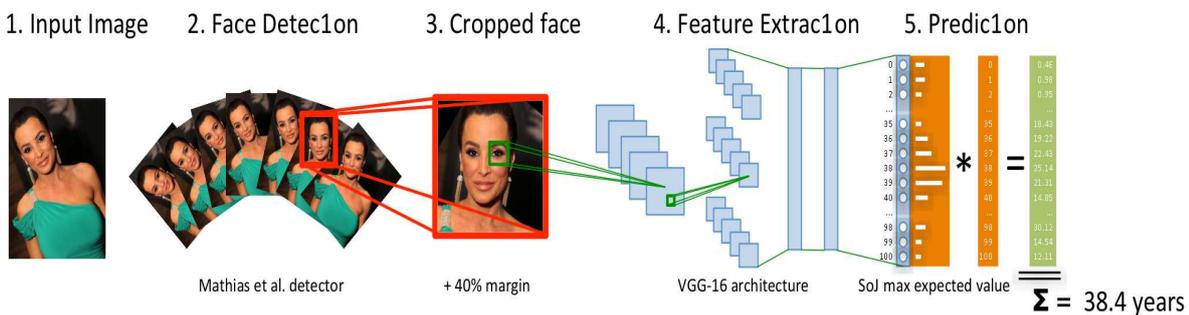


Figure 2.25: Pipeline of DEX method for apparent age estimation [56].

There are other deep-learning-based approaches for facial age estimation besides the above ones such as [58–65].

2.3.2.4 Other approaches

Lanitis *et al.* [66] used Active Appearance Model (AAM) [67, 68] to build an aging function to predict how an individual might look in the future. In [69], they tried different classifier for facial age estimation including Quadratic Functions, Shortest Distance Classifier, Supervised Neural Networks, and Unsupervised Neural Networks.

Geng *et al.* [10] proposed AGing pattErn Subspace (AGES) method for automatic age estimation, by modeling the aging pattern, which is defined as a sequence of personal aging face images, by learning a representative subspace. Figure 2.26 gives an example of the vectorization of the aging pattern. The AGES method has been improved in [38, 70].

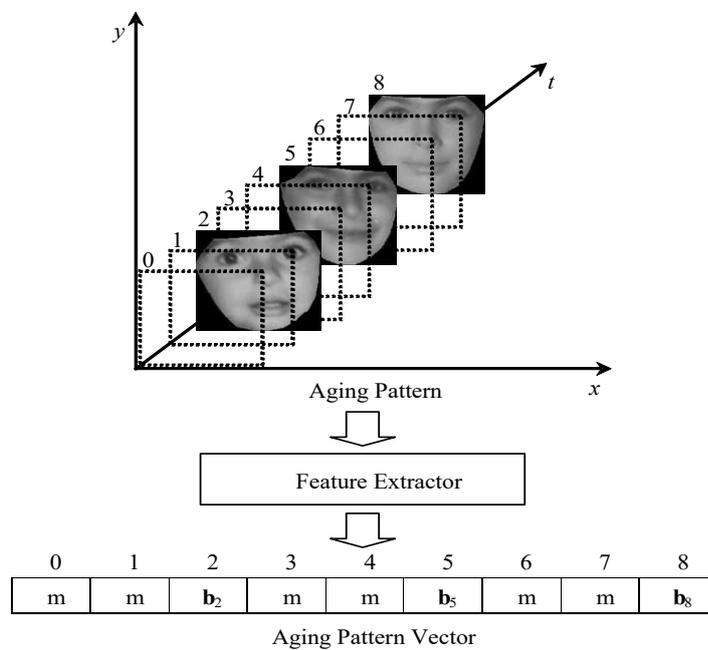


Figure 2.26: Vectorization of the aging pattern [10].

Yan *et al.* [71] presented a patch-based regression framework where the input image is encoded as an ensemble of orderless coordinate patches, the global distribution of which is described by Gaussian Mixture Models (GMM). After that, the patch-kernel is designed for characterizing the Kullback-Leibler divergence between the derived models for any two images, and its discriminating power is further enhanced by a weak learning. Finally, kernel regression is used for human age estimation.

2.4 Gender classification

Facial gender classification task is an interesting topic with a high application potential which attracts more attention in face analysis literature since the gender is one of the most basic information of the human beside the ID and the age. Gender classification is a binary classification problem. Its information considered as one of the facial demographic attributes (age, gender, and ethnicity) which belong to soft biometrics that provides ancillary information of an individual's identity information.

2.4.1 Evaluation metrics

In facial gender classification field, most of the researchers use classification rates which have mentioned in Section 2.3.1.5, but there are some who use the Receiver Operating Characteristic (ROC) curve to report their results.

The ROC curve is a graphical plot that illustrates the performance of a binary classifier system like the gender classification as its discrimination threshold is varied [72].

2.4.2 Existing works

There are several studies in facial gender classification. We divide this studies into geometry-based, appearance-based, and deep-learning-based approaches.

2.4.2.1 Geometry-based approaches

Burton *et al.* [73] used a discriminant function analysis to classify male and female faces using different facial features obtained from the facial key points which are the simple distances between key points in the pictures, the ratios and angles formed between key points in the pictures, and the 3D distances derived by a combination of full-face and profile photographs. They got promising results in that time on their collection data which contains 91 males and 88 female faces.

Fellous [74] extracted individually 40 key points (Fiducial points) from each face (see Figure 2.27). He used the distances between the key points as facial features. The horizontal distances were normalized with respect to the interpupillary distance, whereas vertical distances were normalized with respect to the distance of the eye-midpoint to the philtral ridges midpoint.

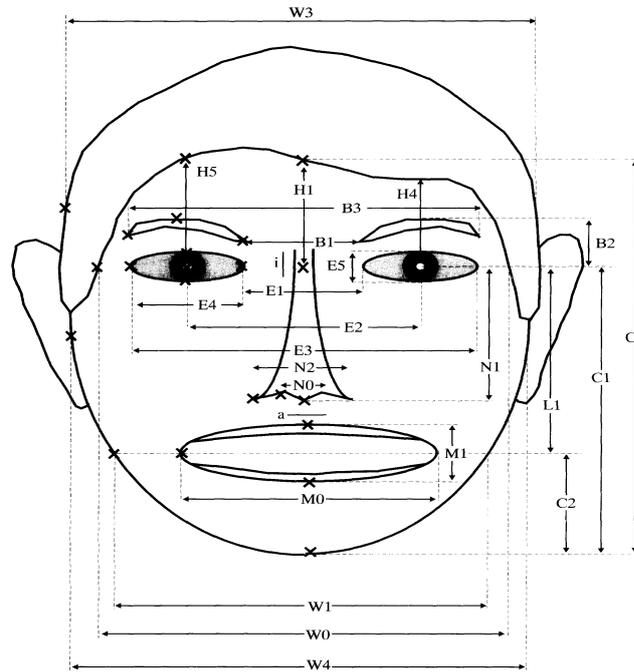


Figure 2.27: Fiducial points and distances [74].

Han *et al.* [75] extracted manually a handful of 3D facial key points from the corresponding geometry meshes. These facial key points broadly define the shape of the eyes, nose, mouth, and cheek. The ratio of the surface area and volume between the points was used as features. SVM classifier was fed with these features to distinguish between male and female.

2.4.2.2 Appearance-based approaches

Jabid *et al.* [76] proposed a novel texture descriptor inspired by LBP called LDP. The LDP descriptor used to represent facial image for gender classification. They divided the face area into small regions, from which LDP histograms are extracted and concatenated into a single vector to efficiently represent the face image. The SVM classifier was used for the classification phase. They conducted their experiments on FERET database and they achieved 95.05% classification rate. The authors of [77] used LBP and SVM along with Principal Component Analysis (PCA).

Hadid *et al.* [78] presented a review of 13 recent and popular LBP variants such as LPQ and BSIF. Also, they gave a comparative analysis on gender classification when using these LBP variants for gender classification. Their extensive experiments showed that LBP basic can deal well with different problems.

Patel *et al.* [79] presented Multi-Quantized Local Binary Patterns (MQLBP) for facial gender classification by performing multi-level vector quantization of Gray Level Difference (GLD) which used both the sign and magnitude components. Their proposed method was evaluated on four datasets (FERET, PAL, CASIA and FEI) through extensive experiments. Moreover, the proposed method has advantages such as higher discrimination power, improved noise robustness and better generalization capability.

Castrillón-Santana *et al.* [80] applied collection of local descriptors for facial gender classification problem. These descriptors are LBP, LBP, Local Ternary Patterns (LTP), Local Derivative Patterns (LDPs), Weber Local Descriptor (WLD), LPQ, HOG, Intensity based Local Binary Patterns (NILBP), Local Salient Patterns (LSP), and Local Oriented Statistics Information Booster (LOSIB). they carried out two techniques to classify the gender. In the first one, the different histograms of the local descriptors are stacked together then classified by SVM. In the second technique, a score was extracted from every local descriptor histogram. These scores are fed to SVM to classify the person gender from his face image. The experiments were conducted on EGA database. In [81], they studied the state-of-the-art accuracies in large datasets MORPH, LFW and IoG.

Lai *et al.* [11] applied Canny edge-detection operator on the input image then used LBDP for feature extraction. Figure 2.28 shows their proposed facial features extraction framework.

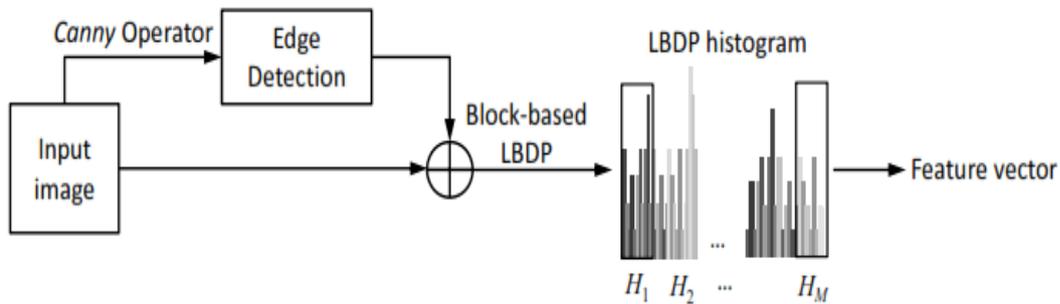


Figure 2.28: LBDP facial features extraction framework [11].

Moeini and Mozaffari [82] proposed a novel approach where the face detected by Viola-Jones face detector from the input image. After that, they divided the face into $8 \cdot 8$ sub-blocks. From each sub-block, the gray pixel values and LBP features are extracted. Then, two separate dictionaries for male and female genders are defined for representing the gender in facial images. Also, two dictionary learning methods are proposed to learn the defined dictionaries in training process. Then, the Sparse Representation Classification (SRC) is adopted for classification in the testing process. Finally, a gender label is given by the probability decision.

2.4.2.3 Deep-learning-based approaches

Juefei-Xu *et al.* [12] designed a progressive convolutional neural network training paradigm to enforce the attention shift during the learning process. The entire training procedure involves $(k + 1)$ epoch groups from epoch group 0 to k , where each epoch group corresponds to one particular blur level. Figure 2.29 illustrates the progressive CNN training with attention. In the training phase. They used images from 5 different databases which are JNET, olympic2012, mugshotDB, pdx2, and Pinellas. However, only two databases (AR Face and PCSO) has been used in the testing phase. In their experiments, they followed two scenarios which are Occlusion Robustness and Low Resolution Robustness. They obtained promising and good results. In the Occlusion Robustness scenario, their proposed network gave 93.12% classification rate. For the Low Resolution Robustness scenario, their proposed network gave 95.67% classification rate.

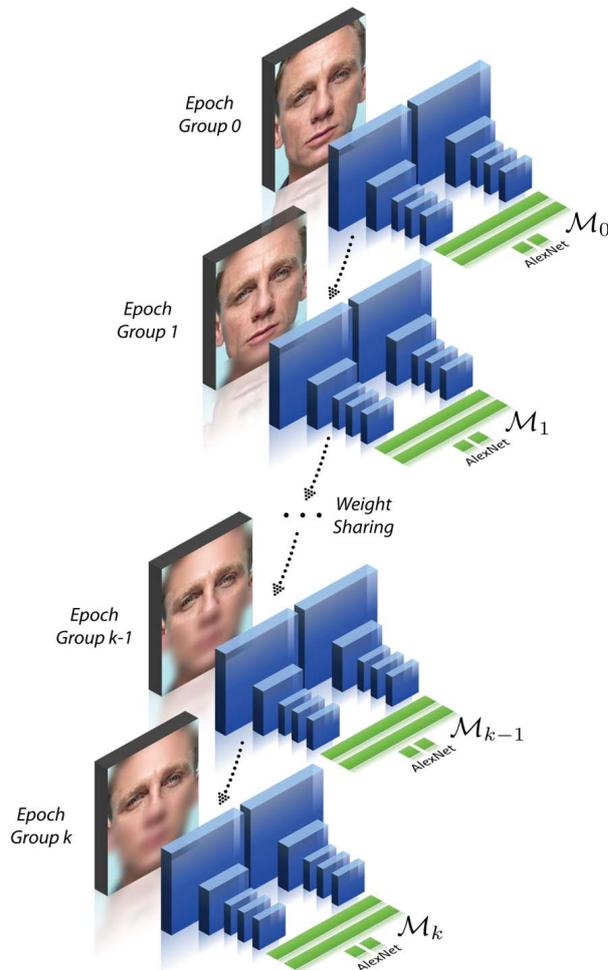


Figure 2.29: The progressive CNN training with attention. [12].

Antipov *et al.* [83] designed a powerful and complex CNN called "Starting CNN" which is a simplification of the CNN proposed by Simonyan and Zisserman [57]. This CNN was minimized progressively to be an optimal CNN for gender recognition from faces. They conducted their experiments on CASIA WebFace and LFW databases where CASIA WebFace used for training and validation. For LFW, it was used for testing. They got 97.31% classification rate on LFW database.

Tian *et al.* [13] developed a 16-layer, yet lightweight, neural network which boosts efficiency while maintaining high accuracy as it looks. Their network is pruned also from the CNN proposed by Simonyan and Zisserman [57] (see Figure 2.30). By using either SVM or Bayesian classification on top of the reduced CNN features, they were able to achieve an accuracy of 92.4% on LFW and 98.0% on CelebA.

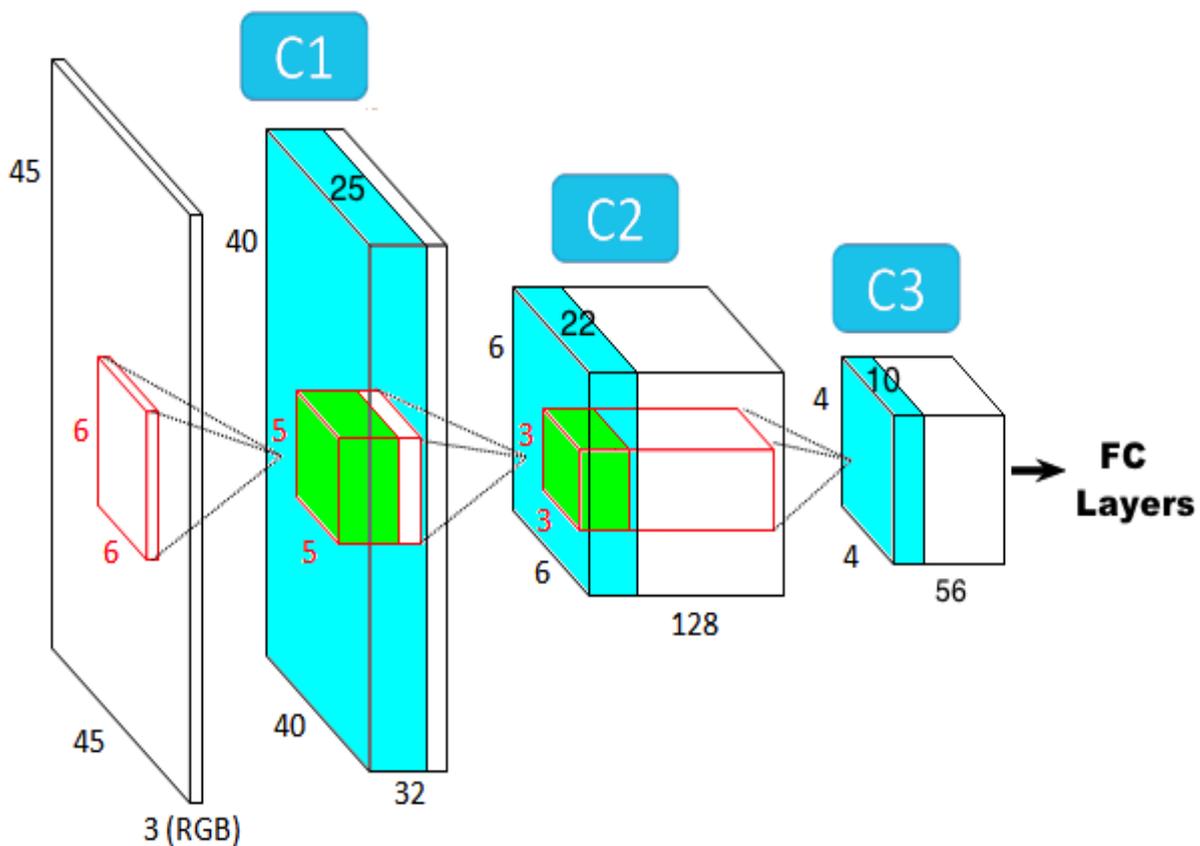


Figure 2.30: Tian *et al.* [13] demonstration of pruning on filter level for their CNN.

2.5 Ethnicity classification

Facial ethnicity classification is the least interested attribute among the demographic attributes. In most researches, it found with the other demographic attributes (age and gender). As we know, facial ethnicity or race classification is a multi-class problem. Therefore, the only known used evaluation term used in literature is the accuracy (classification rates). Lu and Jain [84] presented a facial ethnicity classification scheme using Linear Discriminant Analysis (LDA) to classify face images into two classes (Asian vs. non-Asian). They conducted their experiments on private database which has 2,630 face images and equally balanced. Hosoi *et al.* [85] developed a novel approach for ethnicity classification from facial images. They used Gabor Wavelets Transformation (GWT) and retina sampling to extract facial features. These features are fed to SVM to classify the face image into Asian, African, or European. They achieved approximately 94% for ethnicity estimation under various lighting conditions.

2.6 Demographic estimation

Human demographic estimation has become an active research topic in computer vision and pattern recognition. However, most of the researches focused on a single demographic attribute especially age and gender. There are few researches on all demographic attributes (age, gender and ethnicity) such as [86–93].

In [86], the authors used LBP histogram as features. The steep decent method was applied to find an optimal reference template. Given a local patch, the Chi-square distance between a sample and the optimal reference template is used as a measure of confidence belonging to the reference class. This was applied for all demographic attributes (age, gender and ethnicity). For their experiments, the authors considered the FERET, PIE, and snapshot databases.

The LBP was also investigated in [89] for automatic demographic classification from human faces which includes age categorization, gender recognition and ethnicity classification. They focused on the LBP based spatiotemporal method as a baseline system for combining spatial and temporal information. Furthermore, the correlation between the face images through manifold learning has been exploited. Figure 2.31 illustrates their age classification scheme.

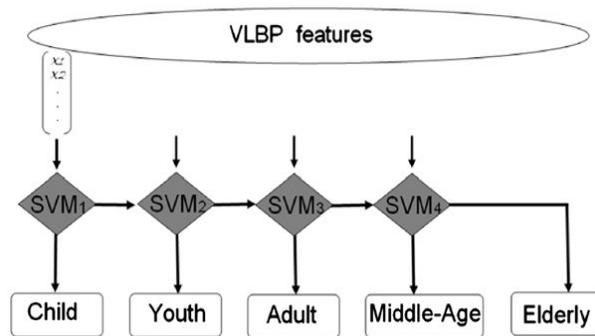


Figure 2.31: Hadid *et al.* age classification scheme [89].

Guo and Mu [87] studied age estimation performance under the variations of race and gender. They proposed a hierarchical demographic estimator with two layers (see Figure 2.32). Their approach gave MAE of 4.45 years on MORPH II database. In [88], they investigated Canonical Correlation Analysis (CCA), regularized Canonical Correlation Analysis (rCCA), Kernel Canonical Correlation Analysis (KCCA), Partial Least Squares (PLS) and Kernel Partial Least Squares (KPLS) methods for demographic estimation. KCCA gave them the best results with 98.4% of gender, 98.9% of ethnicity and MAE of 3.98 years. In [91] presented a framework that can estimate demographic attributes (including age, gender and ethnicity) jointly. They extracted BIF features which are projected by either CCA, PLS or their variants like in [88]. For the stage of estimation, they use SVM and SVR. They conducted their experiments on MORPH II database and got MAE of 3.92 years when using regularized Kernel Canonical Correlation Analysis (rKCCA) with SVM.

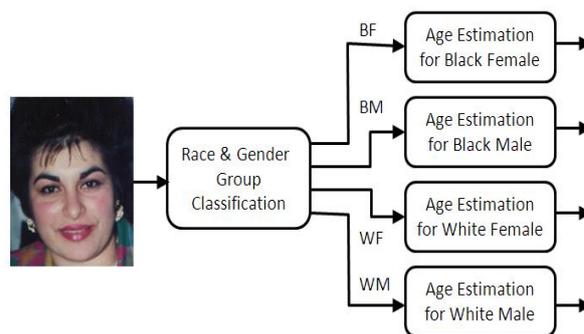


Figure 2.32: The proposed hierarchical demographic estimator of Guo and Mu [87].

Han *et al.* [92] used selected BIF features using Demographic Informative Features (DIF) to estimate the age, gender and ethnicity attributes (see Figure 2.33). To further improve the performance, they designed a quality assessment method to detect low-quality face images.

They evaluate their approach on different databases (FG-NET, FERRET, MORPH II, PCSO and LFW), the performance was well compared to most demographic estimation approaches.

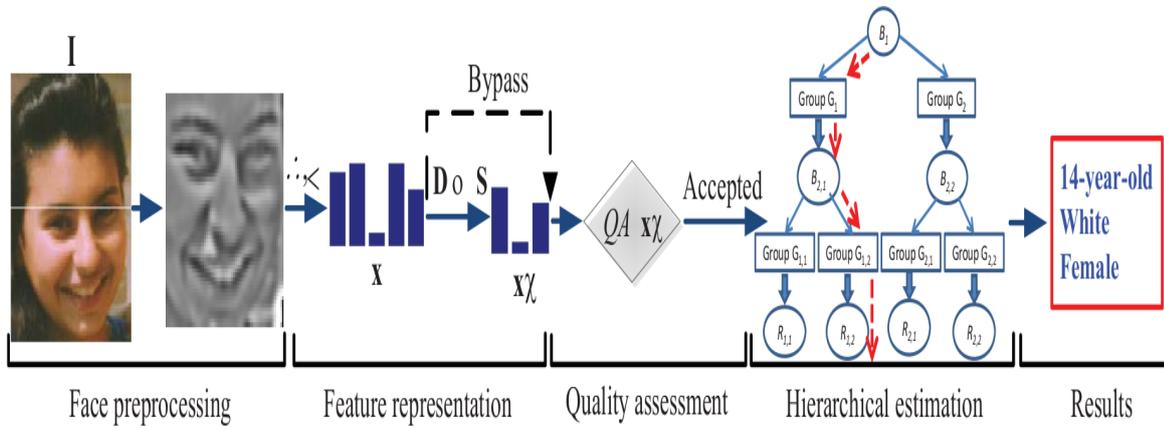


Figure 2.33: Han *et al.* demographic estimation scheme [92].

Yi *et al.* [93] proposed multi-scale CNN which take a face image as input, this image cropped into many local aligned patches. All patches are fed to the multi-scale convolutional network. The response of each patch is combined at the full connected layer to estimate the age, gender and ethnicity, Figure 2.34 illustrates that. They conducted their experiments on MORPH II database and got 98.6% ethnicity rate, 97.9% gender rate, and 3.63 years of MAE.

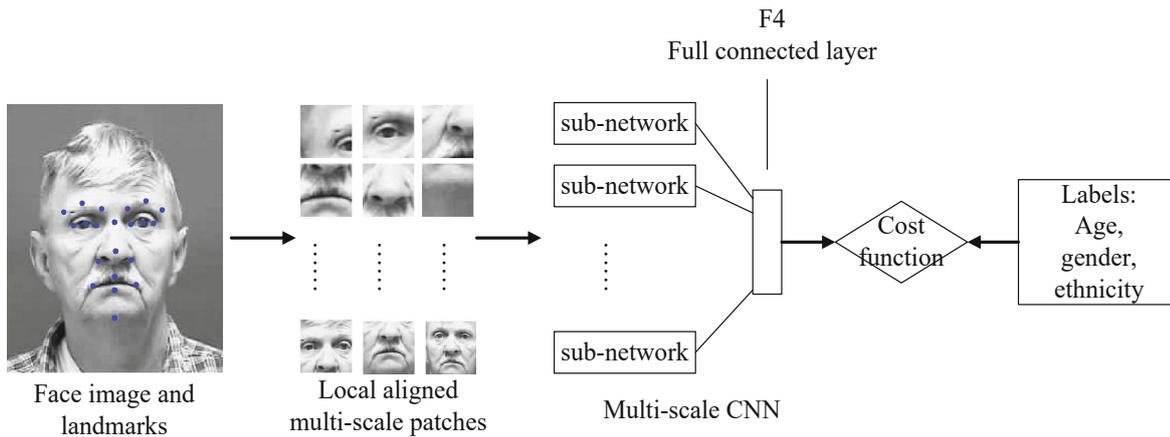


Figure 2.34: The structure of Yi *et al.* [93] network which can do age estimation, gender classification and ethnicity classification well simultaneously.

We summarize the works which include all the demographic attributes as well as their strength and weakness in Table 2.9.

Table 2.9: A comparison of demographic estimation approaches.

Approach	Strength	Weakness	Database
LBP + Real AdaBoost [86]	➤ Suited for real-time applications	➤ Poor performance	FERET PIE
Manifold Learning [89]	➤ Ability to work on different face analysis tasks	➤ High computational cost	Private
BIF+rKCCA+SVM [91]	➤ Good performance	➤ High computational cost	MORPH II
BIF+SVM [92]	➤ Achieving good and stable results in different databases ➤ Suited for real-time applications	➤ Commercial software for face and eye detection	FG-NET MORPH II PCSO LFW FERET
CNN Yi <i>et al.</i> [93]	➤ Achieving good results ➤ Suited for real-time applications	➤ High cost	MORPH II
PML+BSIF+LPQ+SVR [14]	➤ Ability to work on different face analysis tasks ➤ Achieving good and stable results in different databases ➤ Suited for real-time applications	➤ Vulnerable to profile faces and wild poses.	MORPH II PAL IoG LFW FERET ChaLearn2016

2.7 Conclusion

In this chapter, we introduced the most common databases used in facial age estimation, gender classification and ethnicity classification. Also, we explored the different application which can be benefit from the demographic estimation systems. Moreover, an overview is given about facial demographic estimation techniques and a comparison between the latest ones. Furthermore, the different evaluation terms are given for the three demographic attributes.

EXTRACTING DEMOGRAPHIC TRAITS FROM FACE IMAGES

Contents

	Page
3.1 Introduction	38
3.2 Face preprocessing	38
3.2.1 Facial parts detection	38
3.2.2 Pose correction	39
3.2.3 Face region selection	40
3.3 Feature extraction using PML	41
3.3.1 Texture descriptors	41
3.3.2 PML face representations	47
3.3.3 Dimensionality reduction	48
3.4 Hierarchical demographic estimation	49
3.5 Conclusion	50

3.1 Introduction

In this chapter, we will describe the principle of the proposed approach. Our approach is consisting of three main stages, face preprocessing, feature extraction and selection, and demographic estimation. In the first stage, we proposed a face alignment to rectify the face before feature extraction. For feature extraction, we propose a PML face representation. We tested also this proposed method using four texture descriptors. A hierarchical estimator with three layers has been proposed where we classify ethnicity first, then gender and finally age (see Figure 3.1).

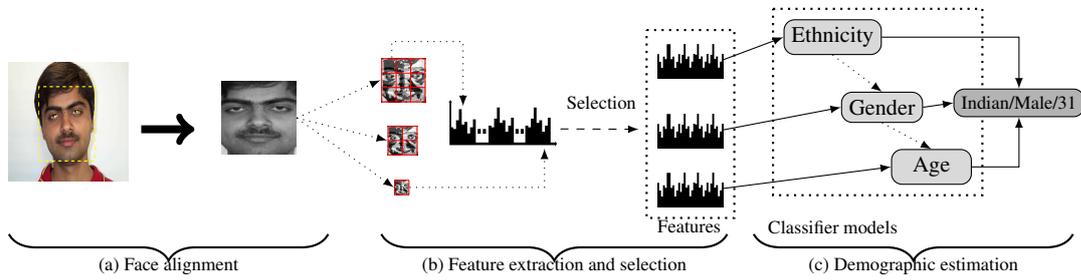


Figure 3.1: General structure of the proposed approach [14].

3.2 Face preprocessing

The face preprocessing is the most important phase in facial soft biometrics systems. Our face preprocessing phase consists of three steps : (i) facial parts detection (ii) pose correction; (iii) face region selection.

3.2.1 Facial parts detection

Firstly, the input image must be converted into a gray-scale image. The formula Eq.3.1 is used to convert the RGB value of each pixel into its gray-scale value. Where R, G, B correspond to the color of the pixel, respectively [94].

$$GRAY = 0.2989 \cdot R + 0.5870 \cdot G + 0.1140 \cdot B \quad (3.1)$$

After that we apply the cascade object detector that uses Viola-Jones algorithm [95] to detect people's faces. Then detecting eyes of each face using Kazemi algorithm [96]. This last is the most used for face landmarks detection. It performs face alignment in milliseconds and achieves accuracy superior to most public algorithms. Figure 3.2 illustrates the facial parts detection step.

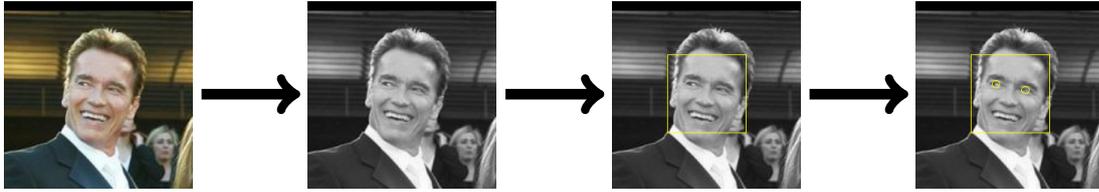


Figure 3.2: Facial parts detection step.

After face detection, the right and the left eye positions are calculated using the following points: (37 and 40) and (43 and 46) respectively (see Figure 3.3). These points are used for the pose correction phase.

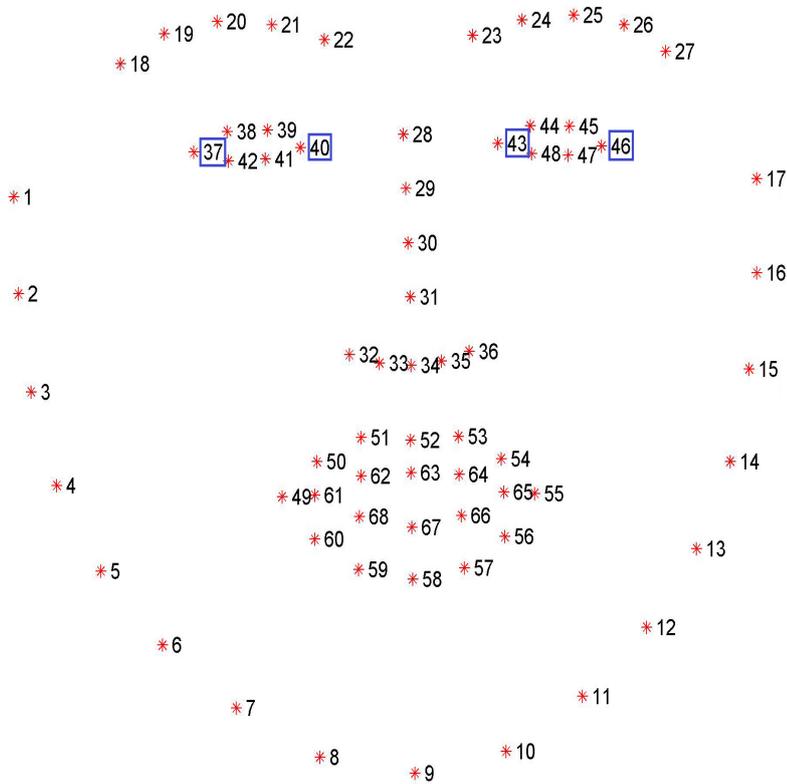


Figure 3.3: The 68 facial landmarks used in Kazemi algorithm.

3.2.2 Pose correction

To correct the pose, we apply 2D transformation based on the eyes center to correct the pose. Like in [14, 54] we rotate clockwise the face by an angle θ around the image center

(C_x, C_y) . The angle is calculated using the formula in Eq. 3.2.

$$\theta = \tan^{-1}\left(\frac{L_y - R_y}{L_x - R_x}\right) \quad (3.2)$$

Then, we find the new coordinates of the eyes center using the formulas Eq. 3.3, , Eq.3.4, Eq.3.5, and Eq.3.6.

$$R'_x = C_x + (R_x - C_x).cos(\theta) - (R_y - C_y).sin(\theta) \quad (3.3)$$

$$R'_y = C_y + (R_x - C_x).sin(\theta) + (R_y - C_y).cos(\theta) \quad (3.4)$$

$$L'_x = C_x + (L_x - C_x).cos(\theta) - (L_y - C_y).sin(\theta) \quad (3.5)$$

$$L'_y = C_y + (L_x - C_x).sin(\theta) + (L_y - C_y).cos(\theta) \quad (3.6)$$

Where $R'(x, y)$, $L'(x, y)$, and $C(x, y)$ are the new right eye, the new left eye, and the center of image points. Figure 3.4 shows how the pose correction step works.

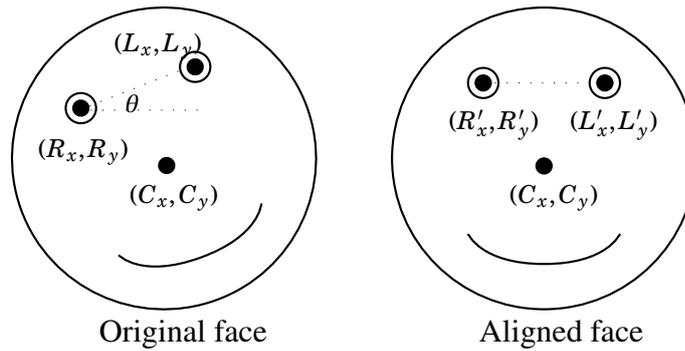


Figure 3.4: Facial pose correction step.

3.2.3 Face region selection

To select the region of interest, we calculate the distance d between the new coordinates of the center points of the eyes where $d = |R'_x - L'_x|$. Then we crop the image based on factor

values k_{side} , k_{top} , and k_{bottom} [14]. In our work and by experiments, we set the value of these factors as follow: $k_{side} = 0.5$, $k_{top} = 1$, and $k_{bottom} = 1.75$. Figure 3.5 illustrates the face Region Of Interest (ROI) which will be used in the feature extraction phase.

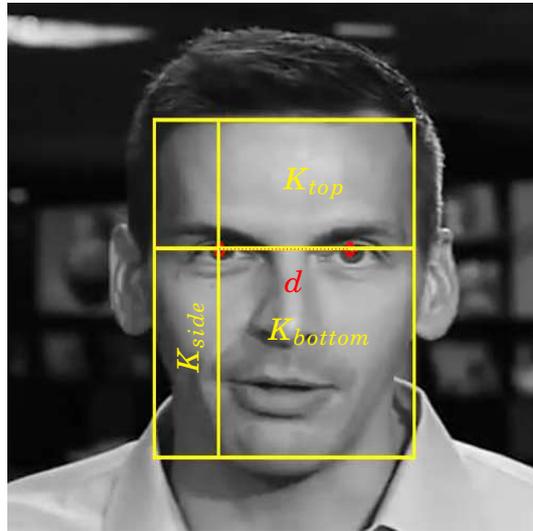


Figure 3.5: Face region selection step.

3.3 Feature extraction using PML

Feature extraction stage has been the most studied topic among the rest stages due to its effective effect on the demographic estimation systems performance. We divided our feature extraction stage into three parts. In the first part, we will present the different texture descriptors used in our work. Then, the proposed PML face representation will be illustrated. Finally, the proposed technique for dimensionality reduction will be presented.

3.3.1 Texture descriptors

In the literature of image classification field, there are a lot of image descriptor which can be used in facial demographic estimation or has been already used in this field. In our work, we used four widely known image descriptors which are LBP, LPQ, LDP, and BSIF.

3.3.1.1 Local Binary Patterns

LBP is very efficient method for analysis of two dimensional textures, which used the pixels of an image by thresholding the neighborhood of each pixel and considers the result as a binary

number. The LBP was used widely in many applications such as face descriptor. The face can be seen as a composition of micro-patterns such as edges, spots and flat areas (see Figure 3.6) which are well described by the LBP operator [97].

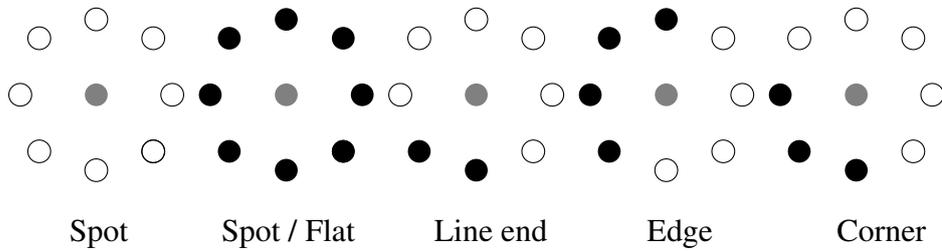


Figure 3.6: Different texture primitives detected by the LBP [98].

The LBP operator was firstly introduced by Ojala *et al.* [99]. It assigns a label for every pixel of an image except the pixels at the border, using the pixel value as threshold for the 3x3 neighborhood and consider the result as binary number (see Figure 3.7). The histogram of these different labels can then be used as features to describe the texture of the image.

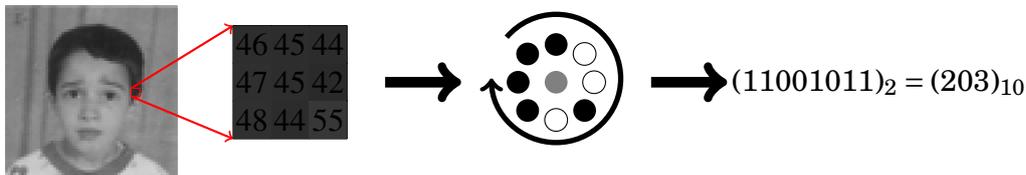


Figure 3.7: The basic operator of LBP.

The LBP operator was lately extended to deal with textures at different scales, using a circular neighborhood and bi-linearly interpolating values at non-integer pixel coordinates allow any radius and number of pixels in the neighborhood. $LBP_{P,R}$ is almost used for pixel neighborhoods and it refers to P sampling points on a circle of radius R . The value of the LBP code of a pixel (x_c, y_c) is given by:

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c)2^p \tag{3.7}$$

where g_c corresponds to the gray value of the center pixel (x_c, y_c) , g_p refers to gray values of P equally spaced pixels on a circle of radius R , and s defines a thresholding function as

follows:

$$s(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{otherwise.} \end{cases} \quad (3.8)$$

Another extension to the original operator is the uniform patterns [100, 101], which can be used to reduce the length of the feature vector and implement a simple rotation-invariant descriptor. This extension was inspired by the fact that some binary patterns occur more commonly in texture images than others. A local binary pattern is called uniform if the binary pattern contains at most two binary transitions (0-1 or 1-0). For example, 01000000 (2 transitions) is a uniform pattern, 01011001 (5 transitions) is not [102]. The number of LBP histogram bins can be calculated based on this formula:

$$N_{bins} = P \cdot (P - 1) + 3 \quad (3.9)$$

where P is the number of neighborhood pixels around the central pixel. In the computation of the LBP histogram, the histogram has a separate bin for every uniform pattern, and all non-uniform patterns are assigned to a single bin which is 59. Using uniform patterns, the length of the feature vector for a 3x3 window reduces from 256 (2^8) to 59. Also, the length of the feature vector for a 5x5 window is reduced from 65536 (2^{16}) to 243.

3.3.1.2 Local Directional Pattern

LDP is an eight-bit binary code assigned to each pixel of an input grayscale image. LDP pattern is calculated by comparing the relative edge response value of a pixel in different directions. The eight directional edge response values of a particular pixel are calculated using Kirsch masks in eight different orientations ($M_0 - M_7$) centered on its own position [103, 104]. These masks are shown in Figure 3.8.

$$\begin{bmatrix} -3 & -3 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & 5 \end{bmatrix} M_0(\uparrow) \quad \begin{bmatrix} -3 & 5 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & -3 \end{bmatrix} M_1(\searrow) \quad \begin{bmatrix} 5 & 5 & 5 \\ -3 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix} M_2(\leftarrow) \quad \begin{bmatrix} 5 & 5 & -3 \\ 5 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix} M_3(\swarrow)$$

$$\begin{bmatrix} 5 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & -3 & -3 \end{bmatrix} M_4(\downarrow) \quad \begin{bmatrix} -3 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & 5 & -3 \end{bmatrix} M_5(\nwarrow) \quad \begin{bmatrix} -3 & -3 & -3 \\ -3 & 0 & -3 \\ 5 & 5 & 5 \end{bmatrix} M_6(\rightarrow) \quad \begin{bmatrix} -3 & -3 & -3 \\ -3 & 0 & 5 \\ -3 & 5 & 5 \end{bmatrix} M_7(\nearrow)$$

Figure 3.8: Kirsch compass kernel.

By applying eight masks, eight edge response values will be obtained m_0, m_1, \dots, m_7 , each one represents the edge significance in its respective direction. The response values are not equally important in all directions. In order to generate the LDP codewords, a k value must be given. Then, the top k values of $|m_j|$ are set to 1, and the rest $8 - k$ values of $|m_j|$ are set to 0. LDP code for each pixel is calculated using the formulas below:

$$LDP_k = \sum_{i=0}^7 b_i(m_i - m_k) \cdot 2^i \quad (3.10)$$

$$b_i(a) = \begin{cases} 1 & \text{if } a \geq 0 \\ 0 & \text{otherwise.} \end{cases} \quad (3.11)$$

where m_k is the k -th most significant directional response. After computing the LDP code for each pixel (r, c) , the LDP histogram H of the image I is represented using this equation:

$$H(\tau) = \sum_{r=1}^M \sum_{c=1}^N f(LDP_k(r, c), \tau) \quad (3.12)$$

where τ is the LDP code value. The number of LDP histogram bins is calculated as follow:

$$N_{bins} = \frac{8!}{k! \cdot (8 - k)!} \quad (3.13)$$

A new variant of LDP called LDPv has been introduced in [105] where the variance σ is introduced as an adaptive weight to adjust the contribution of the LDP code in the histogram generation. The new histogram of LDPv is calculated using the formulas below:

$$LDPv(\tau) = \sum_{r=1}^M \sum_{c=1}^N w(LDP_k(r, c), \tau) \quad (3.14)$$

$$w(LDP_k(r, c), \tau) = \begin{cases} \sigma(LDP_k(r, c)) & \text{if } LDP_k(r, c) = \tau \\ 0 & \text{otherwise.} \end{cases} \quad (3.15)$$

$$\sigma(LDP_k(r, c)) = \frac{1}{8} \sum_{i=0}^7 (m_i - \bar{m})^2 \quad (3.16)$$

where \bar{m} is the average of all directional responses m_i calculated for a pixel (r, c) .

3.3.1.3 Local Phase Quantization

New texture descriptor called LPQ was proposed in [106]. It is based on the application of Short Term Fourier Transform (STFT). The advantage in STFT is that the phase of the low frequency coefficients is insensitive to centrally symmetric blur. The spatial blurring is represented by a convolution between the image intensity and a Point Spread Function (PSF). The LPQ descriptor uses the local phase information extracted by the 2-D Discrete Fourier Transform (DFT) or, more precisely, a STFT computed over a rectangular $M - b$ by $M - M$ neighborhood N_x at each pixel position x of the image $f(x)$ defined by this formula:

$$F(u, x) = \sum_{y \in N_x} f(x - y) e^{-j2\pi u^T y} = w_u^T f_x \quad (3.17)$$

where w_u is the basis vector of the 2-D DFT at frequency u , and f_x is another vector containing all M^2 image samples from N_x [107]. The local Fourier coefficients are computed at four frequency points $u_1 = [a, 0]^T$, $u_2 = [0, a]^T$, $u_3 = [a, a]^T$, and $u_4 = [a, -a]^T$, where a is a scalar frequency below the first zero crossing of $H(u)$ that satisfies the condition $H(u_i) > 0$ [108]. So a vector obtained for each pixel, will be built like in this formula:

$$F_x = [F(u_1, x), F(u_2, x), F(u_3, x), F(u_4, x)] \quad (3.18)$$

The phase information in the Fourier coefficients is recorded by observing the signs of the real and imaginary parts of each component in $F(x)$. This is done by using a simple scalar quantization which presented in this formula:

$$q_j = \begin{cases} 1 & \text{if } g_j \geq 0 \\ 0 & \text{otherwise.} \end{cases} \quad (3.19)$$

where g_j is the j -th component of the vector $G(x) = [Re\{F(x)\}, Im\{F(x)\}]$. The resulting eight binary coefficients q_j represent the binary code pattern. This code will be converted to decimal number between 0-255. From that, the LPQ histogram must have 256 bins [108].

3.3.1.4 Binarized Statistical Image Features

BSIF is a new image texture descriptor proposed by Kannala and Rahtu [109]. It is inspired by LBP and LPQ texture descriptors. The idea behind BSIF is to automatically learn a fixed set of filters from a small set of natural images, instead of using hand-crafted filters such as in LBP

and LPQ. The set of filters is learnt from a training set of natural image patches by maximizing the statistical independence of the filter responses.

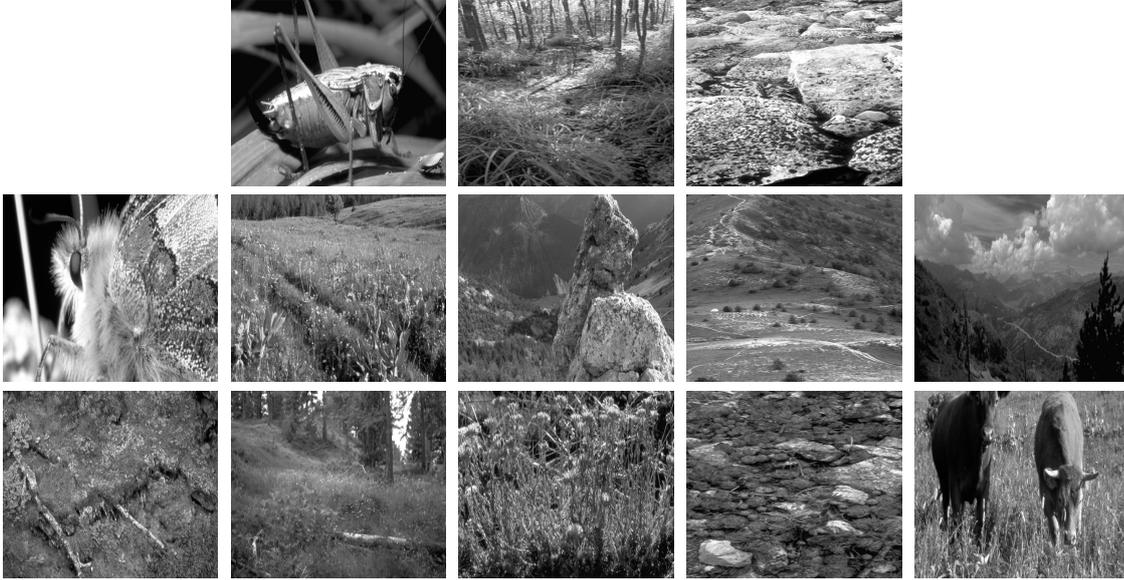


Figure 3.9: 13 natural images used for filters learning of BSIF descriptor [15].

Given an image patch X of size $l \cdot l$ pixels and a linear filter W_i of the same size, the filter response s_i is obtained by :

$$s_i = \sum_{i,v} W_i(i,v)X(u,v) = w_i^T x, \quad (3.20)$$

where vectors w and x contain the pixels of W_i and X . The binarized feature b_i is obtained by:

$$b_i = \begin{cases} 1 & \text{if } s_i > 0 \\ 0 & \text{otherwise.} \end{cases} \quad (3.21)$$

The filters W_i are learnt using Independent Component Analysis (ICA) by maximizing the statistical independence of s_i . The same filters learnt from a set of 13 natural images (see Figure 3.9). The number of histogram bins obtained by BSIF descriptor is calculated using this formula:

$$N_{bins} = 2^{N_f} \quad (3.22)$$

where N_f is the number of filters used to obtain the BSIF code.

3.3.2 PML face representations

The most common face representation in computer vision is a regular grid of fixed size regions which we call it MB representation. MB face representation divides the image into n^2 blocks where n is the intended level of MB. Figure 3.10 illustrates the MB face representations.

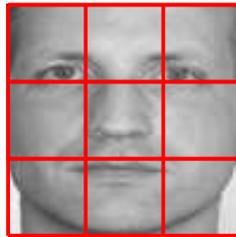


Figure 3.10: MB face representation (level 3).

Recently, a similar representation called ML representation used in the age estimation and gender classification topics [6, 110]. ML face representation is a spatial pyramid representation which constructed by sorted series of MB representations. The ML face representation level n is constructed from level 1, 2, ..., n MB face representations. Figure 3.11 illustrates the ML face representations.

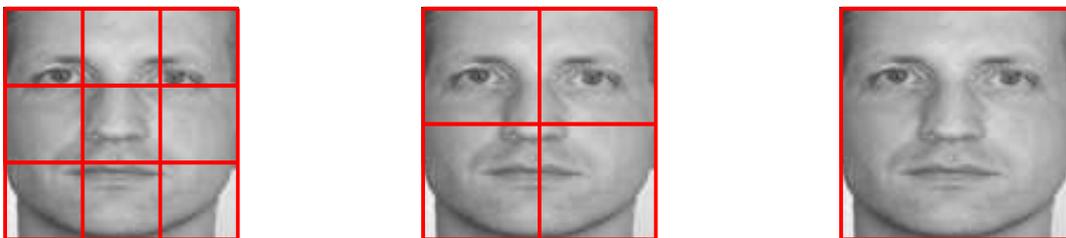


Figure 3.11: ML face representation (level 3).

Inspired by ML [6, 110, 111] and Pyramid-Based Multi-Scale [112] representations, we present the PML [14] representation of the face image in order to extract the local texture features using different descriptors. Unlike, the ML representation, the PML representation adopts an explicit pyramid representation of the original image. This pyramid represents the image at different scales. For each such level or scale, a corresponding MB representation is used. PML sub-blocks have the same size which determined by the image size and the chosen level. Figure 3.12 illustrates the PML face representations.

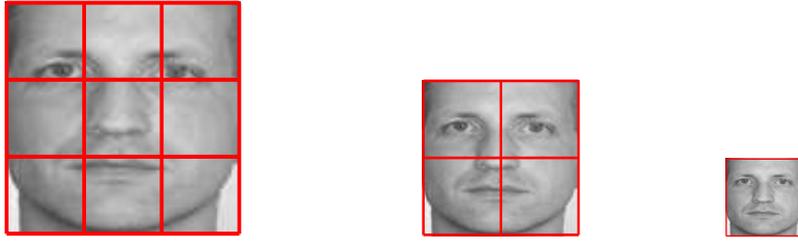


Figure 3.12: PML face representation (level 3).

The main idea of the PML is to extract features from different division of each pyramid level. In other words, given n , w , h as PML level, original image width and original image height respectively, we assume that the original face ROI corresponds to level n , the pyramid levels (n , $n - 1$, $n - 2$, ..., 1) are built as follows. The image of level $n - 1$ is obtained from image at level n by resizing it to $(w' = w \cdot \frac{(n-1)}{n}, h' = h \cdot \frac{(n-1)}{n})$. At each level n , the obtained image is divided into n^2 blocks. Figure 3.13 illustrates the PML principle for a pyramid of four levels. This figure corresponds to the Local Phase Quantization (LPQ) descriptor.

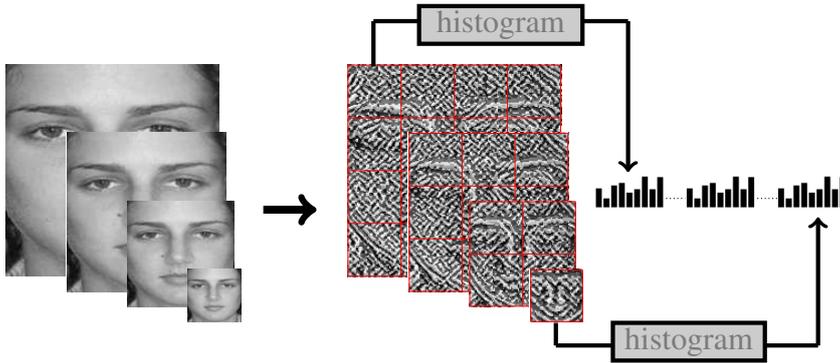


Figure 3.13: Pyramid Multi-Level Local Phase Quantization example of level 4.

3.3.3 Dimensionality reduction

Due to the huge number of features obtained from the different face representation, we propose to reduce the number of features using feature selection method based on Fisher's score, which quantifies the discriminating power of features. This score is given by [113]:

$$W_i = \frac{\sum_{j=1}^C N_j \cdot (m_j - \bar{m})^2}{\sum_{j=1}^C N_j \cdot \sigma_j^2} \quad (3.23)$$

where W_i is the score of feature i , C is the number of classes, N_j is the number of samples in class j , \bar{m} is the feature mean, m_j and σ_j^2 are the mean and the variance of the class j in intended feature.

The features are sorted according to the value of their scores, i.e. the sorting is carried out by descending order. After sorting the feature by the weight, we select the first 5% of the features for each demographic attribute since empirically this was found to be as a good choice for the trade-off efficiency-accuracy. This choice was made after we conducted some experiments which are presented in Section 4.5.

We point out that Fisher Score is computed independently for each demographic attribute. The finally selected features are 5% for each demographic attribute. In the case of age, each class is an interval of 5 years without overlapping. We found that the separate use of selected subset of features (by the demographic attribute estimator) was found to be more accurate than using all selected subsets of features [14].

We apply then a Min-Max scaling method for feature normalization on the subset of the training using the selected features. This method rescales the range of features between [0, 1]. The general formula of the Min-Max scaling is given by:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (3.24)$$

3.4 Hierarchical demographic estimation

The last stage in our proposed approach is the hierarchical demographic estimator where the decision for the demographic traits will be made. As we know, there are three demographic attributes which are age, gender, and ethnicity. Each one of these attributes has its own classification problem. Starting with ethnicity attribute which is a multi-class problem, then the gender attribute which is a binary classification where we have only male or female as an output result. So, gender and ethnicity estimation are classification problems requiring a classifier for each one. On the other hand, the age estimation task can be formulated as a regression or classification problem.

Our proposed approach consists of estimating ethnicity firstly, then gender, after that estimating age based on the estimated ethnicity and gender values. We choose ethnicity as root based on the fact that, according to our experiments, ethnicity attribute estimation has the highest accuracy among all three demographic attributes. In the case of the absence of ethnicity, gender will become the root of the demographic estimator. Figure 3.14 illustrates the adopted hierarchical demographic estimator.

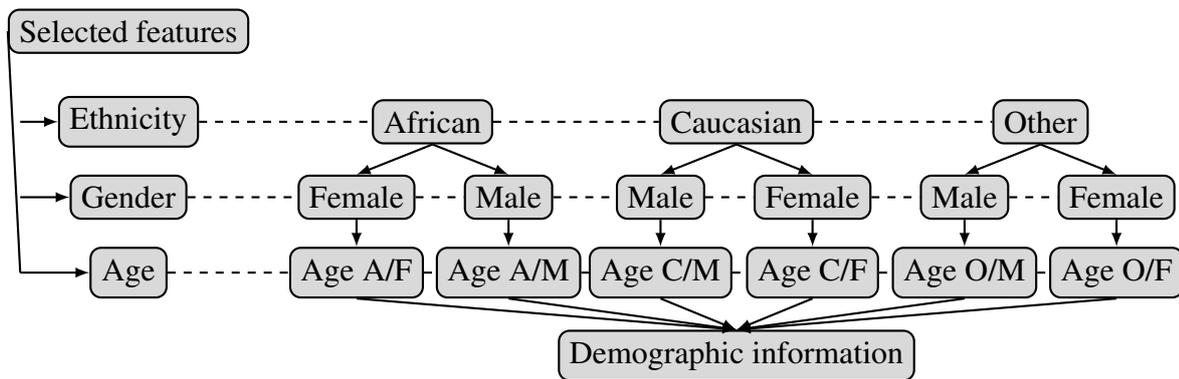


Figure 3.14: The proposed hierarchical demographic estimator.

In the figure above, the age layer can be implemented using either SVM for classification or SVR for estimation. SVMs are supervised learning models with associated learning algorithms that analyze data use for classification and regression analysis a support vector machine constructs a hyperplane or set of hyperplanes in a high- or infinite-dimensional space, which can be used for classification, regression, or other tasks. Intuitively, a good separation is achieved by the hyperplane that has the largest distance to the nearest training-data point of any class (so-called functional margin), since in general the larger the margin the lower the generalization error of the classifier [114].

The demographic estimation was performed using LIBLINEAR [115] which is a package for large-scale regularized linear classification and regression. Tuning the best parameters of the classifier or the regressor was based on grid search strategy combined with cross-validation.

3.5 Conclusion

In this chapter, we have presented a description of our proposed approach. This latter is consisting of face preprocessing, feature extraction, and demographic estimator. Our face preprocessing handles frontal and near-frontal faces easily which help the system to maintain the important information of the face. In the feature extraction stage, the proposed PML face presentation has been illustrated along with the four image texture descriptors used in our work. We finished with the demographic estimation stage where we presented the proposed hierarchical estimator layer by layer.

RESULTS AND DISCUSSION

Contents

	Page
4.1 Introduction	52
4.2 Settings	52
4.3 Effect of the face preprocessing	53
4.4 Effect of the face representation on ethnicity classification	56
4.5 Effect of features ratio on the CPU time	62
4.6 Effect of demographic attributes on each other	63
4.7 Single database evaluation	65
4.7.1 MORPH II	65
4.7.2 PAL	68
4.7.3 IoG	70
4.7.4 LFW	71
4.7.5 FERET	72
4.8 Challenge database evaluation	73
4.9 Cross-database evaluation	76
4.9.1 Age estimation	77
4.9.2 Gender classification	77
4.9.3 Ethnicity classification	78
4.10 Real-time application	78
4.11 Conclusion	79

4.1 Introduction

This chapter contains the description of all databases and its protocols used in our work. Also, the obtained results of our experiments are given in details. These experiments include single database, cross database, challenge database, effect of demographic attributes on each other, effect of features ratio on CPU time, effect of features ratio on classification rates, and the effect of the layers order of the proposed hierarchical estimator. Many discussions have been given concerning each experiment. A comparison with the state of the art is also given.

4.2 Settings

The experimental settings for the five publicly available databases is explained below.

- MORPH II: Following the BeFIT¹ recommendations, we use 5-fold cross-validation evaluation, the folds are selected in such a way to prevent algorithms from learning the identity of the persons in the training set by making sure that all images of individual subjects are only in one fold at a time. In addition to the distribution of age, gender and ethnicity distributions in the folds are similar to the distributions in the whole database.
- PAL: For the evaluation of the approach, we conduct 5-fold cross-validation, our distribution of folds is selected based on age, gender and ethnicity. Knowing that we took into account the variation of the face expressions.
- IoG: We perform evaluations using the entire database with 5-fold cross-validation testing scheme. In this case, the first layer of the hierarchical estimator is the gender, and the second layer is age. Also, the age layer use SVM instead of SVR.
- LFW: For evaluating our approach, we follow the BeFIT benchmarking protocol by conducting a 5-fold cross-validation scheme. We do not use the hierarchical estimator in this case due to the absence of age and ethnicity attributes for this database.
- FERET: To evaluate our approach, we use only the fa subset (regular frontal image) and the fb subset (alternative frontal image) that contain 2722 images of 994 subjects. in this case, the demographic attributes are ethnicity and gender. We adopted a 5-fold cross-validation testing scheme in our experiments for FERET database.

¹<http://fipa.cs.kit.edu/425.php>

4.3 Effect of the face preprocessing

To tackle the problem of preprocessing, we conducted some experiments using the 2D proposed face preprocessing and two other 3D face preprocessing approaches. These 3D face preprocessing approaches are based on the work of Hassner *et al.* [116]. The 3D face preprocessing used for alignment of out-plane rotation in order to get a frontal appearance of a face in which out-of-plane rotation is canceled. The set of 2D points are provided by the work of Kazemi and Sullivan [96]. This set of 2D points is then used by the efficient 3D frontalization method described in [116], this method has the possibility of using the soft symmetry. The results of our experiments are summarized in Table 4.1.

Table 4.1: Results of demographic estimation on four databases with three scenarios: 2D face alignment, 3D face alignment without symmetry, and 3D face alignment with face symmetry.

Database	Face	Age		Gender	Ethnicity
		MAE	ACC		
PAL	2D ^a	5.0	NA	98.2%	99.3%
	3D ^b	4.9	NA	98.3%	99.3%
	3D Sym ^c	5.3	NA	97.9%	99.1%
MORPH II	2D	3.5	NA	98.9%	99.3%
	3D	3.6	NA	98.5%	99.1%
	3D Sym	4.0	NA	97.5%	98.6%
IoG	2D	NA	68.2%	90.0%	NA
	3D	NA	68.2%	90.1%	NA
	3D Sym	NA	68.4%	90.3%	NA
Chalearn	2D	0.2874 ^d		NA	NA
	3D	0.2905 ^d		NA	NA
	3D Sym	0.2853 ^d		NA	NA

(a) - 2D alignment.

(b) - 3D frontalization without soft symmetry.

(c) - 3D frontalization with soft symmetry.

(d) - Instead of MAE, the EPS score is used.

From these results, we observe a slight improvement on PAL, on IoG, and on Chalearn as expected. At the same time, there is a slight deterioration on MORPH II. There are several plausible explanations for this deterioration such as the images quality and the 3D model (Hassner *et al.* [116]) used which can introduce some artifacts to the frontal face textures. Soft symmetry was good process for IoG and Chalearn but was not for PAL and MORPH II. From the results, we can say that the extra computation needed by the 3D frontalization method (both with soft symmetry and without it) have not led to a significant improvement in age,

gender, and ethnicity estimation. Furthermore, for Chalearn datasets which contains faces with significant out-of-plane rotation, seems to benefit very slightly from 3D face frontalization.

We included some examples of the cropped images for the three cases 2D alignment, 3D frontalization without soft symmetry and 3D frontalization with soft symmetry. Figures 4.1, 4.2, 4.3, and 4.4 show that.



Figure 4.1: Aligned face images from PAL database. The upper part represent the 2D face alignment (a 2D similarity transform). The middle part represent the 3D frontalization without symmetry .The lower part represent the out-plane rotation.



Figure 4.2: Aligned face images from MORPH II database. The upper part represent the 2D face alignment (a 2D similarity transform). The middle part represent the 3D frontalization without symmetry .The lower part represent the out-plane rotation.



Figure 4.3: Aligned face images from IoG database. The upper part represent the 2D face alignment (a 2D similarity transform). The middle part represent the 3D frontalization without symmetry .The lower part represent the out-plane rotation.

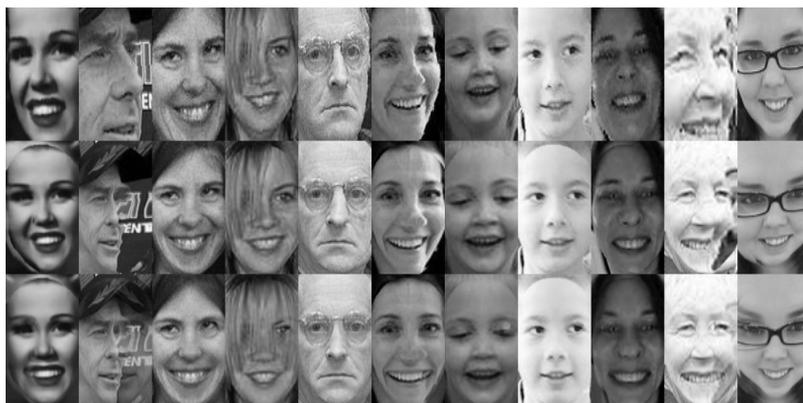


Figure 4.4: Aligned face images from Chalearn database. The upper part represent the 2D face alignment (a 2D similarity transform). The middle part represent the 3D frontalization without symmetry .The lower part represent the out-plane rotation.

It is worth knowing that the CPU time for 2D alignment is small, it is also small for some 3D face frontalization methods. For that we choose a 3D face frontalization method which needs $\approx 30\text{ms}$ for each face (2D face alignment needs $\approx 10\text{ms}$ for each face) to preserve the ability to work on real time application. The used method for 3D face frontalization is efficient. This method succeeds in providing a plausible face appearance that cancel the out-of-plane rotation. On the other hand, this method could not be considered as the most accurate method for generating a 3D frontal face that preserves the person specific shape since a common 3D surface is used.

4.4 Effect of the face representation on ethnicity classification

classification

To study the effect of the face representation, the number of levels used and the feature ratio used on the ethnicity classification we took 2000 images from MORPH II database, 500 images from FERET, and 500 images from PAL database to create a new collection of images, this collection distributed equally by age, gender and ethnicity. We conduct 12 experiments on this collection starting by using MB face representation which is applied on LBP, LDP, LPQ, and BSIF images. Figures 4.5, 4.6, 4.7, and 4.8 show the results of 10 different levels (1-10) for MB face representation and features ratio between 10% and 100%. Then, the same thing for ML and PML face representations which is applied on the latter four image texture descriptors. Figures 4.9, 4.10, 4.11, and 4.12 show the results when using ML face representation. For PML face representation, the results are shown in the figures 4.13, 4.14, 4.15, and 4.16.

From all these experiments, we can conclude that increasing the features number make the system more accurate. Also, we can observe that the PML face representation overcome both MB and ML face representations in term of ethnicity classification rates.

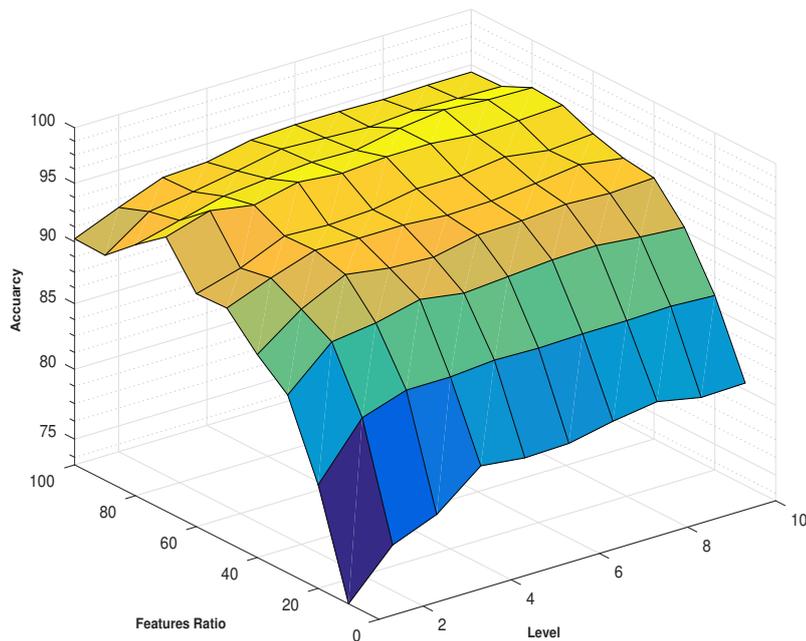


Figure 4.5: Ethnicity classification results using LBP descriptor and MB face representation.

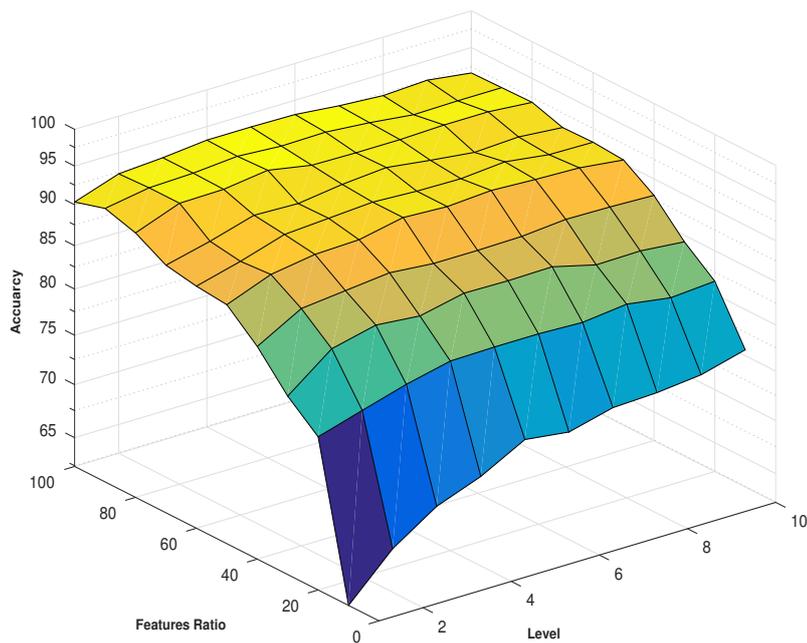


Figure 4.6: Ethnicity classification results using LDP descriptor and MB face representation.

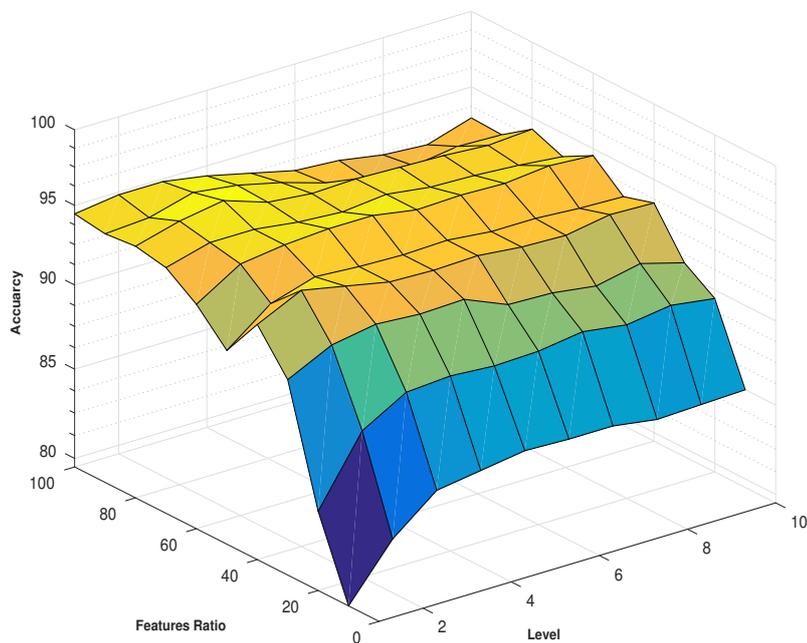


Figure 4.7: Ethnicity classification results using LPQ descriptor and MB face representation.

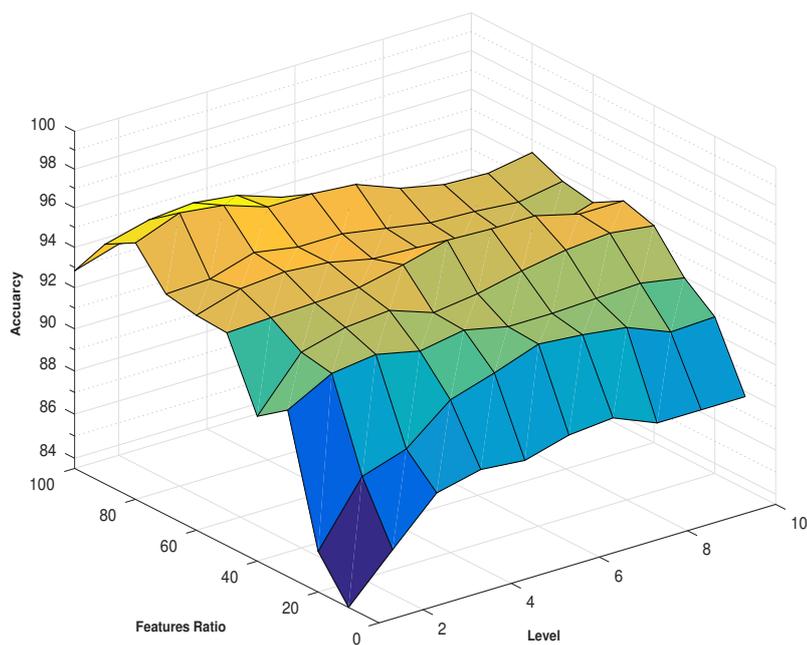


Figure 4.8: Ethnicity classification results using BSIF descriptor and MB face representation.

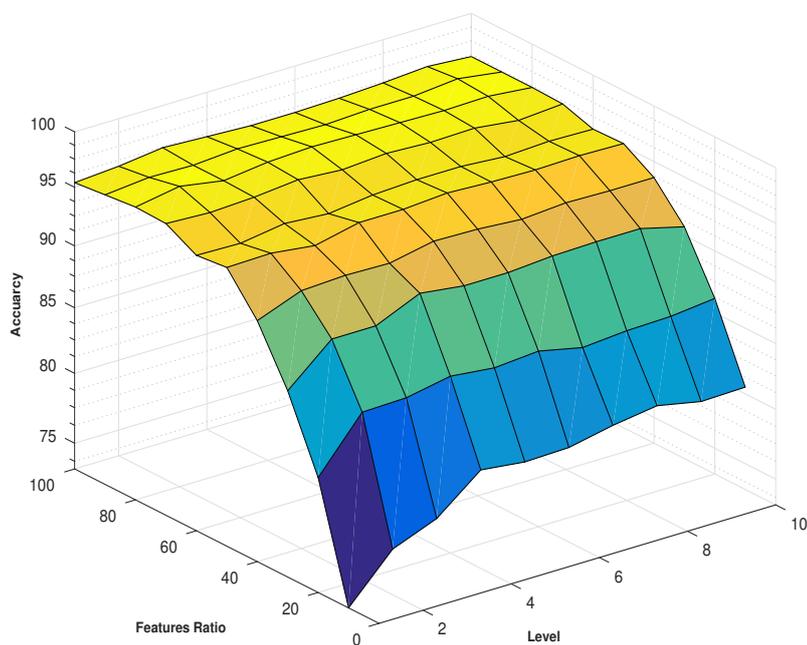


Figure 4.9: Ethnicity classification results using LBP descriptor and ML face representation.

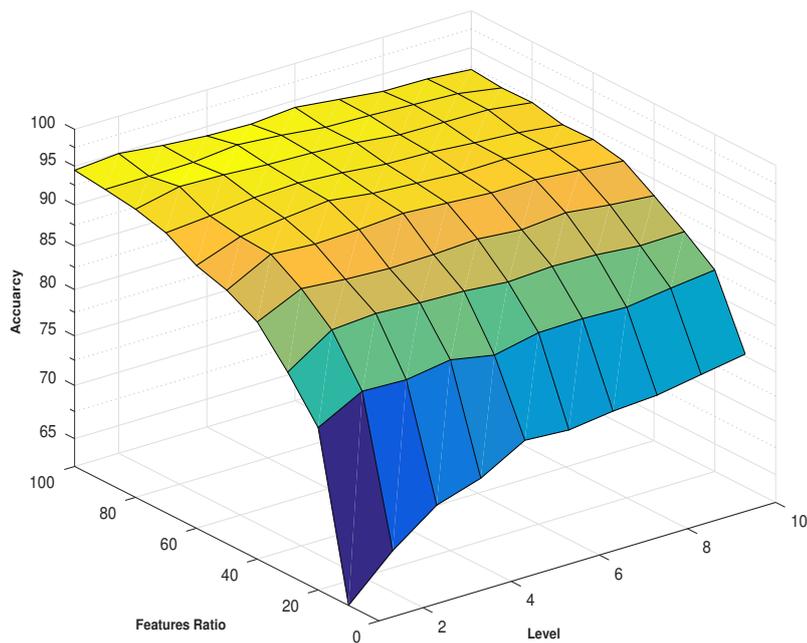


Figure 4.10: Ethnicity classification results using LDP descriptor and ML face representation.

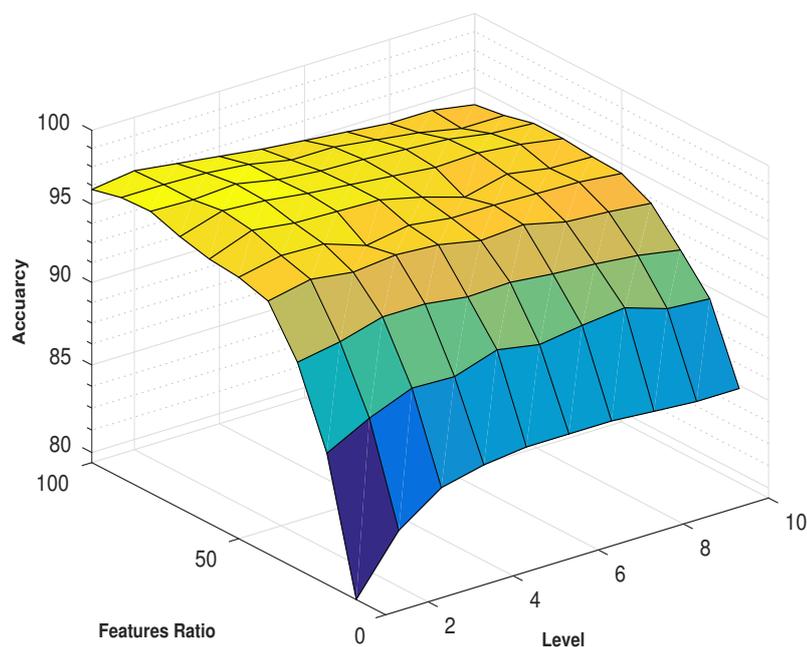


Figure 4.11: Ethnicity classification results using LPQ descriptor and ML face representation.

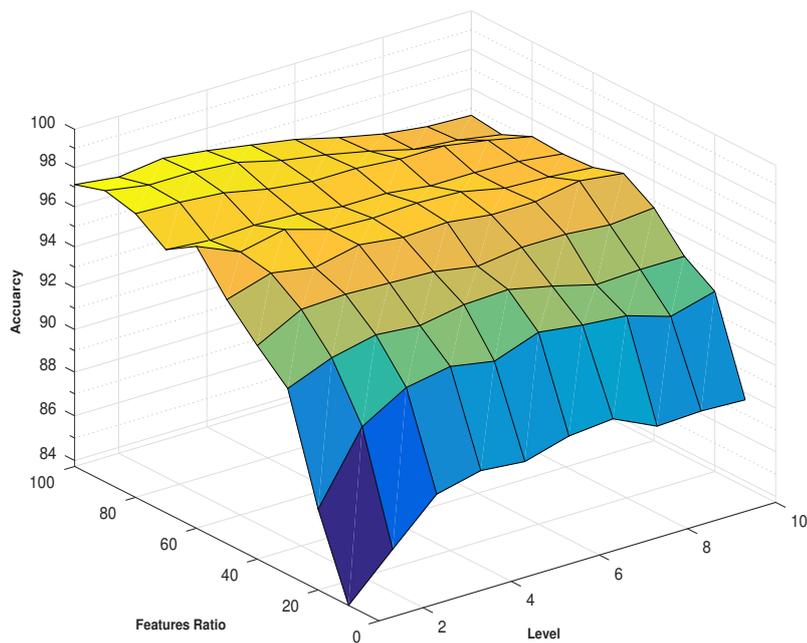


Figure 4.12: Ethnicity classification results using BSIF descriptor and ML face representation.

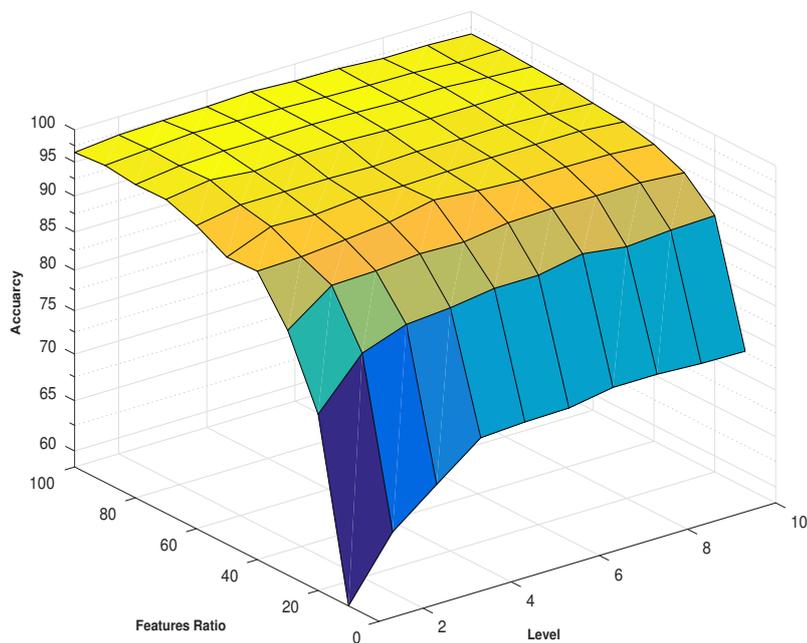


Figure 4.13: Ethnicity classification results using LBP descriptor and PML face representation.

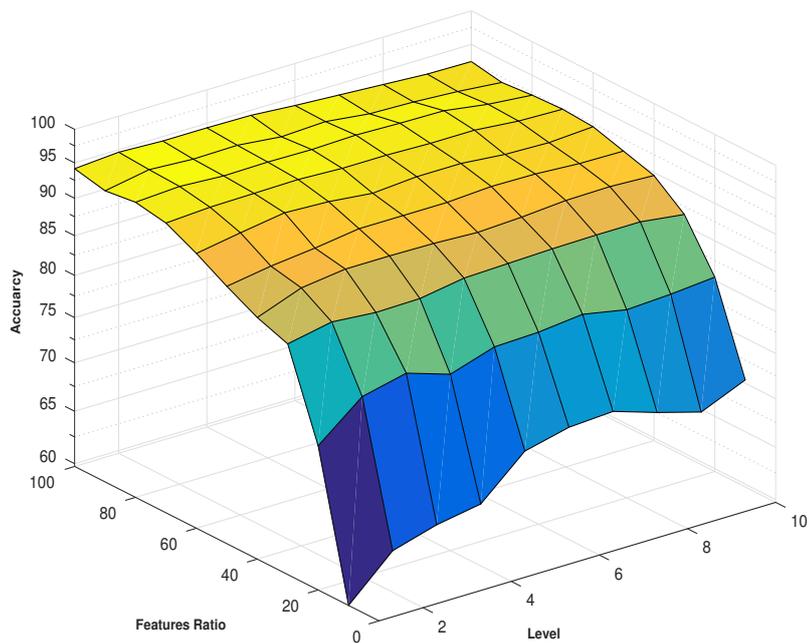


Figure 4.14: Ethnicity classification results using LDP descriptor and PML face representation.

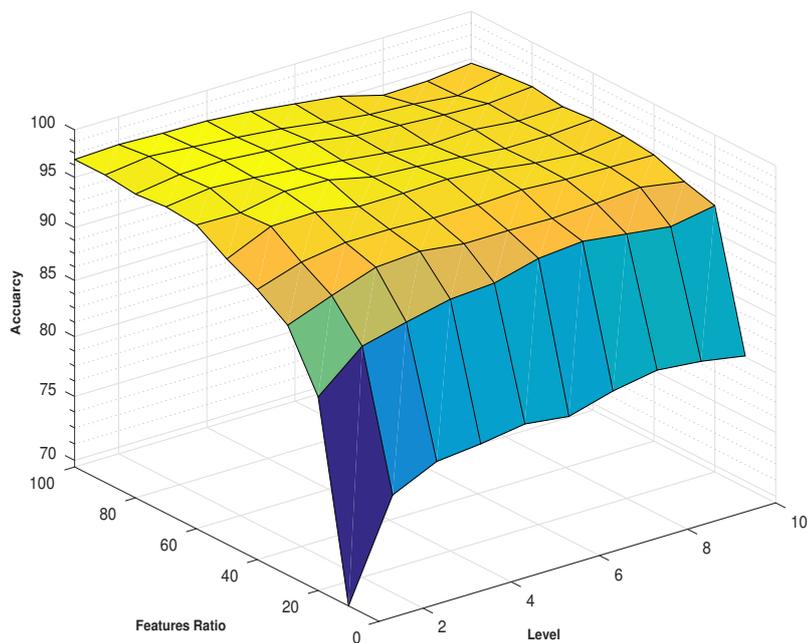


Figure 4.15: Ethnicity classification results using LPQ descriptor and PML face representation.

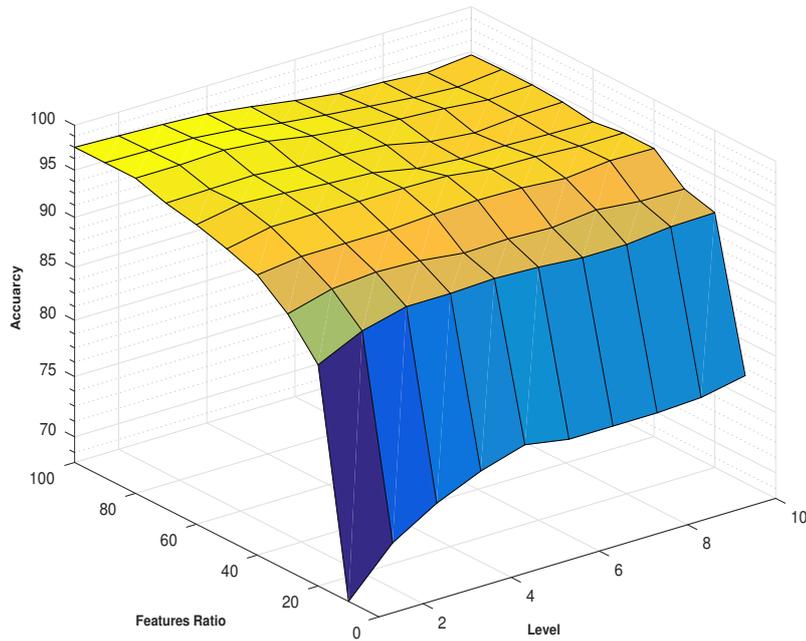


Figure 4.16: Ethnicity classification results using BSIF descriptor and PML face representation.

4.5 Effect of features ratio on the CPU time

In this experiment, we used only the first four folds of the PAL database for training and the last fold for test. Table 4.2 summarizes the obtained results that justify our choice.

Table 4.2: Accuracy of demographic estimation as a function of different features ratios. The last row depicts the CPU time (in seconds) of the training phase as a function of different features ratios.

Features ratio (%)	1%	5%	10%	20%	50%	60%	75%	100%
Age (MAE in years)	6.5	5.0	5.0	5.0	4.8	4.8	4.7	4.7
Gender (%)	93.7%	98.2%	98.5%	98.9%	99.0%	99.0%	99.1%	99.2%
Ethnicity (%)	95.0%	99.3%	99.4%	99.4%	99.5%	99.5%	99.5%	99.5%
CPU time (s)	49	77	91	108	134	142	165	189

In this table, we show the accuracy of estimating the three demographic attributes (age, gender, and ethnicity) as a function of features ratio (after Fisher score-based ranking). We choose 7 level PML the efficiency-accuracy trade-off of proposed approach based on our experiments. The features correspond to a 7 level PML (BSIF+LPQ) which has 280 blocks, each with 256 bins. Thus, we have 71680 features in total. In all experiments, the adopted PML has 7 levels.

The last row of the Table 4.2 illustrates the CPU time (in seconds) needed for training the model. This CPU time corresponds to the feature selection process and to the training of the demographic attribute classifiers. As can be seen, the choice of 5% can be considered as a good trade-off between accuracy and computational complexity. Indeed, if only 5% of the features are used the MAE increases 0.3 years, the classification accuracy of gender drops by 1%, and that of ethnicity by 0.2%. At the same time, the CPU time of the training phase decreased from 189 s to 77 s.

Table 4.3 depicts the CPU time of the testing phase associated with one image. This CPU time corresponds to the estimation part. Besides the results associated with the training phase, the test results as well justify our choice of 5% of the selected features. Indeed, the CPU time of 7 ms per image enables the system to be used in real-time applications. Our test was done on a laptop with 4 core Intel Xeon Processor E3-1535M v5 (8M Cache, 2.90 GHz) and 64GB RAM [14].

Table 4.3: CPU time associated with the testing phase as a function of different features ratios.

Features ratio (%)	1%	5%	10%	20%	50%	60%	75%	100%
CPU time (ms)	3	7	14	23	49	58	71	92

4.6 Effect of demographic attributes on each other

In the proposed hierarchical estimator, the age is estimated after estimating the ethnicity and gender attributes. We have performed an experimental evaluation to investigate the effect of knowing one or more demographic attributes on the estimation of the other.

To this end, we split MORPH II database into six subsets that correspond to the six combinations gender-ethnicity. Then, a five-fold cross-validation is conducted in which the age of each test image is estimated using the corresponding SVR. Table 4.4 illustrates the MEA obtained by using the six SVRs. In this case, the ground-truth of ethnicity and gender attributes of the test image are used.

Table 4.4: Age estimation for different ethnicities and gender on MORPH II database. The ground-truth ethnicity and gender are used by the age estimator.

Gender	Ethnicity	Images number	Age (MAE)
Male	Black	36804	3.2
	Other	1843	3.5
	White	7999	3.0
Female	Black	5758	4.4
	Other	129	4.7
	White	2601	3.9

According to Table 4.6, the obtained MAE is 3.5 years. This corresponds to the use of the hierarchical estimator. However, when the ground-truth ethnicity and gender labels are used the same MAE becomes 3.4 years (see Table 4.4). Since we know the misclassified images in the ethnicity layer, it is possible to study the influence of these misclassified 408 images through the ethnicity classifier. To this end, we calculate the absolute age error for the case where the ground truth ethnicity is used for these 408 images. The corresponding MAE is 3.3 years. On the other side, where the full hierarchical system is used (i.e, the predicted ethnicity is propagated) the MAE associated with these 408 images is 5.2 years. Figure 4.17 shows the curve of the age error of the 408 misclassified images in both cases (ground-truth ethnicity and predicted ethnicity).

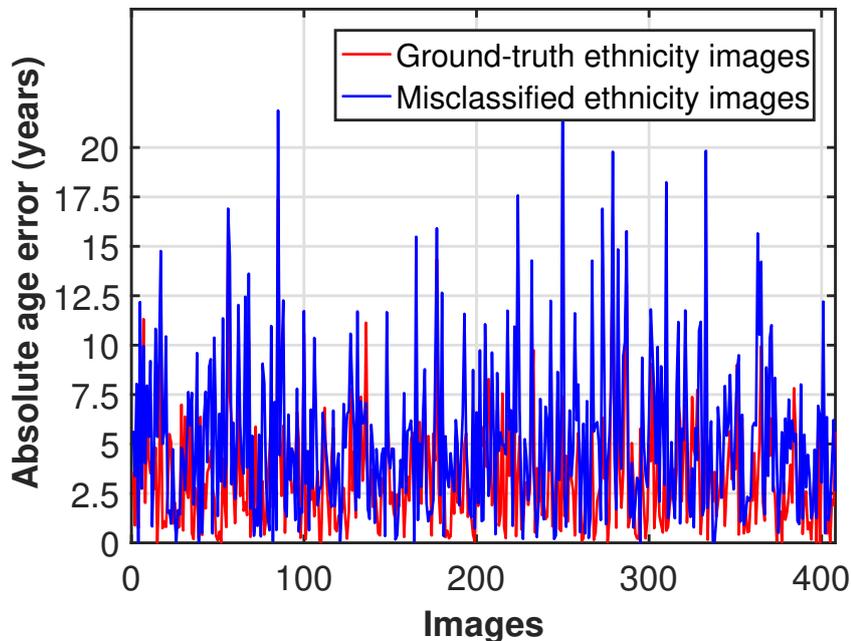


Figure 4.17: Absolute age error of misclassified ethnicity images and ground-truth ethnicity images.

We also studied the effect of misclassified images in the ethnicity layer on the gender layer. Table 4.5 shows the results of gender classification when we use the ethnicity layer of the hierarchical estimator and the ground-truth ethnicity. From these results, we can observe that the possible misclassification in the ethnicity layer can slightly deteriorate the performance of gender and age estimation.

Table 4.5: Gender classification using predicted and ground-truth ethnicities.

	Ground-truth ethnicity	Predicted ethnicity
Gender rate	99.2%	98.9%

4.7 Single database evaluation

4.7.1 MORPH II

Figure 4.18 presents the accuracy of gender classification obtained by different face representations (MB, ML and PML) and different texture descriptors (BSIF, LPQ and BSIF+LPQ). The accuracy is given as a function of the level. In this figure, the level is related to the face representation used. For example, when PML is used the level indicates the pyramid level. We can observe that the performance of the PML is better than MB and ML using the same descriptor.

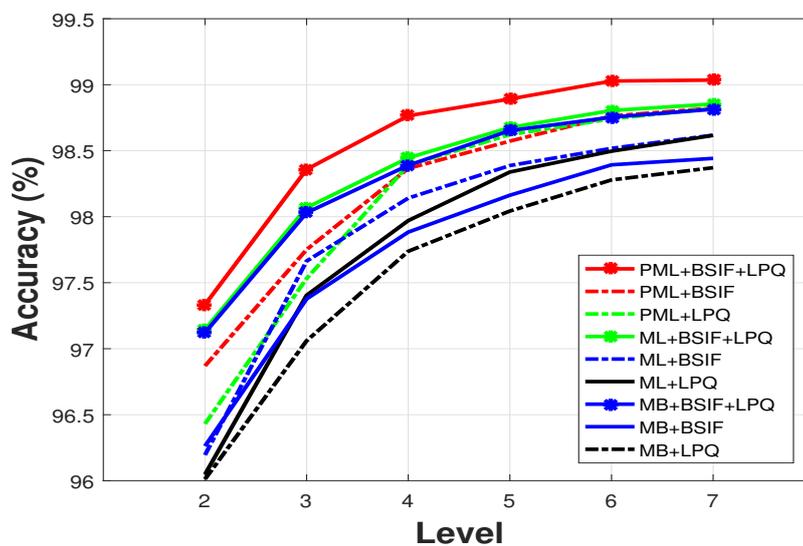


Figure 4.18: Comparison of different face representations using 7 levels in gender classification on MORPH II database.

Moreover, in Figure 4.19 the performance of the PML is better than both MB and ML using 7 levels for each face representation.

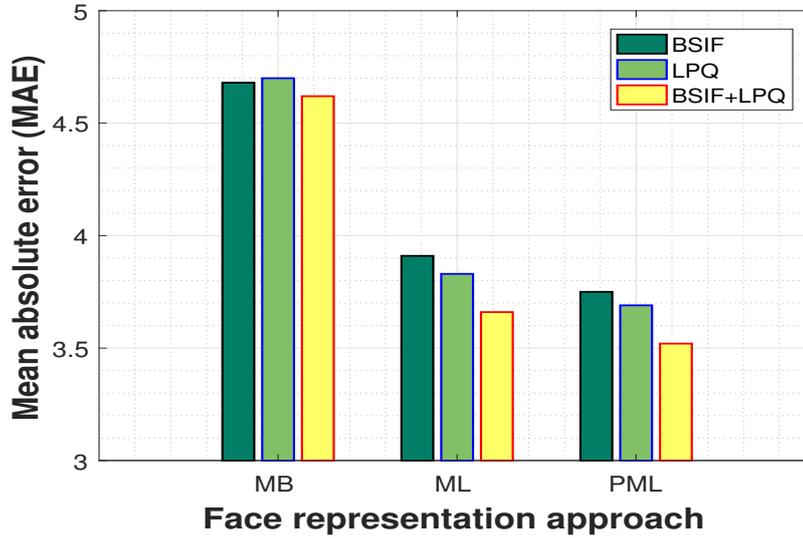


Figure 4.19: Comparison of different face representations in age estimation on MORPH II database.

Table 4.6 illustrates the age estimation performance of the proposed approach as well as of that of some competing approaches. In [51], they proposed a new classifier called Ordinal Hyperplanes Ranker (OHRank). They randomly split the data into 80% for training and 20% for testing over 30 trails. In [52], Conditional Probability Neural Network (CPNN) estimator was proposed which significantly improved their results. They conducted their experiments using the 10-fold cross validation strategy. In [14, 92, 117] used the 5-fold cross validation strategy either on training or validation.

Table 4.6: A comparison of the proposed approach with other demographic estimation approaches on MORPH II database.

Approach	Performance		
	Age (MAE)	Gender	Ethnicity
AAM+OHRank [51]	6.1	NA	NA
BIF+CPNN [52]	4.8	NA	NA
BIF+KPLS [50]	4.2	98.2%	98.9%
BIF+KCCA [88]	4.0	96.0%	98.9%
CNN [117]	3.9	NA	NA
BIF+DIF+SVM [92]	3.6	97.6%	99.1%
BSIF+LPQ+PML+SVM (ours) [14]	3.5	98.9%	99.3%

Figure 4.20 and 4.21 show the performance of the proposed approach on MORPH II. Figure 4.20 illustrates the correlation between the estimated ages and their ground-truth values.

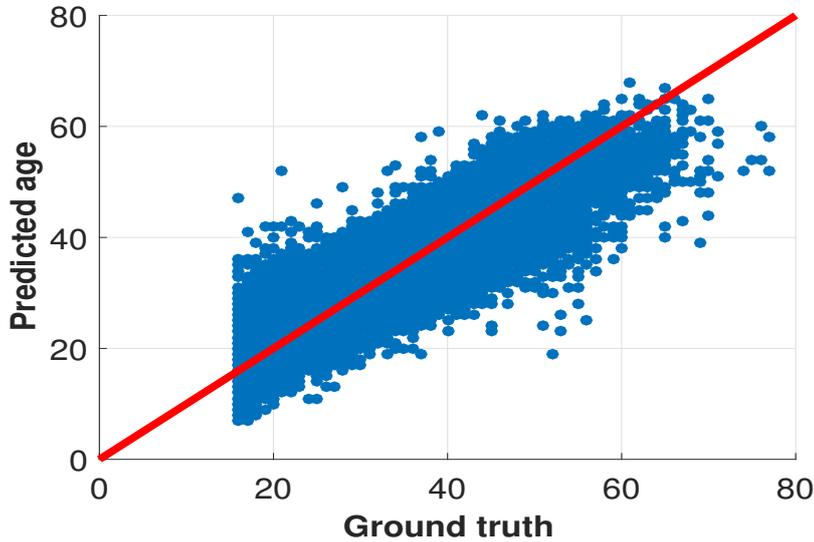


Figure 4.20: Correlations between True ages and Predicted ages by the proposed approach on MORPH II database.

Figure 4.21 illustrates the CS associated with three face descriptors.

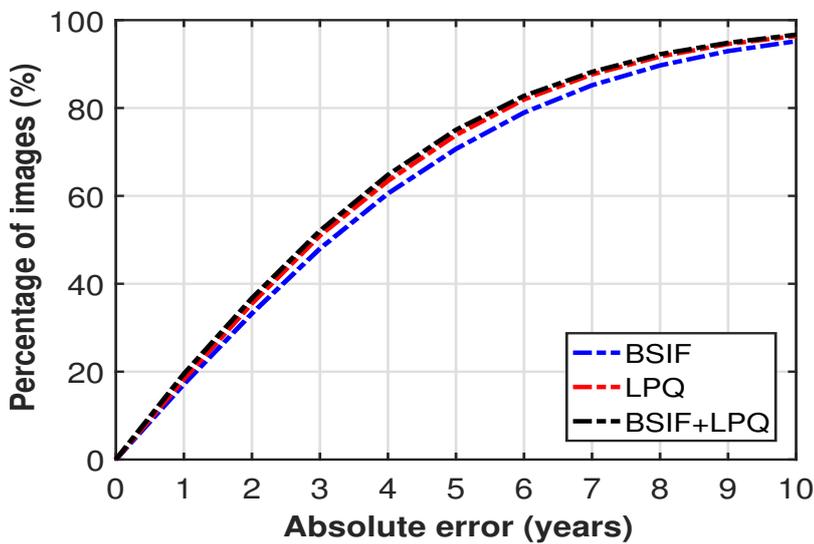


Figure 4.21: Cumulative scores obtained by the proposed approach on MORPH II database.

Figure 4.22 illustrates some examples for good and poor demographic estimation obtained by the proposed approach with MORPH II database. For age estimation, a poor estimation is declared for a test face if the absolute age error is equal to or greater than one year.



Figure 4.22: Examples of good and poor demographic estimation on MORPH II database.

4.7.2 PAL

Figure 4.23 shows the results of age estimation obtained with PAL dataset. The first row in the figure illustrates the correlation between the predicted ages and their ground truth values. The other row represents the cumulative score associated with three face descriptors.

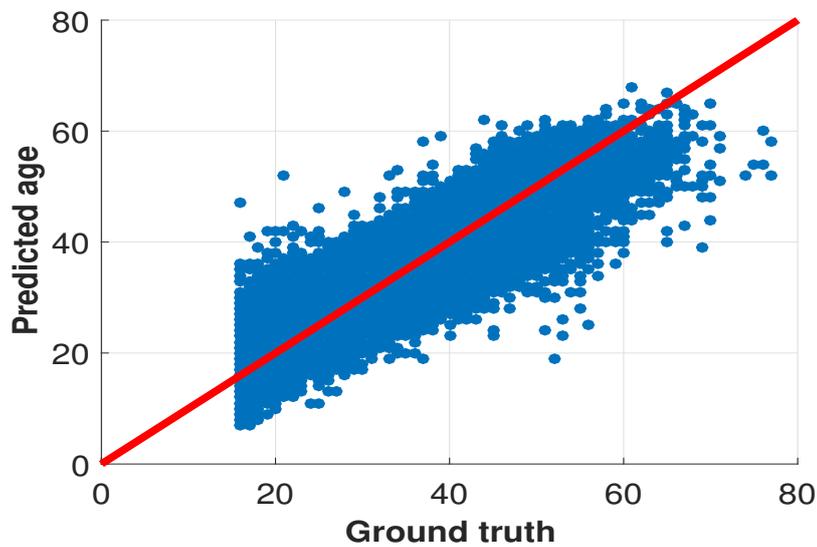


Figure 4.23: Correlations between True ages and Predicted ages by the proposed approach on PAL database.

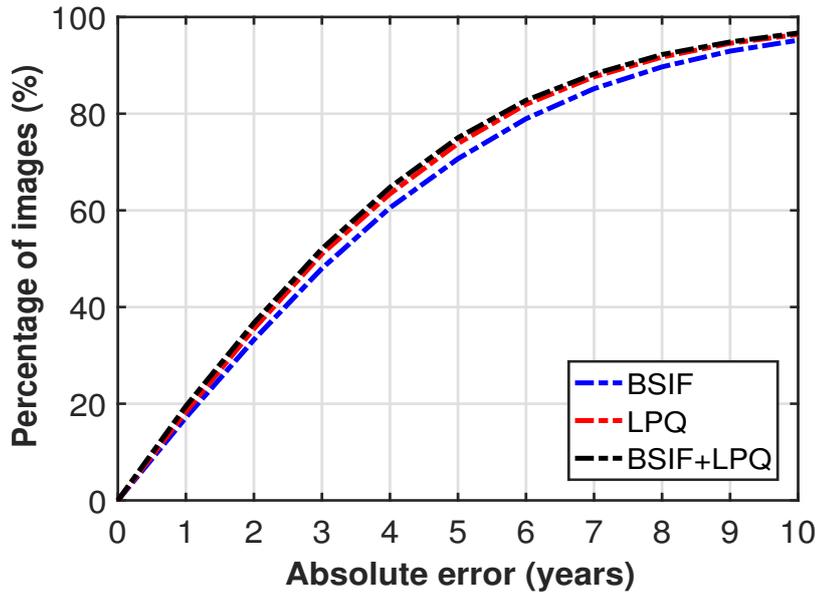


Figure 4.24: Age estimation performance by the proposed approach on PAL database.

The comparison of our approach with the state-of-the-art approaches is shown in Table 4.7. The authors of [118–120] use part of the PAL database which corresponds to the frontal images without expression. Furthermore, they conduct K-fold cross-validation testing scheme for evaluating the performance of age estimation. In [118], the Gaussian High Pass Filter (GHPF) has been proposed for age estimation. In [119], they used Contourlet Appearance Model (CAM) to extract features, and the classification strategy has not been mentioned. In [54, 121], the authors use the whole frontal database with the same scheme to evaluate the performance of age estimation. For gender classification, Patel et al. [79] conduct 5-fold cross-validation without mentioning the number of images used.

Table 4.7: A comparison of the proposed approach with other demographic estimation approaches on PAL database.

Approach	Performance		
	Age (MAE)	Gender	Ethnicity
GHPF+SVR [118]	8.4	NA	NA
CAM+SVR [119]	6.0	NA	NA
ML-LBP+GABOR+SVR [121]	6.5	NA	NA
LBP+BSIF+SVR [54]	6.2	NA	NA
MQLBP+SVM [79]	NA	95.5%	NA
AAM+GABOR+LBP [120]	5.4	NA	NA
BSIF+LPQ+PML+SVM (ours) [14]	5.0	98.2%	99.3%

Figure 4.25 shows some examples of good and poor demographic estimation in PAL database.



Figure 4.25: Examples of good and poor demographic estimation on PAL database.

4.7.3 IoG

Table 4.8 illustrates the confusion matrix associated with age group classification. Table 4.9 illustrates the age estimation performance of the proposed approach as well as of that of some competing approaches. In this table, the column Age+ considers the estimated age group as correct if it corresponds to either the ground-truth age group or to its nearest neighbor age group. Figure 4.26 illustrates some examples for good and poor demographic estimation obtained by the proposed approach with IoG database.

Table 4.8: Confusion matrix of age classification on IoG database.

P \ G	0-2	3-7	8-12	13-19	20-36	37-65	66+
0-2	712	170	10	2	9	0	0
3-7	177	959	304	57	21	11	0
8-12	2	36	53	14	8	1	0
13-19	0	13	23	40	10	2	0
20-36	52	375	439	1497	13852	3359	151
37-65	10	38	41	77	1138	3384	862
66+	1	4	2	5	10	60	240

P Predicted groups
 G Ground-truth groups

Table 4.9: A comparison of the proposed approach with other demographic estimation approaches on IoG database.

Approach	Performance		
	Age	Age+	Gender
Appearance+Context [28]	42.9%	78.1%	74.1%
boosted LBP+SVM [46]	55.9%	87.7%	77.4%
CLBP_M +SVM	51.7%	88.7%	NA
Gabor+Preserving Locality (PLO) [77]	48.5%	88.0%	NA
BIF+SVM [90]	68.1%	NA	87.1%
LBP+FPLBP [29]	66.6%	95.3 %	88.6%
LBP+CH+SIFT [47]	63.0%	NA	91.6%
ML-LBP+SVM [111]	55.1%	88.2%	78.3%
Best Local-DNN [122]	NA	NA	90.6%
BSIF+LPQ+PML+SVM (ours) [14]	68.2%	95.2%	90.0%



Figure 4.26: Examples of good and poor demographic estimation on IoG database.

4.7.4 LFW

The comparison with the state-of-the-art approaches is reported in Table 4.10. Most of the authors follow the BeFIT benchmarking protocol by conducting a 5-fold cross-validation scheme. We include some images with their prediction of gender in Figure 4.27.

Table 4.10: A comparison of the proposed approach with other gender classification approaches on LFW database.

Approach	Performance Gender
LBP+PCA+SVM [123]	97.2%
LBP+SVM [124]	94.8%
Fusion [125]	98.0%
BIF+SVM [90]	95.4%
SIFT+HOG+Gabor [126]	98.0%
BIF+DIF+SVM [92]	94.0%
CNN [83]	97.3%
BSIF+LPQ+PML+SVM (ours) [14]	98.3%



Figure 4.27: Examples of good and poor demographic estimation on LFW database.

4.7.5 FERET

In table 4.11, we give a comparison between our approach and some state-of-the-art approaches. Figure 4.28 illustrates some examples for good and poor demographic estimation obtained by the proposed approach with FERET database.

Table 4.11: A comparison of the proposed approach with other demographic estimation approaches on FERET database.

Approach	Performance	
	Gender	Ethnicity
LBP + Real AdaBoost [86]	93.3%	NA
LBP + SVM [127]	92.0%	NA
BIF+DIF+SVM [92]	96.8%	NA
FVs+linear SVM [128]	97.7%	NA
BSIF+LPQ+PML+SVM (ours) [14]	99.0%	95.6%



Figure 4.28: Examples of good and poor demographic estimation on FERET database.

4.8 Challenge database evaluation

2016 ChaLearn Looking at People and Faces of the World Challenge² ran three parallel quantitative challenge tracks on RGB face analysis. The first track was about apparent age estimation. The dataset used in this track contains 8,000 images labeled with the apparent age. Figure 4.29 shows some samples of the dataset. Each image has been labeled by multiple individuals, using a collaborative Facebook implementation and Amazon Mechanical Turk. The votes variance is used as a measure of the error for the predictions [41]. The mean is used as the ground truth age. For quantitative evaluation instead of MAE, they used ϵ .



Figure 4.29: Samples from Chalearn2016 dataset.

These are some features which make the dataset a challenge to researchers:

- Images with diverse backgrounds.

²<http://gesture.chalearn.org/2016-looking-at-people-cvpr-challenge>

- Images taken in non-controlled environments.
- Non-labelled face bounding boxes, neither face landmarks.

Nearly 100 participants registered for this challenge, the top 8 participants are Deep learning-based approaches (shown in Table 4.12). OrangeLabs team [65] which are the winners of the apparent age competition used the pretrained VGG-16 convolutional neural network (CNN) to train it on the huge IMDB-Wiki dataset (500k+ face images) for biological age estimation and then fine-tune it for apparent age estimation using the dataset of the challenge. They got an age error of 0.2411 in terms of the score ϵ . Neither the CPU time nor the time complexity was mentioned in their work. Both palm seu team [62] and cmp+ETH [63] teams used the pretrained VGG-16 CNN in their approaches.

We conducted our experiments using the same protocols as the contenders of the challenge did. Compared to OrangeLabs team approach the difference between our results and theirs is about 0.0463 in terms of the score ϵ . Also, according to the Table 4.12 our results come in the second position. The results showed that our approach can give better results even in uncontrolled complex conditions. Thus, it confirms that deep-learning-based approaches are not necessarily the best ones.

Table 4.12: Comparison of the results between the participating teams approaches and our approach in the apparent age-estimation competition.

Position	Team	Test Error
1	OrangeLabs [65]	0.2411
2	palm seu [62]	0.3214
3	cmp+ETH [63]	0.3361
4	WYU CVL	0.3405
5	ITU SiMiT	0.3668
6	Bogazici	0.3740
7	MIPAL SNU	0.4569
8	DeepAge	0.4573
-	Ours	0.2874

The third track was about smile and gender classification. They used a different subset of images from the FotW dataset to create this challenging dataset. This dataset is composed of a training set (6,171 images), a validation set (3,086 images) and a test set (8,505 images). Table 4.13 shows the images distribution of the dataset.

Table 4.13: The images distribution based on facial expression and gender in the dataset.

Attribute	Train	Val	Test
Male	2946	1691	4614
Female	3318	1361	3799
Not sure	93	34	92
Smile	2234	1969	4411
No smile	3937	1117	3849

Nearly 60 participants registered for this competition. The 7 teams that finally submitted predictions are listed in Table 4.14 together with the accuracy of each of the participating teams. The performance evaluation using this database give good results in both smile and gender classification. From the fact of the results of smiles classification (facial expression), we can say that our can be applied to other face-based learning tasks.

SIAT MMLAB team [129] won the challenge, they proposed a deep model composed of two CNNs GNet and SNet, for facial gender and smile classification tasks. Also, they used multi-task and general-to-specific fine-tuning scheme while training their models. Their proposed deep model gave results of 92.69% for gender and 85.83% for smile classifications. Our results came in the third position compared to the ranking table of the challenge. We get 90.36% and 82.15% for gender and smile classification respectively.

Table 4.14: Comparison of the results between the participating teams approaches and our approach in the gender/smiles classification competition.

Team	Gender	Smile	Mean
SIAT MMLAB	0.9269	0.8583	0.8926
IVA NLPR	0.9152	0.8252	0.8702
VISI.CRIM	0.9016	0.8212	0.8614
SMILELAB NEU	0.8999	0.8148	0.8574
Lets Face it!	0.8454	0.8439	0.8446
CMP+ETH	0.7465	0.7189	0.7327
SRI	0.5716	0.5853	0.5784
Our approach	0.9036	0.8215	0.8625

The databases used in the 2016 ChaLearn Looking at People and Faces of the World Challenge are considered as uncontrolled (In the wild) databases. Our system had some poor face alignment which affect the global performance of the proposed method. Figures 4.30 and 4.31 show some poor face alignment for both databases.



Figure 4.30: Poor face preprocessing on Chalearn2016 dataset (age).

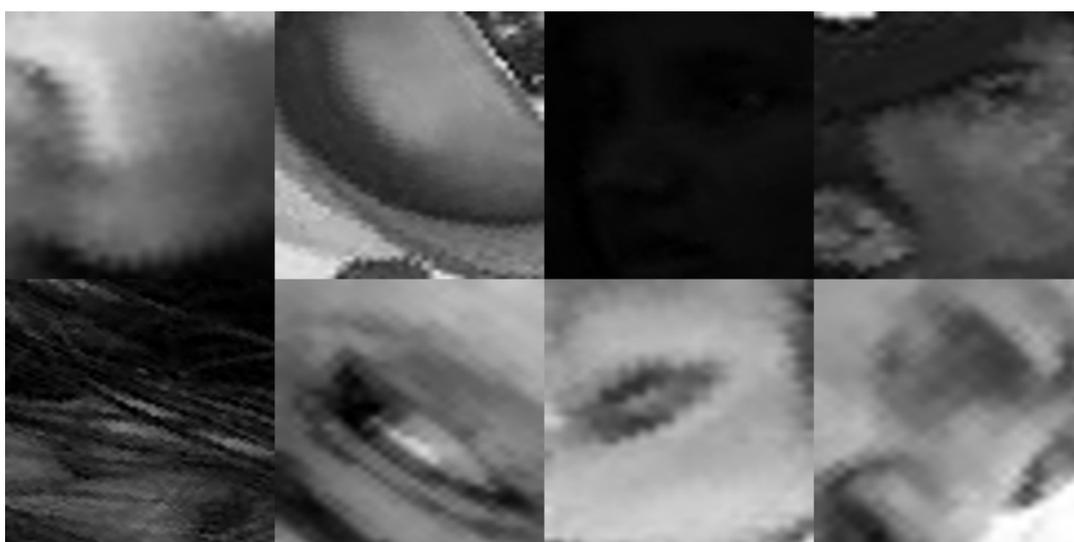


Figure 4.31: Poor face preprocessing on Chalearn2016 dataset (gender and smile).

4.9 Cross-database evaluation

In order to have a quantitative evaluation on multiple databases, we performed cross-database experiments on the demographic attributes.

4.9.1 Age estimation

For the age estimation task, we use two databases MORPH II and PAL since both provide the age, gender and ethnicity labels. This enables us to use the full hierarchical estimator (with three layers). The protocol adopted in this experiment is to use an entire database as the training subset and the other as the testing subset and vice-versa. The results are given in Table 4.15.

Table 4.15: The performance (MEA in years) of the cross-database evaluation in the case of age estimation.

Testing \ Training	PAL	MORPH II
PAL	/	6.1
MORPH II	5.5	/

We can observe that when MORPH II database is used as a training subset (PAL is used as a test dataset) the performance is better than the opposite case. This can be explained by the fact that MORPH II database is very rich and can better capture variation in the demographic attributes. Also, the performance on PAL deteriorated from 5.0 to 5.6. This is due to the fact that the two databases are different.

4.9.2 Gender classification

In the case of gender classification, we use all proposed databases to this end. We exclude the hierarchical estimation and use just the gender classifier. The protocol used in the task is the same which was used in the age estimation task. In table 4.16, we present the results. We can observe that whenever the IoG database is used in the training or testing subsets, the gender classification performance is not good compared to the other databases. This may be due to the distribution of IoG database which contains a lot of children images, unlike the others. Both IoG and LFW images are captured in the wild. Therefore, their results are quasi symmetric. In most cases, MORPH II and LFW have the best results which can be due to their richness in images.

Table 4.16: Cross-database evaluation in the case of gender classification.

Testing \ Training	PAL	MORPH II	IoG	LFW	FERET
PAL	/	87.6%	65.3%	82.1%	87.3%
MORPH II	89.0%	/	75.7%	90.3%	94.0%
IoG	75.6%	72.2%	/	79.1%	70.6%
LFW	84.5%	89.5%	81.0%	/	92.8%
FERET	91.1%	89.8%	71.5%	88.1%	/

4.9.3 Ethnicity classification

Like in age estimation and gender classifications tasks, we use the same protocol for ethnicity classification. We use three databases (MORPH II, PAL and FERET). The results are presented in Table 4.17. The results of cross-database evaluation are good given the fact that there is a difference in the distribution and in the images of these three databases.

Table 4.17: Cross-database evaluation in the case of ethnicity classification.

Testing \ Training	PAL	MORPH II	FERET
PAL	/	81.1%	91.6%
MORPH II	88.0%	/	87.7%
FERET	95.3%	85.2%	/

4.10 Real-time application

We implement our approach as a C++ application using OpenCV³ and DLIB libraries⁴. We used PML BSIF features. The structure of the system is composed of: (i) face preprocessing which includes detection and normalization, (ii) feature extraction and selection, and (iii) the demographic estimator which contains one SVM model for ethnicity classification, 3 SVM models for gender classification and 6 SVRs for age estimation. We emphasize that, at testing phase, the system will apply 2 SVM models and 1 SVR model each time. The CPU time associated with each part is given in Table 4.18. The application works at 30 fps, if there is more than one face it is best to ignore some frames. The test is carried out on a laptop DELL

³<http://opencv.org/>

⁴<http://dlib.net/>

7510 Precision (Xeon Processor E3-1535M v5 ,8M Cache, 2.90 GHz, 64GB RAM, Ubuntu 16.04).

Among the approaches which worked on the joint demographic attributes, only [91] and [92] reported the CPU time associated with their developed methods. Guo [91] has reported a 623 ms for the testing phase for one image excluding the time of feature extraction. Their test was on Lenovo Thinkpad X61 (Intel Core 2, Duo CPU T8100, @2.1GHz, 4G RAM, Windows Vista 64). In the other hand, Han [92] has got 90 ms on execution time for one image without excluding any part of the prototype. They used a laptop with 2.9 GHz dual-core Intel Core i7 processor and 8G RAM for the testing the system.

We emphasize the fact that the CPU time of proposed state-of-the-art demographic estimation methods is not always reported in the literature. Indeed, most researchers in this field were focusing on improving the accuracy of their methods and systems.

Table 4.18: CPU time (ms) of the proposed approach.

Face preprocessing	Feature extraction and selection	Demographic estimation	Total
9	12	7	28

This application is developed to support the facial recognition system by using this information as facial soft biometric traits. Moreover, the application can be used in security for the governments or to get statistics in social media and commercial centers.

4.11 Conclusion

In this chapter, we presented the results obtained from our approach using different protocols and different databases. Moreover, we studied the effect of different face representation and different features ratio on the results of demographic estimation. A comparison with the state-of-the-art in demographic estimation is given and prove that our approach gives stable and good results in the most known databases.

CONCLUSIONS AND FUTURE WORKS

Contents

	Page
5.1 Conclusions	81
5.2 Future works	82

5.1 Conclusions

In this thesis, we were interested in the problems of facial age estimation, facial gender classification, and facial ethnicity classification. We have described the facial soft biometrics and the facial demographic attributes. We illustrated also the ability of soft biometrics when it comes to supporting the biometrics systems. Also, we have given other application of demographic attributes besides supporting the recognition system. We presented different databases with a different variation such as illumination, pose, and quality. We have presented all the terms used in age estimation, gender classification and, ethnicity or race classification.

A state of the art of recent methods and previous work has been given. We categorized these approaches based on the image representation and the estimation algorithms, the categories are anthropometric-based, image-based, deep-learning-based, and other approaches. Our approach belongs to image-based approaches which is mainly based on texture descriptors. In the literature review, we show the similarities and the differences between the methods which includes all the demographic attributes. Besides this, we show the strength and the weakness of each method. This allowed us to choose the appropriate tools for our problem.

As we explained in Chapter 3, the proposed approach is composed mainly of three parts which are face preprocessing, feature extraction and selection, and demographic estimation. Our face preprocessing method is based on face alignment using the eyes center and it has been proved its good performance when dealing with frontal and near-frontal faces. That has been shown through our experiments about the effect of 2D and 3D face preprocessing on the demographic estimation system. Although, our proposed 2D face preprocessing method is vulnerable to profile faces and wild poses.

Regarding the feature extraction and selection part, we did extensive experiments. First, we tested the following image texture descriptors: LBP, LDP, LPQ, and BSIF. From these experiments, we came out with the results that LPQ and BSIF overcome LBP and LDP in term of classification rates for different scenarios. Also, we found that the mixed features between BSIF and LPQ are the best pair when dealing with PMLs. Second, we presented and tested different face representation (MB, ML, and PML). Our proposed PML face representation is inspired from MB and ML, where the original face image is explicitly transformed and represented by several scaled MB images. This representation showed better results in all the demographic attributes cases compared to MB and ML face representations. Last, we had conducted many experiments concerning the effects of feature selection on our demographic estimation approach. We found that using 5% until 15% from the sorted features is a good choice for the trade-off efficiency-accuracy.

About the last part of our proposed approach which is the hierarchical demographic estimator. We conducted some experiments to see the effects of each layer on the performance of the system. Firstly, we tested each layer of the hierarchical demographic estimator alone and we noticed a decreasing performance of the system in age, gender, and ethnicity. Finally, we investigated the effects of the order of the layers. This latter shows that using ethnicity layer as root gives the highest accuracy among all other demographic attributes.

In brief, the main contributions which gave us stable and good results are the 2D face pre-processing, the proposed PML face representation and the proposed hierarchical demographic estimator. When comparing our results with the existing methods, this has shown that our approach achieves promising results. In addition, the implementation of our approach in C++ proved that it is suited to real time applications. Finally, we believe that our proposed approach can also be applied in several facial analysis applications.

5.2 Future works

Biometrics systems becoming very important in human daily life either for protection or other goals. As an active research area, there are many new research ideas to develop. We envision the use of deep features obtained from pre-trained neural networks. We also envision the improvement of the hierarchical demographic estimator by improving the architecture of the hierarchy and by introducing more sophisticated feature selection methods. Moreover, we envision the use of color information, especially for ethnicity classification. Therefore, the possibility of using other space colors. Concerning the application, we suggest the use of multiple threads with a synchronization between the different threads. Finally, we propose the including of face tracker in order to predict the demographic information from video sequences and the use of temporal features.

BIBLIOGRAPHY

- [1] A. Dantcheva, C. Velardo, A. D'Angelo, and J.-L. Dugelay, "Bag of soft biometrics for person identification," *Multimedia Tools Appl.*, vol. 51, pp. 739–777, Jan. 2011.
- [2] Y. H. Kwon and N. da Vitoria Lobo, "Age classification from facial images," *Computer Vision and Image Understanding*, vol. 74, no. 1, pp. 1 – 21, 1999.
- [3] A. Gunay and V. V. Nabiyev, "Automatic detection of anthropometric features from facial images," in *2007 IEEE 15th Signal Processing and Communications Applications*, pp. 1–4, June 2007.
- [4] J. Ylioinas, A. Hadid, and M. Pietikäinen, "Age classification in unconstrained conditions using lbp variants," in *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, pp. 1257–1260, Nov 2012.
- [5] H. Han, C. Otto, and A. K. Jain, "Age estimation from face images: Human vs. machine performance," in *2013 International Conference on Biometrics (ICB)*, pp. 1–8, June 2013.
- [6] S. E. Bekhouche, A. Ouafi, A. Benlamoudi, A. Taleb-Ahmed, and A. Hadid, "Facial age estimation and gender classification using multi level local phase quantization," in *Control, Engineering Information Technology (CEIT), 2015 3rd International Conference on*, pp. 1–4, May 2015.
- [7] C. Yan, C. Lang, T. Wang, X. Du, and C. Zhang, "Age estimation based on convolutional neural network," in *Proceedings of the 15th Pacific-Rim Conference on Advances in Multimedia Information Processing — PCM 2014 - Volume 8879*, (New York, NY, USA), pp. 211–220, Springer-Verlag New York, Inc., 2014.
- [8] G. Levi and T. Hassner, "Age and gender classification using convolutional neural networks," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2015.
- [9] X. Wang, R. Guo, and C. Kambhamettu, "Deeply-learned feature for age estimation," in *2015 IEEE Winter Conference on Applications of Computer Vision*, pp. 534–541, Jan 2015.
- [10] X. Geng, Z.-H. Zhou, Y. Zhang, G. Li, and H. Dai, "Learning from facial aging patterns for automatic age estimation," in *Proceedings of the 14th ACM International Conference on Multimedia*, MM '06, (New York, NY, USA), pp. 307–316, ACM, 2006.
- [11] C.-C. Lai, C.-H. Wu, S.-T. Pan, S.-J. Lee, and B.-H. Lin, *Gender Recognition Using Local Block Difference Pattern*, pp. 45–52. Cham: Springer International Publishing, 2017.

- [12] F. Juefei-Xu, E. Verma, P. Goel, A. Cherodian, and M. Savvides, “Deepgender: Occlusion and low resolution robust facial gender classification via progressively trained convolutional neural networks with attention,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 136–145, June 2016.
- [13] Q. Tian, T. Arbel, and J. J. Clark, “Efficient gender classification using a deep lda-pruned net,” *arXiv preprint arXiv:1704.06305*, 2017.
- [14] S. Bekhouche, A. Ouafi, F. Dornaika, A. Taleb-Ahmed, and A. Hadid, “Pyramid multi-level features for facial demographic estimation,” *Expert Systems with Applications*, vol. 80, pp. 297 – 310, 2017.
- [15] A. Hyvärinen, J. Hurri, and P. Hoyer, *Natural Image Statistics: A Probabilistic Approach to Early Computational Vision*. Computational Imaging and Vision, Springer London, 2009.
- [16] H. Rhodes, *Alphonse Bertillon, father of scientific detection*. Abelard-Schuman, 1956.
- [17] A. K. Jain, K. Nandakumar, X. Lu, and U. Park, *Integrating Faces, Fingerprints, and Soft Biometric Traits for User Recognition*, pp. 259–269. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004.
- [18] A. K. Jain, S. C. Dass, and K. Nandakumar, *Soft Biometric Traits for Personal Recognition Systems*, pp. 731–738. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004.
- [19] D. A. Reid and M. S. Nixon, “Using comparative human descriptions for soft biometrics,” in *2011 International Joint Conference on Biometrics (IJCB)*, pp. 1–6, Oct 2011.
- [20] U. Park and A. K. Jain, “Face matching and retrieval using soft biometrics,” *IEEE Transactions on Information Forensics and Security*, vol. 5, pp. 406–415, Sept 2010.
- [21] T. Correa, A. W. Hinsley, and H. G. de Zuniga, “Who interacts on the web?: The intersection of users’ personality and social media use,” *Computers in Human Behavior*, vol. 26, no. 2, pp. 247 – 253, 2010.
- [22] N. Kumar, P. Belhumeur, and S. Nayar, *FaceTracer: A Search Engine for Large Collections of Images with Faces*, pp. 340–353. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008.
- [23] “Fg-net (face and gesture recognition network).” <http://www.prima.inrialpes.fr/FGnet/>, 2004. Accessed: 2016-09-30.
- [24] K. Ricanek Jr. and T. Tesafaye, “Morph: A longitudinal image database of normal adult age-progression,” in *Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition, FGR ’06*, (Washington, DC, USA), pp. 341–345, IEEE Computer Society, 2006.
- [25] P. J. Phillips, H. Moon, P. Rauss, and S. A. Rizvi, “The feret evaluation methodology for face-recognition algorithms,” in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 137–143, Jun 1997.
- [26] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, “Labeled faces in the wild: A database for studying face recognition in unconstrained environments,” Tech. Rep. 07-49, University of Massachusetts, Amherst, October 2007.

- [27] M. Minear and D. C. Park, "A lifespan database of adult facial stimuli," *Behavior Research Methods, Instruments, & Computers*, vol. 36, no. 4, pp. 630–633, 2004.
- [28] A. C. Gallagher and T. Chen, "Understanding images of groups of people," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 256–263, June 2009.
- [29] E. Eidinger, R. Enbar, and T. Hassner, "Age and gender estimation of unfiltered faces," *IEEE Transactions on Information Forensics and Security*, vol. 9, pp. 2170–2179, Dec 2014.
- [30] R. Rothe, R. Timofte, and L. Van Gool, "Deep expectation of real and apparent age from a single image without facial landmarks," *International Journal of Computer Vision*, pp. 1–14, 2016.
- [31] M. GONZALEZ-ULLOA and E. S. FLORES, "SENILITY OF THE FACE—BASIC STUDY TO UNDERSTAND ITS CAUSES AND EFFECTS," *Plastic and Reconstructive Surgery*, vol. 36, pp. 239–246, aug 1965.
- [32] D. Berry and L. McArthur, "Perceiving character in faces: the impact of age-related craniofacial changes on social perception.," *Psychological bulletin*, vol. 100, no. 1, p. 3, 1986.
- [33] X. Geng, Y. Fu, and K. Smith-Miles, "Automatic facial age estimation," in *11th Pacific Rim International Conference on Artificial Intelligence*, pp. 1–130, 2010.
- [34] G. Petra, "Introduction to human age estimation using face images," *The Journal of Slovak University of Technology*, vol. 21, pp. 24–30, 2013.
- [35] M. G. Rhodes, "Age estimation of faces: a review," *Applied Cognitive Psychology*, vol. 23, pp. 1–12, jan 2009.
- [36] A. K. Jain, S. C. Dass, and K. Nandakumar, "Can soft biometric traits assist user recognition?," 2004.
- [37] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," *International Journal of Forecasting*, vol. 22, no. 4, pp. 679 – 688, 2006.
- [38] X. Geng, Z. H. Zhou, and K. Smith-Miles, "Automatic age estimation based on facial aging patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, pp. 2234–2240, Dec 2007.
- [39] J. M. Bland and D. G. Altman, "Measurement error and correlation coefficients.," *BMJ: British Medical Journal*, vol. 313, no. 7048, p. 41, 1996.
- [40] S. Escalera, J. Fabian, P. Pardo, X. Baró, J. González, H. J. Escalante, D. Misevic, U. Steiner, and I. Guyon, "Chalearn looking at people 2015: Apparent age and cultural event recognition datasets and results," in *2015 IEEE International Conference on Computer Vision Workshop (ICCVW)*, pp. 243–251, Dec 2015.
- [41] S. Escalera, M. Torres Torres, B. Martinez, X. Baro, H. Jair Escalante, I. Guyon, G. Tzimiropoulos, C. Corneou, M. Oliu, M. Ali Bagheri, and M. Valstar, "Chalearn looking at people and faces of the world: Face analysis workshop and challenge 2016," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2016.

- [42] Y. H. Kwon and N. da Vitoria Lobo, "Age classification from facial images," in *1994 Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 762–767, Jun 1994.
- [43] T. Alley, *Social and Applied Aspects of Perceiving Faces*. Resources for ecological psychology, Lawrence Erlbaum Associates, 1988.
- [44] E. Vezzetti and F. Marcolin, "3d human face description: landmarks measures and geometrical features," *Image and Vision Computing*, vol. 30, no. 10, pp. 698 – 712, 2012. 3D Facial Behaviour Analysis and Understanding.
- [45] A. Gunay and V. V. Nabiyev, "Automatic age classification with lbp," in *2008 23rd International Symposium on Computer and Information Sciences*, pp. 1–4, Oct 2008.
- [46] C. Shan, "Learning local features for age estimation on real-life faces," in *Proceedings of the 1st ACM International Workshop on Multimodal Pervasive Video Analysis, MPVA '10*, (New York, NY, USA), pp. 23–28, ACM, 2010.
- [47] E. Fazl-Ersi, M. E. Mousa-Pasandi, R. Laganière, and M. Awad, "Age and gender recognition using informative features of various types," in *2014 IEEE International Conference on Image Processing (ICIP)*, pp. 5891–5895, Oct 2014.
- [48] J. Ylioinas, A. Hadid, X. Hong, and M. Pietikäinen, *Age Estimation Using Local Binary Pattern Kernel Density Estimate*, pp. 141–150. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013.
- [49] G. Guo, G. Mu, Y. Fu, and T. S. Huang, "Human age estimation using bio-inspired features," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 112–119, June 2009.
- [50] G. Guo and G. Mu, "Simultaneous dimensionality reduction and human age estimation via kernel partial least squares regression," in *CVPR 2011*, pp. 657–664, June 2011.
- [51] K. Y. Chang, C. S. Chen, and Y. P. Hung, "Ordinal hyperplanes ranker with cost sensitivities for age estimation," in *CVPR 2011*, pp. 585–592, June 2011.
- [52] X. Geng, C. Yin, and Z. H. Zhou, "Facial age estimation by learning from label distributions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, pp. 2401–2412, Oct 2013.
- [53] G. Guo, Y. Fu, C. R. Dyer, and T. S. Huang, "Image-based human age estimation by manifold learning and locally adjusted robust regression," *IEEE Transactions on Image Processing*, vol. 17, pp. 1178–1188, July 2008.
- [54] S. Bekhouche, A. Ouafi, A. Taleb-Ahmed, A. Hadid, and A. Benlamoudi, "Facial age estimation using bsif and lbp," in *Proceeding of the first International Conference on Electrical Engineering ICEEB'14*, 2014.
- [55] Wikipedia, "Convolutional neural network — wikipedia, the free encyclopedia," 2017. [Online; accessed 6-May-2017].
- [56] R. Rothe, R. Timofte, and L. V. Gool, "Dex: Deep expectation of apparent age from a single image," in *2015 IEEE International Conference on Computer Vision Workshop (ICCVW)*, pp. 252–257, Dec 2015.

- [57] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
- [58] X. Liu, S. Li, M. Kan, J. Zhang, S. Wu, W. Liu, H. Han, S. Shan, and X. Chen, "Agenet: Deeply learned regressor and classifier for robust apparent age estimation," in *2015 IEEE International Conference on Computer Vision Workshop (ICCVW)*, pp. 258–266, Dec 2015.
- [59] Z. Kuang, C. Huang, and W. Zhang, "Deeply learned rich coding for cross-dataset facial age estimation," in *2015 IEEE International Conference on Computer Vision Workshop (ICCVW)*, pp. 338–343, Dec 2015.
- [60] R. Ranjan, S. Zhou, J. C. Chen, A. Kumar, A. Alavi, V. M. Patel, and R. Chellappa, "Unconstrained age estimation with deep convolutional neural networks," in *2015 IEEE International Conference on Computer Vision Workshop (ICCVW)*, pp. 351–359, Dec 2015.
- [61] R. C. Malli, M. Aygün, and H. K. Ekenel, "Apparent age estimation using ensemble of deep learning models," in *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 714–721, June 2016.
- [62] Z. Huo, X. Yang, C. Xing, Y. Zhou, P. Hou, J. Lv, and X. Geng, "Deep age distribution learning for apparent age estimation," in *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 722–729, June 2016.
- [63] M. Uricár, R. Timofte, R. Rothe, J. Matas, and L. V. Gool, "Structured output svm prediction of apparent age, gender and smile from deep features," in *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 730–738, June 2016.
- [64] F. Gürpınar, H. Kaya, H. Dibeklioglu, and A. A. Salah, "Kernel elm and cnn based facial age estimation," in *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 785–791, June 2016.
- [65] G. Antipov, M. Baccouche, S. A. Berrani, and J. L. Dugelay, "Apparent age estimation from face images combining general and children-specialized deep learning models," in *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 801–809, June 2016.
- [66] A. Lanitis, C. J. Taylor, and T. F. Cootes, "Toward automatic simulation of aging effects on face images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 442–455, Apr 2002.
- [67] T. F. Cootes, G. J. Edwards, and C. J. Taylor, *Active appearance models*, pp. 484–498. Berlin, Heidelberg: Springer Berlin Heidelberg, 1998.
- [68] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, pp. 681–685, Jun 2001.
- [69] A. Lanitis, C. Draganova, and C. Christodoulou, "Comparing different classifiers for automatic age estimation," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 34, pp. 621–628, Feb 2004.

- [70] X. Geng, K. Smith-Miles, and Z.-H. Zhou, "Facial age estimation by nonlinear aging pattern subspace," in *Proceedings of the 16th ACM International Conference on Multimedia*, MM '08, (New York, NY, USA), pp. 721–724, ACM, 2008.
- [71] S. Yan, X. Zhou, M. Liu, M. Hasegawa-Johnson, and T. S. Huang, "Regression from patch-kernel," in *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8, June 2008.
- [72] Wikipedia, "Receiver operating characteristic — wikipedia, the free encyclopedia," 2017. [Online; accessed 3-May-2017].
- [73] A. M. Burton, V. Bruce, and N. Dench, "What's the difference between men and women? evidence from facial measurement," *Perception*, vol. 22, no. 2, pp. 153–176, 1993. PMID: 8474841.
- [74] J.-M. Fellous, "Gender discrimination and prediction on the basis of facial metric information," *Vision Research*, vol. 37, no. 14, pp. 1961 – 1973, 1997.
- [75] X. Han, H. Ugail, and I. Palmer, "Gender classification based on 3d face geometry features using svm," in *2009 International Conference on CyberWorlds*, pp. 114–118, Sept 2009.
- [76] T. Jabid, M. H. Kabir, and O. Chae, "Gender classification using local directional pattern (ldp)," in *2010 20th International Conference on Pattern Recognition*, pp. 2162–2165, Aug 2010.
- [77] C. Li, Q. Liu, J. Liu, and H. Lu, "Learning ordinal discriminative features for age estimation," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2570–2577, June 2012.
- [78] A. Hadid, J. Ylioinas, M. Bengherabi, M. Ghahramani, and A. Taleb-Ahmed, "Gender and texture classification: A comparative analysis using 13 variants of local binary patterns," *Pattern Recognition Letters*, vol. 68, Part 2, pp. 231 – 238, 2015. Special Issue on "Soft Biometrics".
- [79] B. Patel, R. Maheshwari, and R. Balasubramanian, "Multi-quantized local binary patterns for facial gender classification," *Computers & Electrical Engineering*, vol. 54, pp. 271 – 284, 2016.
- [80] M. Castrillón-Santana, M. D. Marsico, M. Nappi, and D. Riccio, "Meg: Texture operators for multi-expert gender classification," *Computer Vision and Image Understanding*, vol. 156, pp. 4 – 18, 2017. Image and Video Understanding in Big Data.
- [81] M. Castrillón-Santana, J. Lorenzo-Navarro, and E. Ramón-Balmaseda, "Descriptors and regions of interest fusion for in- and cross-database gender classification in the wild," *Image and Vision Computing*, vol. 57, pp. 15 – 24, 2017.
- [82] H. Moeini and S. Mozaffari, "Gender dictionary learning for gender classification," *Journal of Visual Communication and Image Representation*, vol. 42, pp. 1 – 13, 2017.
- [83] G. Antipov, S.-A. Berrani, and J.-L. Dugelay, "Minimalistic cnn-based ensemble model for gender prediction from face images," *Pattern Recognition Letters*, vol. 70, pp. 59 – 65, 2016.
- [84] X. Lu and A. K. Jain, "Ethnicity identification from face images," 2004.

- [85] S. Hosoi, E. Takikawa, and M. Kawade, "Ethnicity estimation with facial images," in *Sixth IEEE International Conference on Automatic Face and Gesture Recognition, 2004. Proceedings.*, pp. 195–200, May 2004.
- [86] Z. Yang and H. Ai, *Demographic Classification with Local Binary Patterns*, pp. 464–473. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007.
- [87] G. Guo and G. Mu, "Human age estimation: What is the influence across race and gender?," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*, pp. 71–78, June 2010.
- [88] G. Guo and G. Mu, "Joint estimation of age, gender and ethnicity: Cca vs. pls," in *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, pp. 1–6, April 2013.
- [89] A. Hadid and M. Pietikäinen, "Demographic classification from face videos using manifold learning," *Neurocomputing*, vol. 100, pp. 197 – 205, 2013. Special issue: Behaviours in video.
- [90] H. Han and A. K. Jain, "Age, gender and race estimation from unconstrained face images," Tech. Rep. MSU-CSE-14-5, Department of Computer Science, Michigan State University, East Lansing, Michigan, July 2014.
- [91] G. Guo and G. Mu, "A framework for joint estimation of age, gender and ethnicity on a large database," *Image and Vision Computing*, vol. 32, no. 10, pp. 761 – 770, 2014. Best of Automatic Face and Gesture Recognition 2013.
- [92] H. Han, C. Otto, X. Liu, and A. K. Jain, "Demographic estimation from face images: Human vs. machine performance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, pp. 1148–1161, June 2015.
- [93] D. Yi, Z. Lei, and S. Z. Li, *Age Estimation by Multi-scale Convolutional Network*, pp. 144–158. Cham: Springer International Publishing, 2015.
- [94] M. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis, and Machine Vision*. PWS Pub., 1999.
- [95] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, vol. 1, pp. I–511–I–518 vol.1, 2001.
- [96] V. Kazemi and J. Sullivan, "One millisecond face alignment with an ensemble of regression trees," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1867–1874, June 2014.
- [97] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, pp. 2037–2041, Dec 2006.
- [98] M. Pietikäinen, A. Hadid, G. Zhao, and T. Ahonen, *Computer Vision Using Local Binary Patterns*. Computational Imaging and Vision, Springer London, 2011.

- [99] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognition*, vol. 29, no. 1, pp. 51 – 59, 1996.
- [100] T. Ojala, M. Pietikäinen, and T. Mäenpää, *Gray Scale and Rotation Invariant Texture Classification with Local Binary Patterns*, pp. 404–420. Berlin, Heidelberg: Springer Berlin Heidelberg, 2000.
- [101] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 971–987, Jul 2002.
- [102] Wikipedia, "Local binary patterns — wikipedia, the free encyclopedia," 2017. [Online; accessed 1-June-2017].
- [103] T. Jabid, M. H. Kabir, and O. Chae, "Local directional pattern (ldp) for face recognition," in *2010 Digest of Technical Papers International Conference on Consumer Electronics (ICCE)*, pp. 329–330, Jan 2010.
- [104] T. Jabid, M. H. Kabir, and O. Chae, "Local directional pattern (ldp) - a robust image descriptor for object recognition," in *Proceedings of the 2010 7th IEEE International Conference on Advanced Video and Signal Based Surveillance, AVSS '10*, (Washington, DC, USA), pp. 482–487, IEEE Computer Society, 2010.
- [105] M. H. Kabir, T. Jabid, and O. Chae, "Local directional pattern variance (ldpv): a robust feature descriptor for facial expression recognition.," *Int. Arab J. Inf. Technol.*, vol. 9, no. 4, pp. 382–391, 2012.
- [106] V. Ojansivu and J. Heikkilä, *Blur Insensitive Texture Classification Using Local Phase Quantization*, pp. 236–243. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008.
- [107] T. Ahonen, E. Rahtu, V. Ojansivu, and J. Heikkila, "Recognition of blurred faces using local phase quantization," in *2008 19th International Conference on Pattern Recognition*, pp. 1–4, Dec 2008.
- [108] V. Ojansivu, E. Rahtu, and J. Heikkila, "Rotation invariant local phase quantization for blur insensitive texture analysis," in *2008 19th International Conference on Pattern Recognition*, pp. 1–4, Dec 2008.
- [109] J. Kannala and E. Rahtu, "Bsf: Binarized statistical image features," in *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, pp. 1363–1366, Nov 2012.
- [110] D. T. Nguyen, S. R. Cho, and K. R. Park, *Future Information Technology: FutureTech 2014*, ch. Human Age Estimation Based on Multi-level Local Binary Pattern and Regression Method, pp. 433–438. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014.
- [111] S. Bekhouche, A. Ouafi, A. Benlamoudi, A. Taleb-Ahmed, and A. Hadid, "Automatic age estimation and gender classification in the wild," in *Proceeding of the International Conference on Automatic control, Telecommunications and Signals ICATS'15*, 2015.
- [112] W. Wang, W. Chen, and D. Xu, "Pyramid-based multi-scale lbp features for face recognition," in *Multimedia and Signal Processing (CMSP), 2011 International Conference on*, vol. 1, pp. 151–155, May 2011.
- [113] H. Lohninger, *Teach/Me - Data Analysis: Single User Edition*. Springer Berlin Heidelberg, 1999.

- [114] Wikipedia, “Support vector machine — wikipedia, the free encyclopedia,” 2017. [Online; accessed 1-June-2017].
- [115] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin, “LIBLINEAR: A library for large linear classification,” *Journal of Machine Learning Research*, vol. 9, pp. 1871–1874, 2008.
- [116] T. Hassner, S. Harel, E. Paz, and R. Enbar, “Effective face frontalization in unconstrained images,” in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.
- [117] I. Huerta, C. Fernández, C. Segura, J. Hernando, and A. Prati, “A deep analysis on age estimation,” *Pattern Recognition Letters*, vol. 68, Part 2, pp. 239 – 249, 2015. Special Issue on “Soft Biometrics”.
- [118] S. E. Choi, Y. J. Lee, S. J. Lee, K. R. Park, and J. Kim, “A comparative study of local feature extraction for age estimation,” in *2010 11th International Conference on Control Automation Robotics Vision*, pp. 1280–1284, Dec 2010.
- [119] K. Luu, K. Seshadri, M. Savvides, T. D. Bui, and C. Y. Suen, “Contourlet appearance model for facial age estimation,” in *2011 International Joint Conference on Biometrics (IJCB)*, pp. 1–8, Oct 2011.
- [120] A. Günay and V. V. Nabiyev, *Age Estimation Based on Hybrid Features of Facial Images*, pp. 295–304. Cham: Springer International Publishing, 2016.
- [121] D. T. Nguyen, S. R. Cho, K. Y. Shin, J. W. Bang, and K. R. Park, “Comparative study of human age estimation with or without preclassification of gender and facial expression,” *The Scientific World Journal*, vol. 2014, pp. 1–15, 2014.
- [122] J. Mansanet, A. Albiol, and R. Paredes, “Local deep neural networks for gender recognition,” *Pattern Recognition Letters*, vol. 70, pp. 80 – 86, 2016.
- [123] P. Dago-Casas, D. González-Jiménez, L. L. Yu, and J. L. Alba-Castro, “Single- and cross- database benchmarks for gender classification under unconstrained settings,” in *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, pp. 2152–2159, Nov 2011.
- [124] C. Shan, “Learning local binary patterns for gender classification on real-world face images,” *Pattern Recognition Letters*, vol. 33, no. 4, pp. 431 – 437, 2012. Intelligent Multimedia Interactivity.
- [125] J. E. Tapia and C. A. Perez, “Gender classification based on fusion of different spatial scale features selected by mutual information from histogram of lbp, intensity, and shape,” *IEEE Transactions on Information Forensics and Security*, vol. 8, pp. 488–499, March 2013.
- [126] H. Ren and Z. N. Li, “Gender recognition using complexity-aware local features,” in *2014 22nd International Conference on Pattern Recognition*, pp. 2389–2394, Aug 2014.
- [127] E. Mäkinen and R. Raisamo, “An experimental comparison of gender classification methods,” *Pattern Recognition Letters*, vol. 29, no. 10, pp. 1544 – 1556, 2008.
- [128] W. Zhang, M. L. Smith, L. N. Smith, and A. Farooq, “Gender and gaze gesture recognition for human-computer interaction,” *Computer Vision and Image Understanding*, vol. 149, pp. 32 – 50, 2016.

- [129] K. Zhang, L. Tan, Z. Li, and Y. Qiao, "Gender and smile classification using deep convolutional neural networks," in *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 739–743, June 2016.