

2014-09-29

Wavelet-based Recognition of Facial Expressions and Faces

Poursaberi, Ahmad

Poursaberi, A. (2014). Wavelet-based Recognition of Facial Expressions and Faces (Doctoral thesis, University of Calgary, Calgary, Canada). Retrieved from <https://prism.ucalgary.ca>. doi:10.11575/PRISM/28676
<http://hdl.handle.net/11023/1823>

Downloaded from PRISM Repository, University of Calgary

UNIVERSITY OF CALGARY

Wavelet-based Recognition of Facial Expressions and Faces

by

Ahmad Poursaberi

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE

DEGREE OF DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN ELECTRICAL AND COMPUTER ENGINEERING

CALGARY, ALBERTA

SEPTEMBER, 2014

© Ahmad Poursaberi 2014

Abstract

The proposed research contributes to the development of biometric systems and its applications. The facial biometric, as a part of the underlying technology, is a focus of this thesis. Specifically, it is aimed at developing the new concepts of biometrics systems, such as systems with situational awareness, interviewing, and human-machine interaction support system that analyze human behavioral biometrics. The systems make use of facial expression data, besides the traditional visual-spectrum data for face recognition. A new technique for facial expression recognition (FER), in particular, is targeted in this thesis as one of the important components in human behavior analysis. Combined with face recognition (FR), it is a basis for a multi-biometric decision support tool. Most of the available approaches for the recognition of faces and expressions only consider either the expression-invariant face recognition or the facial expression recognition regardless of identity. This research proposes a facial biometric analysis using a new facial feature detection based on a modified Frangi filter. The joint FR/FER uses new feature extraction technique called Gauss-Laguerre wavelets, which provides rich and unique features suitable for expressions classification and identity recognition. In addition, this study considers facial biometric in infrared spectrum, which provides additional information (such as face region temperature) which is required, in particular, at a border check point. Moreover, recognition of expressions from video is investigated to show the extendibility of the proposed technique for video processing as well.

Keywords: biometrics, facial expression recognition, Gauss-Laguerre wavelets, Frangi filter, face recognition, infrared biometrics, and behavioral biometrics.

Acknowledgements

This dissertation would not have been done without my supervisor, Dr. Svetlana Yanushkevich, who walked me through this research. Dear Svetlana, I would like to thank you for all the continuous supports, mentally, emotionally, and scientific, unwavering enthusiasm for my research, and numerous hours you spent discussing my results, reviewing my publications, and scholarship applications. I really appreciate your patience, encouragement and invaluable knowledge. You helped me from the very beginning of my PhD to come up with the right and novel ideas, provided right information and guidelines in the field of facial expression recognition, and righteously corrected me in many ways from time to time. I know I was off-track for a short time during my thesis but you helped and supported me a lot to get back on track and finish my research in a better way. I want to express my deepest appreciation to you. I have really enjoyed my research I was privileged to do in your group and under your supervision. Thank you for the financial supports you've provided to accomplish my dream and get my PhD degree.

Also many thanks to Dr. Marina Gavrilova, my co-supervisor, for her generous scientific and emotional supports. Dear Marina, I remember you accepted to be my co-supervisor and let me to be one of your graduate students, although I was in the middle of my PhD. Thank you for all the time you spent with me to read my papers and prepare my applications.

Besides my advisor, I would like to express my deep gratitude and respect to the rest of my thesis committee: Dr. Henry Leung, Dr. Vassil Dimitrov, Dr. Qiao Sun, and Dr. Issa Traore for the time they spent to read my thesis and also provided insightful comments.

Special thanks to the Alberta Innovates Technology Future (AITF) research grant for the financial support to conduct this research.

I feel a deep sense of gratitude for my parents, for their unconditional and constant love and support. Thank you for all the good things you've taught me in my life, for all the sacrifices you made to raise me, for your prayers, and caring for educating and preparing me for my future.

Above all, I come to address my greatest appreciation to my lovely wife, Behnaz Bakherad, for her endless love, constant support, and for continued patience. It is difficult to find words to express my feeling to you. I am so blessed to have you beside me to enjoy every second of life. I could not have done my PhD without your sacrifices in life and being my only good reason when I was tired and frustrated during my PhD. Your dedication and persistent confidence in me helped me to accomplish my thesis.

Dedication

“To my beautiful wife, Behnaz Bakherad”

Table of Contents

Abstract	ii
Acknowledgements	iii
Dedication	v
Table of Contents	vi
List of Tables	ix
List of Figures and Illustrations	xi
List of Abbreviations	xv
List of Symbols	xvii
 CHAPTER ONE: INTRODUCTION	 1
1.1 Background	1
1.2 Key challenges	2
1.3 Objectives	3
1.4 Contributions	4
1.5 Thesis organization	6
1.6 Publications	7
 CHAPTER TWO: FACIAL EXPRESSION RECOGNITION: LITERATURE REVIEW	 8
2.1 Introduction	8
2.2 Basic emotions	12
2.2.1 Happiness	12
2.2.2 Fear	13
2.2.3 Disgust	14
2.2.4 Surprise	14
2.2.5 Sadness	14
2.2.6 Anger	14
2.3 Facial Action Coding System (FACS)	15
2.4 Facial Animation Parameters (FAPs)	17
2.5 Automatic facial expression analysis modeling	24
2.6 Conclusion	25
 CHAPTER THREE: FACIAL FEATURE DETECTION	 27
3.1 Introduction	27
3.2 Face detection	36
3.2.1 Viola-Jones face detector	37
3.3 The proposed method	39
3.3.1 Facial point detection	40
3.3.2 Filter modification	42
3.3.3 Eye detection	43
3.3.4 Detection of other facial features	46
3.4 Experiments and results	47
3.5 Conclusions	50
 CHAPTER FOUR: FEATURE EXTRACTION FOR FACIAL EXPRESSION RECOGNITION	 52

4.1 Appearance-based modelling.....	52
4.1.1 Optical flow	52
4.1.2 Pixel intensity values	53
4.1.3 Dimensionality reduction	53
4.1.4 Gabor filters	54
4.1.5 Wavelets	56
4.2 Geometry-based modeling.....	56
4.3 Gauss-Laguerre wavelets	59
4.4 Implementation and comparison.....	63
4.4.1 Global GL feature selection.....	65
4.4.2 Global textural and geometric feature selection	69
4.5 Conclusions.....	80
CHAPTER FIVE: EXPRESSION-INVARIANT FACE RECOGNITION	82
5.1 Introduction.....	82
5.2 Proposed method.....	85
5.3 Experimental results	87
5.3.1 Experiments on JAFFE database.....	89
5.3.2 Experiments on CK database.....	92
5.4 Conclusions.....	96
CHAPTER SIX: FACIAL EXPRESSION RECOGNITION FROM INFRARED IMAGES	99
6.1 Previous works.....	99
6.2 Proposed Method	101
6.2.1 Database	102
6.2.2 Feature extraction	103
6.2.3 Classification	106
6.3 Fusion.....	106
6.4 Implementation and results	110
6.4.1 Experiment on IR images	110
6.4.2 Fusion experiment	113
6.5 Conclusions.....	116
CHAPTER SEVEN: APPLICATIONS	118
7.1 Situational Awareness System (SAS).....	118
7.1.1 Facial biometrics for situational awareness.....	122
7.1.2 Decision making support	123
7.1.3 Dialogue support	125
7.2 Face biometrics in human-machine interfaces.....	130
7.3 Potential biomedical applications of face biometrics	131
7.4 Conclusion	133
CHAPTER EIGHT: CONCLUSIONS AND FUTURE WORKS.....	135
8.1 Conclusion	135
8.2 Future works	138

REFERENCES	140
APPENDIX A: FACIAL EXPRESSION RECOGNITION IN VIDEO	157

List of Tables

Table 2-1. AU definition. The images borrowed from [55]	18
Table 3-1. Average and separate detection rate	49
Table 3-2. Comparison with some of the existing methods.....	49
Table 4-1. Confusion matrix for JAFFE database	68
Table 4-2. Upper (distance 1-10) and lower (distance 11-15) face geometric distance	76
Table 4-3. Recognition accuracy (%) on the JAFFE database for different approaches.	78
Table 4-4. Confusion matrix for the leave-one-out method (the JAFFE database).....	78
Table 4-5. Comparison with other methods on the JAFFE database.....	79
Table 4-6. Recognition accuracy (%) on the Cohn-Kanade database for different approaches ...	79
Table 4-7. Confusion matrix for the leave-one-out method (the Cohn-Kanade database).....	79
Table 4-8. Comparison of facial expression recognition for the Cohn-Kanade database	80
Table 4-9. Recognition accuracy (%) on the MMI database for different approaches.....	80
Table 5-1. Comparison with different approaches on the JAFFE database.....	92
Table 5-2. Face recognition accuracy (%) on the JAFFE database for various approaches.....	93
Table 5-3. Face recognition accuracy (%) on the JAFFE database for the approaches in [111] [148] and the proposed one.....	93
Table 5-4. Face recognition rates (%) on the CK and CK+ dataset.....	95
Table 5-5. Face recognition rates (%) on the CK dataset	95
Table 5-6. Comparison of various approaches on the CK/CK+ databases for face expression recognition	97
Table 6-1. The confusion matrix – cross validation – OTCBVS.....	112
Table 6-2. The confusion matrix (%) – LOO–NVIE.....	113
Table 6-3. The confusion matrix (%) –CV–NVIE.....	113
Table 6-4. The confusion matrix (%) – LOO– without fusion on visible images	115
Table 6-5. The confusion matrix (%) –LOO– with fusion	115

Table 0-1. The 28 KNN classifiers used for video processing.	160
Table 0-2. The classification rate (%) for different expressions using different KNNs.	162

List of Figures and Illustrations

Figure 1-1. The flowchart of facial biometric analysis in this thesis.....	6
Figure 2-1. Typical facial expression recognition system with four major modules.....	9
Figure 2-2. Different expressions from CK [51] database (top to bottom, left to right): anger, disgust, fear, happiness, sadness, and surprise.....	13
Figure 2-3. Muscles of facial expression. 1, frontalis; 2, orbicularis oculi; 3, zygomaticus major; 4, risorius; 5, platysma; 6, depressor anguli oris [54].	16
Figure 2-4. Upper face AUs and combinations [56].....	17
Figure 2-5. Lower AUs and combinations [56].	22
Figure 2-6. A non-additive combination example [56].	23
Figure 2-7. Facial animation parameters: (a) Examples of facial points defined in FAP MPEG-4, (b) FAPU example based on the defined distances [58].	23
Figure 3-1. The approach proposed in [66] uses GWN features: first row, left to right: face image, GWN representation of a face, GWN features. Second row: sample images with detected features.....	29
Figure 3-2. The approach proposed in [67] uses the mask and geometrical features. Left: the mask used for geometrical feature relations, Right: The given image (a) and all the feature candidates for eight desired features (b-f).	30
Figure 3-3. Micro-features detection proposed in [68]. Left: Eye detector using neural network and image gradient intensities. Right: The defined micro-features around eye.	31
Figure 3-4. The outline of the method presented in [69]. (a) Modified face detector, (b) ROI extraction, (c) Gabor-based feature detection, (d) GentleBoost feature classification, (e) Detected features based on the proposed mask.....	32
Figure 3-5. Left: Point model of 22 fiducial points suggested in [70]. Right: samples of detected features with different pose/expressions.....	33
Figure 3-6. Left: Example of ASM. Right: The face image annotated with landmarks [74].	33
Figure 3-7. Search using Active Shape Model of a face [74].....	34
Figure 3-8. Left: Fiducial points overlaid on a neutral expression face from.....	35
Figure 3-9. Rectangle features defined by Viola-Jones [82].	38
Figure 3-10. An example of rectangle feature where each small rectangle indicates a pixel value. The example is borrowed from [98].....	39

Figure 3-12. Left to right: X-ray image of the peripheral vasculature, calculated vesselness of the image, calculated vesselness after inversion of the grey-scale map, image obtained by subtracting reference image from left image for visualization [63].....	42
Figure 3-13. Top to bottom: Input face images, the output of the Frangi filter, the output of the proposed filter (F1).	44
Figure 3-14. Face mask used for localization of other features based on eye position.	45
Figure 3-15. Fine estimation of eyes position based on binary mask obtained from F1 image. ..	45
Figure 3-16. Other facial feature detection based on geometric information of eyes (F2). Left to right: coarse and fine estimation of eyebrows, lips detection, nostrils detection.	47
Figure 4-1. Gabor filter response: left to right are the absolute, real and imaginary parts of a Gabor filter [113].	56
Figure 4-2. Four types of rectangular Haar wavelet-like features.	57
Figure 4-3. Example of geometry-based approach [62]: lower and upper face geometry features.	58
Figure 4-4. Example of using locations and relative Distances approach [124]: Geometrical parameters of the face, forming the feature vector.	59
Figure 4-5. Example of parameter estimation approach [125]: The deformable model, (b) the facial motion measurements.	59
Figure 4-6. Example of using Models of face musculature approach [43]: Determining of expressions from video sequences. (a) and (b) show expressions of smile and surprise, (c) and (d) show a 3-D model with surprise and smile expressions, and (e) and (f) show the spatio-temporal motion energy representation of facial motion for these expressions...	61
Figure 4-7. (a) Real part of GL function; $n = 4$, $K = 1$, $j = 2$. (b) Real part of GL CHF; with the variation of filter in spatial domain using the fixed scale, $k = 0, 1, \dots, 4$, and $n = 1, 2, \dots, 5$	64
Figure 4-8. Different expressions from JAFFE database (left to right): anger, disgust, fear, happiness, sadness, surprise, and neutral.	66
Figure 4-9. The GL wavelet response for three different filters on neutral and face with expression.	67
Figure 4-10. Recognition rate with respect to number of neurons in hidden layer.	68
Figure 4-11. Evolution of error in neural network during training.....	69

Figure 4-12. Normalization procedure (left to right): (a) input image (from the JAFFE database), (b) the extracted AAM fiducial points, (c) normalized image to have fixed distance between eyes (d) the localized and resized face.	71
Figure 4-13. Geometric feature selection based on fiducial points.	72
Figure 4-14. Variation of coefficient for geometric features versus average recognition rate.	73
Figure 4-15. Examples of various expressions from the Cohn-Kanade database.....	73
Figure 4-16. The sample face expression images from the MMI database.	74
Figure 5-1. Facial component separation. (a) Original face image, (b) the superposition of a neutral component, (c) the expression component [6].....	84
Figure 5-2. Flowchart of the joint FER and FR system. Feature extraction for FER and FR use the GL filter but with different parameter.	86
Figure 5-3. Gauss-Laguerre filter used for FR: Left-to-Right, Top-to-Bottom: Absolute, real, and imaginary part of the filter. The 3D representation of the absolute real and imaginary filter.....	87
Figure 5-4. The textural and geometric features: Left-to-Right, original image, geometrics distances, real part of filtered image, imaginary part of filtered image (©Jeffrey Cohn [51]).....	88
Figure 5-5. The real part of the Gabor filters with five frequencies and eight orientations [120].	88
Figure 5-6. Sample images in JAFFE, CK and MMI databases used in the experiments.	89
Figure 5-7. The normalization process in face recognition: (left to right): input image (from the JAFFE database), normalized image to have fixed distance between eyes, aligned image to have horizontal line between two eyes, resized image to 128×96 pixels.	89
Figure 5-8. Face recognition variation by keeping out different expressions. The top image is from [6] and the bottom image is from our approach.	98
Figure 6-1. Samples of OTCBVS database. From top to bottom: surprise happy, and angry expression.	103
Figure 6-2. Samples of USTC-NVIE database. From top to bottom, left to right: anger, disgust, fear, happy, sad, and surprise expression.	104
Figure 6-3. Flowchart of the proposed method.....	104
Figure 6-4. (a-b) Selected points of interested on face, (c) The distances used as geometric features from the defined mask.	106

Figure 6-5. Flowchart of the proposed method for fusion at feature level.	108
Figure 6-6. Facial points extraction for both spectra.	109
Figure 6-7. Geometric mask used for feature extraction.	109
Figure 6-8. Flowchart of fusion technique.....	111
Figure 6-9. Samples of filtered face image with real and imaginary parts of GL filter with different values for n and j, $k = 4$, and $a = 2$	112
Figure 6-10. Recognition rate for different expressions with and without fusion.	116
Figure 7-1. A proposed situational awareness system uses multiple sensors and various software components to analyze biometric data and provide dialogue-based decision- making support to aid security personnel.	122
Figure 7-2. Structure of the facial biometric assistant.	124
Figure 7-3. Bayesian network implemented in the experiment.	125

List of Abbreviations

Abbreviations	Definition
2D	Two-Dimensional
AAM	Active Appearance Model
AFER	Automatic Facial Expression Recognition
AI	Artificial Intelligence
ASM	Active Shape Model
AU	Action Unit
CHF	Circular Harmonic Function
CHP	Circular Harmonic Pyramid
CHW	Circular Harmonic Wavelets
CK	Cohn-Kanade
CV	Cross Validation
DCS	Component-based Dictionary Learning
DHS	Department of Homeland Security
EMG	Electromyography
FACS	Facial Action Coding System
FAPs	The Facial Animation parameters
FAST	Future Access Screening Technology
FDDL	Fisher Discrimination Dictionary Learning
FER	Facial Expression Recognition
FR	Face Recognition
FFT	Fast Fourier Transform
GL	Gauss-Laguerre
GMM	Gaussian Mixture Model
GWN	Gabor Wavelet Network
HCI	Human-Computer-Interface
HO-SVD	Higher-Order Singular Value Decomposition
ICA	Independent Component Analysis
IR	Infrared

JAFFE	Japanese Female Facial Expression
JSM	Joint Sparsity Model
KNN	K-Nearest Neighbor
LDA	Linear Discriminant Analysis
LOO	Leave-One-Out
MPEG	Moving Picture Experts Group
NN	Neural Network
OTCBVS	Object Tracking and Classification in and Beyond the Visible Spectrum
PASS	Physical Access Security System
PCA	Principal Component Analysis
PDM	Point Distribution Model
RE	Relative Error
SAS	Situational Awareness System
SHORE	High-Speed Object Recognition Engine
SoC	System on Chip
SVD	Singular Value Decomposition
SVM	Support Vector Machine
USTC-NVIE	University of Science and Technology of China - Natural Visible and Infrared facial Expression

List of Symbols

Symbol	Definition
λ	Eigenvalue
\mathcal{R}_β	Blobness
S	Second Order Structureness
Re	Real part
Fr	Modified Frangi Filter
d	Distance
T	Relative Error
v	Velocity
f	Frequency
θ	Direction
δ	Space Constant (Scale)
n	GL Filter's Order
k	GL Filter's Degree
$V_k^n(\)$	Radial Profile
$I(x, y)$	Image
$I_p(r, \theta)$	Polar Representation of an Image
(\tilde{x}, \tilde{y})	Pivot
$L_K^n(r)$	Generalized Laguerre Polynomial
φ	Angle
$a2^j$	Dyadic Scale
k, n	GL Filter's Scale
P_1, \dots, P_{15}	Geometric Distances
X	Image with Expression
X_n	Neutral Face
X_e	Pure Expression
α, β	Textural and Geometric Feature
P	Probability

Chapter One: **INTRODUCTION**

1.1 Background

This thesis concerns with the concept of a multi-functional facial biometric system. Biometric technologies involve multidisciplinary approaches and means for human identification based on his/her behavioral and physiological characteristics called biometrics. The latter includes image (e.g. facial, eye or body image) and signal (e.g. voice) processing. These two aspects (measurement of human physiological parameters, as well as their imaging) make biometric technologies relevant to biomedical image processing and analysis. The facial biometrics, which involves processing of facial images aim at analyzing and recognition of facial expressions, as well facial recognition itself, is outlined in this section. The applications of biometric technologies, which are being developed in the last 20 years, include:

1. Security in physical access systems such as border and immigration services, law enforcement and check-in facilities,
2. Interviewing and human-machine interaction support techniques, in particular, facial expression tracking and recognition,
3. Physiological and behavioral biometric measurements, such as facial image and thermal images, in the biomedical context for pre-diagnostics or assessment.

Facial biometrics carries extraordinary amount of information about human's appearance and physiological and emotional state. In particular, facial expression is an important behavioral biometrics that supports analysis of human behavior in the systems such as interviewing support, or monitoring of human response in human-machine interaction. Facial feature analysis and facial expression analysis in both visual and infrared is the focus of this research.

1.2 Key challenges

In spite of the continuous studies in the area of facial expression recognition in recent years, and considerable progress in developing the efficient algorithms for face recognition, the problem of joint identity and expression recognition is still an unsolved problem. There are three main challenges related to facial biometrics. These challenges are:

1. Separate environment analysis: Available algorithms for joint face and expression recognition usually encode the expression and identity variability of individuals in two different (or independent) “control parameters” and then perform recognition [1] [2] [3]. Face decomposition techniques [4] [5] [6], and bilinear models [7] which separate the expressions and identity components of a face are the famous techniques in this field. However, having different control environments/parameters to deal with the two problems (expressions and identity recognition) which in nature is a “single” problem - as it comes from “a face”-, may result in several problems including: inconsistency between the systems, computational cost, slower response time, etc.
2. Expression variations in face recognition: Different expressions can change the facial structure of any individual and this causes problem in face recognition. The muscular face movements due to facial expressions change the positions of the facial feature points like the eyebrows, eyes, nose, and the mouth corners from their original positions. Even the same expressions but with different emotional intensity will result in different facial features positions. Therefore having a robust facial feature detection that handles these deformations is vital for any facial recognition system.
3. The analysis of the facial biometrics in the context of human behavior interpretation, is still in its infancy. Facial data acquired by advanced biometric-sensing devices, is

processed and analyzed to extract the features, required for decision making regarding the facial expression, or identity of the individual. However, the decision-making support requires conveying this information to the system personnel, in order to support dialogue, or provide recommendation in order to conduct an interview. This assumes information in a semantic (word) form, in order to support the decisions made by the system personnel during screening, monitoring, or interviewing. A system with enhanced functionality addresses a concept of the new generation of biometric-based systems, namely, the systems with situational awareness, and recommender-based systems.

The problem of joint face recognition and facial expression recognition comprises the following problems:

- facial feature detection which are suitable for FR and FER; and
- feature extraction for joint expressions and identity recognition

An additional problem considered in this thesis, is feature selection from infrared images for further fusion with visible spectrum data.

1.3 Objectives

The primary objectives of the proposed research are:

1. Development of a novel approach to joint facial expression and identity recognition, through joint facial feature detection, and new facial appearance and expression encoding for facial expression and face recognition.
2. Design and implementation of the components (software modules) for facial biometric processing, in the context of decision-making support for physical access control systems with situational awareness.

In order to accomplish these objectives, we conducted:

- Extraction of fiducial points for analysis of controlled and uncontrolled facial images (including head rotation, having eyeglasses, illumination variation, etc.), based on the proposed filtering techniques.
- Encoding facial texture and the geometric information to achieve a robust feature sets for facial expression analysis.
- Developing a real-time facial expression analysis.
- Integrating face recognition with the facial expression to speed up the system and reduce computational cost.
- Multiple tests to verify the system performance on different databases with various image qualities.
- Analysis of infrared facial expressions.
- Feasibility study of application of the proposed technique for facial expression recognition from video.
- Fusion of Infrared and visible facial images to improve recognition rate and reliability.

1.4 Contributions

The main contributions of this thesis are:

- Robust facial feature detection under various conditions: this step is the key step to select the proper facial features in order to construct geometric features for both FR and FER. In a sense, this step provides the key points of a face to be encoded for facial biometric. In this thesis, a new technique based on modified Frangi filter is proposed. The results are presented in [8] [9].

- Facial feature representation and selection: the key part of this system is to select and generate effective features which can be used for both FR and FER. The features should be identical in order to keep the integrity of the system. The feature should also be robust against face deformation, and environment variations. A novel feature extraction based on Gauss-Laguerre wavelet is proposed in this thesis; it is published in [10] [11] [12].
- Feature fusion: in order to use the benefits of different imaging spectra, the infrared face images are used for facial expression. There are few works available on the infrared facial expression analysis, as the texture of the face appears to be “flat” in the infrared images, making it difficult to extract some rich features out of it. This research suggests using the same wavelet-based feature extraction approach as applied for visible face images; it is published in [13] [14].
- Decision-making and interview support: this part is presented as a sample protocols generated as a response in “semantic” form, thus assisting the system personnel in the decision-making regarding authorization. A numerical example along with several generated sample questions are provided in this thesis; this part is addressed in the publications [15] [16] [17].
- Facial expression recognition from video: in this thesis, preliminary results on expression recognition out of video are presented. The main focus of this thesis is image analysis for facial biometrics; however, in order to show the extendibility of the proposed system, we applied the system to video as well. Since the nature of video and image is totally different, it is not feasible to use the exact same approach for video analysis; we showed that the key elements for feature extraction stay the same.

Figure 1-1 shows the flowchart of the whole system that is proposed in this thesis.

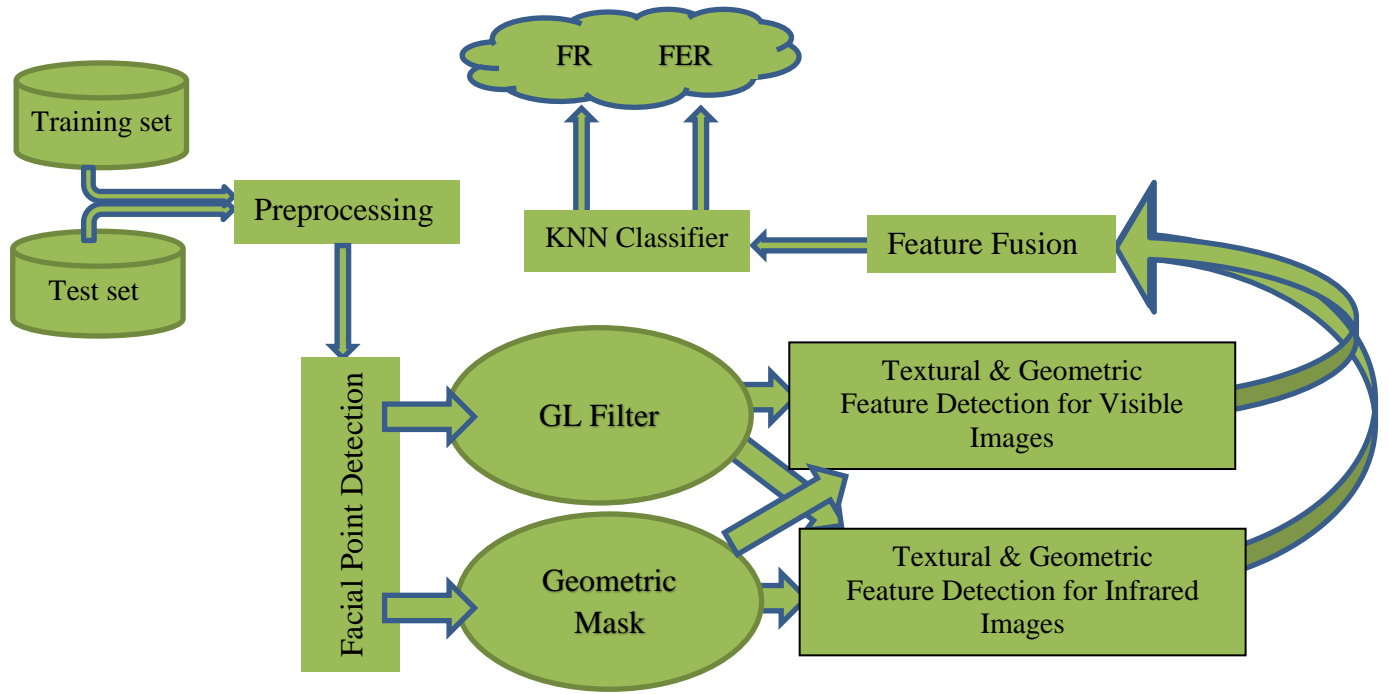


Figure 1-1. The flowchart of facial biometric analysis in this thesis.

1.5 Thesis organization

Chapter 1 outlines the project background and objectives.

Chapter 2 gives an overview of related work, and the introduction to facial expression recognition.

Chapter 3 presents a novel facial feature extraction technique based on Frangi filter; it is accompanied by the experimental results.

Chapter 4 describes the proposed method on facial expression recognition for still images, as well as some preliminary results of application to video processing.

Chapter 5 addresses face recognition for faces with expressions, using the approach presented in chapter 4.

Chapter 6 presents the infrared image analysis applied for facial expression recognition, and the fusion of the infrared and visible facial biometrics.

Chapter 7 contains the concept of biometric-based situational awareness system with the examples of interview supporting protocols.

Chapter 8 concludes this thesis and outlines the future work.

1.6 Publications

A. Poursaberi, M. Spicher, H. Ahmadi, S.N. Yanushkevich, *Global Gauss-Laguerre Wavelets Feature Selection for Facial Expression Recognition*, 8th International Conference on Digital Technologies, Slovak Republic, 2011.

A. Poursaberi, S.N. Yanushkevich, M.L. Gavrilova, *Modified Multiscale Vesselness Filter for Facial Feature Detection*, IEEE 4th International Conference on Emerging Security Technologies, UK, pp. 21-24, 2013.

A. Poursaberi, H. Ahmadi, S.N. Yanushkevich, M.L. Gavrilova, *Gauss-Laguerre Wavelet Textural Feature Fusion with Geometrical Information for Facial Expression Identification*, EURASIP Journal of Image and Video Processing 2012: 17 (2012).

A. Poursaberi, J. Vana, S. Mracek, R. Dvora, S.N. Yanushkevich, M. Drahansky, V. Shmerko, M.L. Gavrilova, *Facial Biometrics for Situational Awareness Systems*, IET Biometrics Journal, vol . 2, issue. 2, 35-47, 2013.

A. Poursaberi, S.N. Yanushkevich, M.L. Gavrilova, *An Efficient Facial Expression Recognition System in Infrared Images*, IEEE 4th International Conference on Emerging Security Technologies, UK, pp. 25-28, 2013.

J. Vana, S. Mrácek, M. Drahanský, **A. Poursaberi**, S.N. Yanushkevich, *Applying Fusion in Thermal Face Recognition*, BIOSIG, Germany, pp. 1-7, 2012.

A. Poursaberi, M.L. Gavrilova, S.N. Yanushkevich, *Fusion of Infrared and Visible Images for Facial Expression Recognition*. (Submitted to IEEE face and Gesture Recognition 2015).

O. Boulanov, M.L. Gavrilova, **A. Poursaberi**, M. Spicher, V.P. Shmerko, S.N. Yanushkevich, *Biometric-Based Intelligent Agent Systems*, Intelligent Systems and Agents 2011, Italy.

K. Lai, **A. Poursaberi**, S.N. Yanushkevich, *One-Shot Facial Feature Extraction Based on Gauss-Laguerre Filter*, CCECE Canadian Conference on electrical and Computer Engineering, 2014, Canada.

A. Poursaberi, S.N. Yanushkevich, M.L. Gavrilova, V.P. Shmerko, P.S.P. Wang. *Situational Awareness through Biometrics*, IEEE Computer 46(5): 102-104 (2013).

Chapter Two: **FACIAL EXPRESSION RECOGNITION: LITERATURE REVIEW**

This chapter presents an overview of facial expression definitions and the available algorithms for expression analysis. A FER system typically consists of four main steps:

1. Preprocessing which includes face detection, face normalization and alignment, and image enhancement,
2. Facial points detection (in both still images and videos) and tracking (in videos),
3. Feature extraction,
4. Facial expression classification.

Figure 2-1 shows the flowchart of a typical system. The face detection step aims to locate the face (or faces) in an image or video. Normally the rectangle around the face determines the face location. Face normalization follows face detection and is done by scaling each image to a fixed size for different databases. After this step, the normalized image has fixed distance between eyes. Some techniques use image enhancement to improve the image quality and remove noises. Face alignment is then applied to the faces to reduce the effects of rotation.

The next step is facial points detection. In many approaches the important facial features like the eyebrows, eyes, nose and the mouth play important roles in FER. Locating these face components requires accurate and fast techniques. The crucial step is facial feature extraction and aims to extract features from the given face image that can be used to classify expressions. These features can be extracted from either the textural or geometric features. Finally a classifier categorizes the input image to one of the expressions.

2.1 Introduction

Automatic facial expression recognition (AFER) is of interest to researchers, because of its importance for facial biometric-based intelligent support systems. It provides a behavioral

measure to assess emotions, cognitive processes and social interaction [18]. Examples of applications of AFER include robotics, human-computer interface (HCIs), behavioral science, animation and computer games, educational software, emotion processing, and fatigue detection. Due to multiple limitations and difficulties such as occlusion, lighting conditions and variation of expressions across the population or even for an individual, having an automatic system helps in creating intelligent visual media for understanding different expressions. Moreover, this understanding helps in building meaningful and responsive HCI interfaces. FER system contains three main functions: face detection and tracking, feature extraction, and expression classification.

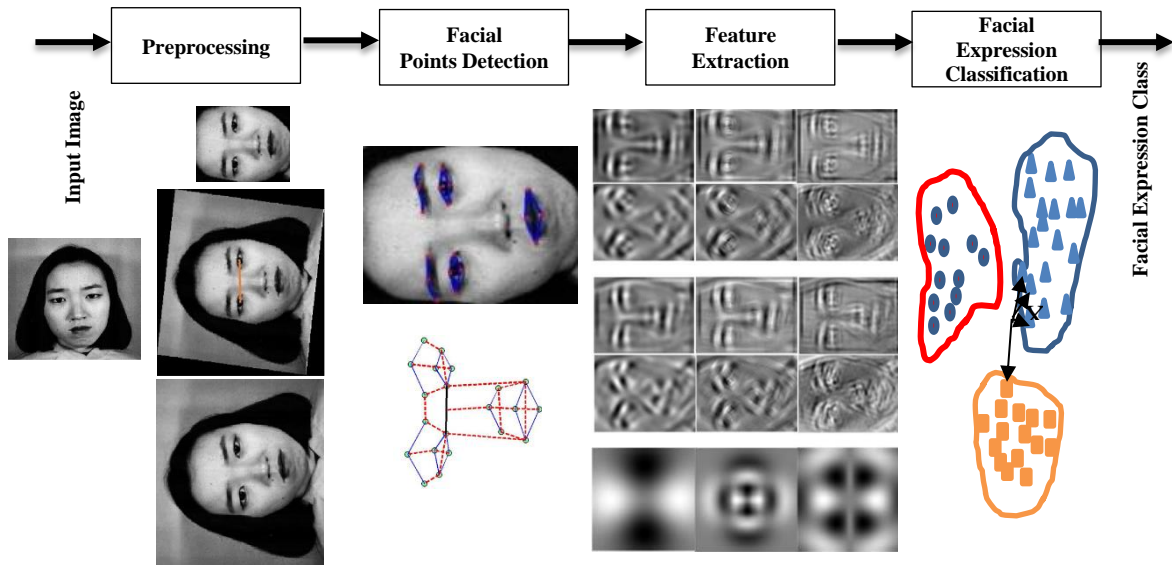


Figure 2-1. Typical facial expression recognition system with four major modules

The first attempt towards the FER was taken in 1978 by Suwa et al. who presented a system for facial expressions analysis from video and tracking twenty points as features [19]. Before that, only two ways existed for FER [20]: (a) human observer based coding system [21] which is subjective, time consuming, and hard to standardize, and (b) electromyography (EMG) based

systems [22] which is invasive (needs sensors on the face). The muscle actions result in various facial behaviors and motions, and later on can be used to represent the corresponding facial expressions. These assumptions became the basis for developing the following systems for coding multiple facial expressions and emotions:

5. The Facial Action Coding System (FACS) — Ekman and Friesen [23].
6. The Facial Animation parameters (FAPs) — MPEG-4 standard, SNHC [24] [25].

The FACS describes facial expressions in terms of action units (AUs) [26]. There are two main approaches to AU recognition:

- Processing 2D static images;
- Processing image sequences.

The first one, which is more difficult than image sequence since less information is available, often uses feature-based methods. Using only one image for expression recognition needs robust and highly distinctive features to cope with variations in human subjects or imaging conditions [27] [28] [20] [29] [30] [31]. There are several methods to process still images. One of them is PCA-based holistic representations combined with feed forward neural networks (NN) for classification proposed by Cottrell and Metcalfe [32]. Chen and Huang [33] used a clustering based feature extraction to recognize only three facial expressions. Eigenface feature extraction accompanied by principal component analysis (PCA) is proposed by Turk and Pentland [34]. Shin et al. [35] used the combination of 2D learning discriminant analysis and support vector machine for classification. Holistic representations and neural networks are applied to the pyramid-structured images by Rahardja et al. [36]. Feng et al. [37] applied local binary pattern for feature extraction, and used a linear programming technique as classifier. Deformable models were utilized by Lanitis et al. [38] to capture variations in shape and grey-level appearance. In

the second approach, an image sequence displays one expression. Neutral face is used as a baseline face, and FER is based on the difference between the baseline and the following input face image. Preliminary work on facial expressions by tracking the motion of twenty identified spots has been done by Suwa et al. [19]. Motion tracking of facial features in image sequences is performed by optical flow, and expressions are classified into six basic classes [39]. Yacoob and Davis [40] extracted local edges around mouth, nose, eyes, and eyebrows and their motion were used as features. Fourier transform was utilized for feature extraction, and a fuzzy C-means clustering was applied to build a spatiotemporal model for each expression in [41]. Bartlett et al. [42] used the combination of optical flow and PCA obtained from the image differences.

Facial coding is normally performed in two different ways: holistic and analytic. In holistic approach, face is treated as a whole. Different methods are presented in this approach including: optical flow [43], Fisher linear discriminates [44], neural network [45], active appearance model [38], and Gabor filters [46]. In analytic approach, local features are used instead of the whole face, namely, fiducial points describe the position of important points on face (e.g., eyes, eyebrows, mouth, nose, etc.), together with the geometry or texture features around these points [47].

Gabor filters are widely used in texture analysis. These filters model simple cells in the primary visual cortex. Zafeiriou and Pitas [48] showed the best performance of Gabor filters in both analytic and holistic approaches. Gabor filters have been used for expression classification in [19]. Although the Gabor filters show high performance in FER, the main problem using this filter is how to select the optimum one, in terms of scale and orientation. The number of filters for convolution depends on the applications, and usually 40 filters (5 scales and 8 orientations) are used [31]. Because of the large number of convolution operations, it needs large amount of

memory and computational cost. Moreover, with the small training samples, the dimensionality is really high [49].

Normally, two types of facial features are used: permanent and transient. Permanent features include eyes, lips, brows and cheeks, and transient features include facial lines, wrinkles and furrows. The eyebrows and the mouth play main roles in facial expressions. Pardas and Bonafonte [50] showed that expressions such as surprise, joy and disgust have much higher recognition rate, since clear motion of the mouth and the eyebrows are involved.

2.2 Basic emotions

There are six different expressions in addition to the neutral expression which are normally discussed in the facial expression recognition. To assign different expressions to specific emotion, we have to know, which class of human emotions can be addressed by each expression. In 1971, Ekman and Friesen [26] suggested six primary emotions, which each of them has a distinctive content together with a unique facial expression. In literatures, these classifications are also addressed as “basic emotions”. The key feature is that these emotions should be universal regardless of human ethnicities and cultures. Happiness, fear, disgust, sadness, surprise and anger are the basic expressions proposed by Ekman and Friesen, although many other categories are also known and being used by scientists, psychologists and people who are concerned with emotion. As mentioned earlier, the well-known categories like FACS and FAPs are standard scientific tools in the field of expression recognition. Figure 2-2 shows sample of each expressions. The following definitions are borrowed from [25].

2.2.1 Happiness

It is the easiest expression which can be recognized in any face (or image) regardless of age, gender or culture. Happiness is an emotional state of well-being described by pleasant emotions.

It is normally followed by a situation (condition) of pleasure, amusement, enjoyment and satisfaction. Finding happy faces even in a crowded scene can be the easiest among all other emotions simply by seeing “smile” on a face. But sometimes even a happy person, although has a smile on his/her face, does not exactly comply a true happiness, but maybe is hiding other emotions and deceiving other people. A good example of this situation can be when a politician or a celebrity smiles in front of cameras on television. Thus, a valid detection of “genuine” happy expressions could be beneficial in some applications.



Figure 2-2. Different expressions from CK [51] database (top to bottom, left to right): anger, disgust, fear, happiness, sadness, and surprise.

2.2.2 Fear

Fear is an unpleasant emotion caused by a threat or imminent danger that induced a change in brain followed by a behavior like running, screaming or hiding. In this case, the specific objects that can produce fear vary. Fear is also categorized as rational (appropriate) and irrational (inappropriate). An example of irrational fear is having phobia to a phenomena. The fear expressions vary a lot based on the degree of anxiety/concern, gender and different cultures. Other symptoms of fear could be heart-pounding, short-breathing, and wide-open eyes.

2.2.3 Disgust

Disgust expressions correspond to feeling of revulsion or intense displeasure stimulated by unpleasant events (behavior) or offensive action. It is the body's reaction to objects that are nauseating like seeing a hair in your food, spoiled foods, or an unpleasant smell. Women generally report greater disgust than men especially during their pregnancy. Since disgust is an emotion with physical reactions to unpleasant or dirty situations, studies have shown that there are cardiovascular and respiratory changes while expressing disgust [52].

2.2.4 Surprise

Surprise expressions normally occur when something unexpected, sudden, or amazing happens. Surprise expression is frequently followed by other emotions such as happiness or fear depending on the event that causes the surprise reaction. For instance, if you have been surprised by your friends on your birthday, a typical after-emotion would be happiness. Other mixed-surprise emotion situation is when someone suddenly appears, and you may fear after being surprised. Thus, surprise can have any valence and can be pleasant/unpleasant, positive/negative. Surprise is different from startle and their expressions are also totally different.

2.2.5 Sadness

Sadness is often the opposite of happiness and is caused by pain, loss, sorrow, discomfort and helplessness. In some cases, sad expression is followed by tears and weeping. Extreme and constant sadness is called depression. Detecting sadness expressions is sometimes difficult due to its variation with ages, gender and culture.

2.2.6 Anger

One of the most daily experienced emotions that people have these days because of daily stresses in modern society. Anger is an emotional reaction to the situation where a person found him/her

offended, misunderstood, wronged, or denied. It can convey messages about hostility, opposition, and potential attack [25]. Anger may have physical symptoms such as increased heart rate, blood pressure, and levels of adrenaline and noradrenaline [53]. The signs of anger expressions can be found in facial expressions, body language, acts of aggression or physiological responses. Anger can be also a good thing if it can give you a way to find a solution to problems or to express negative feelings.

2.3 Facial Action Coding System (FACS)

Facial expressions are the outcome of both facial muscle and autonomic nervous system actions, but the question is whether a quantitative system exists to describe all actions the face can perform. Ekman and Friesen [26] proposed FACS which could measure all visible facial movements. In practice, FACS differentiates every visible change in muscular action but not invisible changes like vascular changes produced by the autonomic nervous system. The main goal of FACS is to provide a “comprehensive” system that measures all possible visually discriminable facial actions. It was the outcome of anatomical analysis of facial movement and how each muscle of the face acts to change visible appearances. Ekman and Friesen distinguished six emotions: anger, disgust, fear, happiness, sadness and surprise, as being “discriminable within any one literate culture”. Sometimes, a neutral expression is considered as a seventh expression. The FACS describes facial expressions in terms of action units (AUs). FACS decomposes the facial expression into individual AUs and explains how to identify different facial expressions based on application of varying facial muscles individually or in groups. One of Ekman and Friesen goals was to build a system, which could distinguish all possible visually distinguishable facial movements. They analyzed the anatomical basis of facial movements, the muscles. Every muscular action results in a facial movement. FACS contains 46

AUs, which are basic facial movements, corresponding to different muscle activities, but more than 7000 combinations among them have been observed. FACS ignores invisible changes and discards visible changes too subtle for reliable distinction. Since FACS was developed to measure only movement of the face, other visible phenomena, such as facial sweating, tears, rashes, pimples and permanent facial characteristics, were all excluded from FACS.

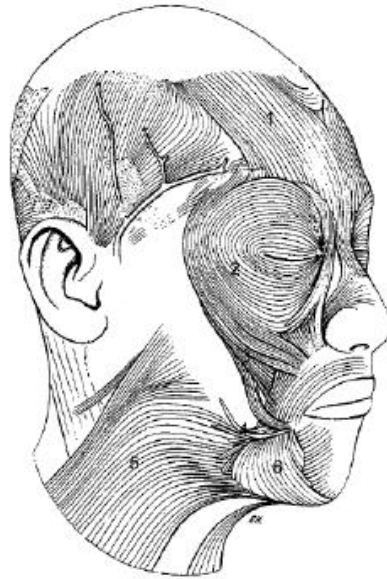


Figure 2-3. Muscles of facial expression. 1, frontalis; 2, orbicularis oculi; 3, zygomaticus major; 4, risorius; 5, platysma; 6, depressor anguli oris [54].

Table 2-1 contains all different AUs definitions. The definition of facial muscles is shown in Figure 2-3 [55]. The AUs can be assigned to Upper face and Lower face. The Upper face normally contains the eyebrows, eyes and forehead. The Lower face includes the lips, nose and chin. Out of all AUs, 30 of them are obtained by contractions of facial muscles where 12 belong to Upper and the rest to Lower face [56]. The combinations of AUs in Upper and Lower are also used for expression recognition. Figure 2-4 and Figure 2-5 show examples of each class which are borrowed from [57].

Action unit combinations may be additive, in which case the combination does not change the appearance of the constituents, or non-additive, such that the appearance of the constituents change [56]. An example is shown in Figure 2-6 for the combination of AU1 and AU4; the eyebrows are raised because of AU1 which affects the AU4 (lowered eyebrows). Thus, AU1+AU4 is an example of a non-additive combination.

























NEUTRAL	AU 1	AU 2	AU 4	AU 5
				
Eyes, brow, and cheek are relaxed.	Inner portion of the brows is raised.	Outer portion of the brows is raised.	Brows lowered and drawn together	Upper eyelids are raised.
AU 6	AU 7	AU 1+2	AU 1+4	AU 4+5
				
Cheeks are raised.	Lower eyelids are raised.	Inner and outer portions of the brows are raised.	Medial portion of the brows is raised and pulled together.	Brows lowered and drawn together and upper eyelids are raised.
AU 1+2+4	AU 1+2+5	AU 1+6	AU 6+7	AU 1+2+5+6+7
				
Brows are pulled together and upward.	Brows and upper eyelids are raised.	Inner portion of brows and cheeks are raised.	Lower eyelids and cheeks are raised.	Brows, eyelids, and cheeks are raised.





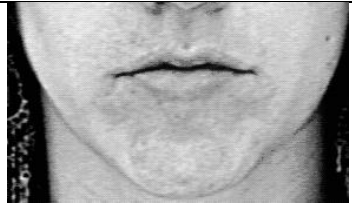



Figure 2-4. Upper face AUs and combinations [56].



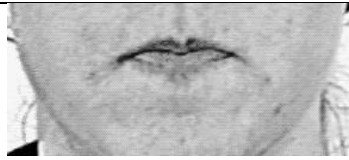



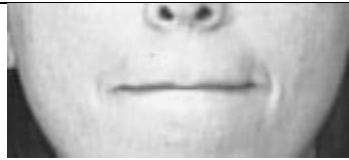

2.4 Facial Animation Parameters (FAPs)




The MPEG-4, which is a multimedia compression standard to separately encode each objects in a scene, contains both face and facial expression modeling. The face is described with a 3D model and the expressions by facial animation parameters (FAPs) [58].

Table 2-1. AU definition. The images borrowed from [55]

AU	Description	Facial muscle	Example image
<u>1</u>	Inner Brow Raiser	<i>Frontalis, pars medialis</i>	
<u>2</u>	Outer Brow Raiser	<i>Frontalis, pars lateralis</i>	
<u>4</u>	Brow Lowerer	<i>Corrugator supercilii, Depressor supercilii</i>	
<u>5</u>	Upper Lid Raiser	<i>Levator palpebrae superioris</i>	
<u>6</u>	Cheek Raiser	<i>Orbicularis oculi, pars orbitalis</i>	
<u>7</u>	Lid Tightener	<i>Orbicularis oculi, pars palpebralis</i>	
<u>9</u>	Nose Wrinkler	<i>Levator labii superioris alaeque nasi</i>	
<u>10</u>	Upper Lip Raiser	<i>Levator labii superioris</i>	
11	Nasolabial Deepener	<i>Zygomaticus minor</i>	

<u>12</u>	Lip Corner Puller	<i>Zygomaticus major</i>	
13	Cheek Puffer	<i>Levator anguli oris (a.k.a. Caninus)</i>	
14	Dimpler	<i>Buccinator</i>	
<u>15</u>	Lip Corner Depressor	<i>Depressor anguli oris (a.k.a. Triangularis)</i>	
16	Lower Lip Depressor	<i>Depressor labii inferioris</i>	
<u>17</u>	Chin Raiser	<i>Mentalis</i>	
18	Lip Puckerer	<i>Incisivii labii superioris and Incisivii labii inferioris</i>	
<u>20</u>	Lip stretcher	<i>Risorius w/ platysma</i>	

22	Lip Funneler	<i>Orbicularis oris</i>	
<u>23</u>	Lip Tightener	<i>Orbicularis oris</i>	
<u>24</u>	Lip Pressor	<i>Orbicularis oris</i>	
<u>25</u>	Lips part	<i>Depressor labii inferioris or relaxation of Mentalis, or Orbicularis oris</i>	
<u>26</u>	Jaw Drop	<i>Masseter, relaxed Temporalis and internal Pterygoid</i>	
<u>27</u>	Mouth Stretch	<i>Pterygoids, Digastric</i>	
28	Lip Suck	<i>Orbicularis oris</i>	
41	Lid droop	<i>Relaxation of Levator palpebrae superioris</i>	

42	Slit	<i>Orbicularis oculi</i>	
43	Eyes Closed	<i>Relaxation of Levator palpebrae superioris; Orbicularis oculi, pars palpebralis</i>	
44	Squint	<i>Orbicularis oculi, pars palpebralis</i>	
45	Blink	<i>Relaxation of Levator palpebrae superioris; Orbicularis oculi, pars palpebralis</i>	
46	Wink	<i>Relaxation of Levator palpebrae superioris; Orbicularis oculi, pars palpebralis</i>	

FAPs [59] allows to animate the synthetic face models which is the results of study on the muscle actions. It models the neutral face with a mesh using 84 well-defined facial feature points and 66 low-level FAPs. In this definition, the high-level features are expressions and low-level features are displacements of a point of the face. FAPs can represent a precise model for the expression evolution on the face based on two mentioned feature classes. This representation includes head motion, tongue, eyes, and mouth movement. FAP units are also used in the modeling which is defined as the fractions of distances between key points on the face. An example is shown in Figure 2-7.

FAPs have been used by many researches, because of its compliance with the MPEG-4 standard and providing good representation of the facial expression development [60]. There are some drawbacks associated with this system. One is its robustness of facial points tracking in video, as the positions are prone to be changed [61]. This approach is stable on still images, and have shown promising results on FER.























<i>NEUTRAL</i>	AU 9	AU 10	AU 12	AU 20
				
Lips relaxed and closed.	The infraorbital triangle and center of the upper lip are pulled upwards. Nasal root wrinkling is present.	The infraorbital triangle is pushed upwards. Upper lip is raised. Causes angular bend in shape of upper lip. Nasal root wrinkle is absent.	Lip corners are pulled obliquely.	The lips and the lower portion of the nasolabial furrow are pulled pulled back laterally. The mouth is elongated.
AU15	AU 17	AU 25	AU 26	AU 27
				
The corners of the lips are pulled down.	The chin boss is pushed upwards.	Lips are relaxed and parted.	Lips are relaxed and parted; mandible is lowered.	Mouth stretched open and the mandible pulled downwards.
AU 23+24	AU 9+17	AU9+25	AU9+17+23+24	AU10+17
				
Lips tightened, narrowed, and pressed together.				
AU 10+25	AU 10+15+17	AU 12+25	AU12+26	AU 15+17
				
AU 17+23+24	AU 20+25			
				

Figure 2-5. Lower AUs and combinations [56].

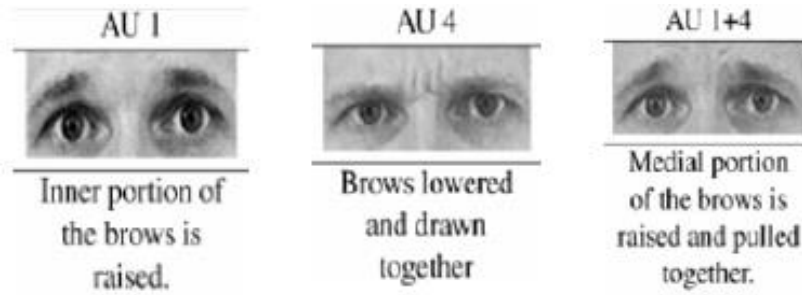


Figure 2-6. A non-additive combination example [56].

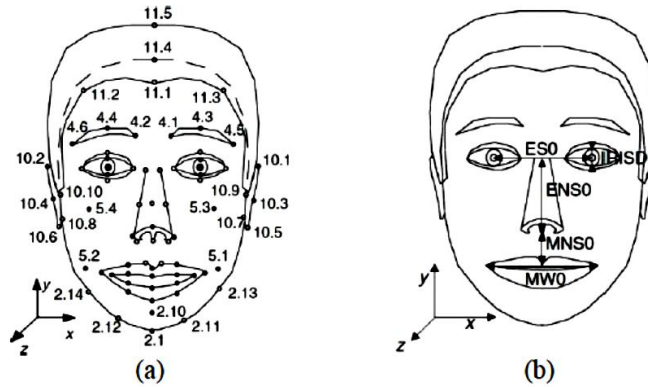


Figure 2-7. Facial animation parameters: (a) Examples of facial points defined in FAP MPEG-4, (b) FAPU example based on the defined distances [58].

Another problem using FAPs is its complexity in interpretation. FAPs can be explained simply as a “geometric mesh model” based on muscle actions. A single muscle action can trigger different FAPs. As an example, in AU12 in FACS (smile), two FAPS (left and right corner of the mouth) are triggered, which, in a complicated model, brings confusion. Another issue is that the system can only encode the “visible” facial feature points, and for some expressions, which just deform the shape of the skin without moving any facial feature points (deepening the nasolabial furrow), it cannot be used [58].

2.5 Automatic facial expression analysis modeling

Automated recognition of AUs, given an image or a video, is not a trivial task. There are two main approaches to AU recognition:

1. Processing of 2D static images
2. Processing of image sequences

Processing of the face images acquired in a varying environment, is very challenging. For example, a 2D face appearance can change significantly under different viewing angles and illuminations. These problems are unavoidable in surveillance applications under uncontrolled environment. In general, two key stages are mandatory for an automatic AU recognition system:

- The facial feature extraction stage, and
- The AU classification stage.

A great number of methods have been proposed in literature [62]. The approaches on AU classification are grouped into:

- Spatial approaches, where AUs are analyzed frame by frame, and
- Spatio-temporal approaches where AUs are recognized over time based on the temporal evolution of facial features.

The difference between these two methods can be either in feature extraction techniques, or the AU classification techniques, or both. However, they all classify each AU or certain AU combinations independently, and statically ignoring the semantic relationships among AUs and their dynamics. The relationship among the AUs, as well as the temporal development of each AU, in order to improve AU recognition was proposed in [62].

It is very rare that a single AU occurs alone in spontaneous facial behavior. Instead, some patterns of AU combinations appear frequently to express natural human emotions. This means,

it is almost impossible to perform some AUs simultaneously, as described in FACS. In addition to the dynamic development of each AU, the relationships among AUs undergo a temporal evolution to reflect the evolution from a neutral state to a weak emotion, then to the apex, and, finally, to a releasing state. The dynamic and semantic relationships can be well modeled, making it capable of representing the relationships among different AUs. The automatic AU recognition system consists of:

- An offline training phase, and
- An online AU recognition phase.

The system training includes training classifier for each AU, in order to correctly model the AU relationships. First, the face and eyes are detected in live video automatically. Then, the face region is divided into upper and lower parts, which are aligned, based on the detected eye positions and the features are extracted using different approaches (e.g., Gabor filters, wavelets) to produce features for each pixel. After that, the classifier combines the features to produce a measurement score for each AU. These scores are then used for final classification.

2.6 Conclusion

This chapter provided an introduction into facial expression recognition, using the established system of expression classification. The six major expressions are usually characterized using the two main coding techniques: FACS and FAPs. Some papers use AU coding for emotion classification, whereas others use classical approaches, such as textural and geometric features. Each of them includes different algorithms to accomplish the facial expression classification. In this thesis, we use the classic approach for feature extraction, while introducing new features. As we will discuss in Chapter 6, the key point is how to integrate both face recognition and expression classification, and perform them simultaneously; this is recently being of interest to

researchers. We will show how the Gauss-Laguerre wavelets can provide rich textural features that are merged into geometric features, suitable for joint FER and FR. In the next chapters, we will review the important works for every single steps of the joint FER and FR, and then present our techniques.

Chapter Three: **FACIAL FEATURE DETECTION**

The aim of this chapter is to propose a solution for facial points detection which are important for the purpose of feature extraction and tracking. The goal of this chapter is to investigate how these facial features can be extracted in an accurate and fast manner. These features will be used in both face recognition and facial expression recognition. The proposed approach used the modified Frangi filter [63] to detect the eyebrows, eyes, nose and the mouth. The original filter is normally used in medical image enhancement to enhance vessel structures to be used for segmentation. The proposed method has been tested on various datasets and with different image qualities.

3.1 Introduction

A facial expression recognition normally involves four different steps: image acquisition, preprocessing, feature extraction, and classification. Preprocessing is performed before feature extraction for FER, in order to increase the system performance. The aim of this step, which includes scaling, intensity normalization, and size equalization, is to have images that only contain a face, expressing certain emotion. Sometimes, histogram equalization is also used to adjust image brightness and contrast. In this chapter, we focus on the preprocessing which contains face detection and facial points detection. In particular, the proposed method in fiducial feature detection is elaborated.

Automatic facial feature detection is a crucial step in many applications of computer vision. Some examples of applications include: face recognition, facial expression analysis, human-computer interaction, and any other new applications that can benefit from facial feature detection. The performance of these applications is highly dependent on the accuracy of detected

points. Therefore, having a reliable and robust algorithm is very important. Among the facial features, eye detection has attracted growing interest in research, since other features can be localized based on eyes location. The position of the eyes is used to normalize face for applications such as face recognition and facial expression recognition. Moreover, it was shown that the accuracy of some famous face recognition methods (e.g. PCA/LDA) is degraded by poor eye detection [64]. Various face feature detection methods have been reported in literature. For eye localization, there are several commonly used methods, including: hand-built templates, Gabor filter, AdaBoost eye regional detection, and support vector machines (SVM). In particular, discriminative methods based on boosting became more popular, because of their potential computational efficiency [65]. Facial feature detection methods can also be categorized generally into several groups [65]: template matching (using intensity and edge information), knowledge-based (using geometrical facial features), feature invariant (using different edge detectors), and appearance-based methods (using gray level of face images). In most approaches, the extraction of facial features depends on the fact that they correspond to low-intensity regions of the face. For instance, in eye detection, this is due to the color of the pupils and eye-sockets. In most approaches, face detection is executed first, in order to apply feature detection techniques within the detected face. The other methods search for facial features within the entire image. In this case, facial geometry and size of features are taken into account to filter the outliers out. These approaches have more errors in terms of localization.

The available approaches for facial feature detection can also be classified into two main categories: texture-based and shape-based methods. In the texture-based method, the local neighborhood around a certain point in a face (e.g., eyebrow's corner) is modeled. However, the shape-based methods utilize all facial points as a shape and try to find (or fit) the right shape for

a given unknown face. This is done based on the pre-trained step using set of labelled faces with the facial points.

Some texture-based approach use: hierarchical wavelet networks [66], log-Gabor wavelet and geometry cross-ratios relationships [67], neural network-based eye detector [68], feature patch template [69], boosted regression coupled with Markov networks [70]. A two-level hierarchical wavelet network for localization of eight facial features is presented in [66]. The method uses a coarse-to-fine approach to localize small features, using cascading sets of Gabor wavelet Networks (GWN) features. At the first level, a wavelet network is used to find the whole face (face matching). A wavelet network consists of a set of wavelets and their weights. Next, a group of 2D Gabor wavelets are utilized to fine-tune the eight feature locations.



Figure 3-1. The approach proposed in [66] uses GWN features: first row, left to right: face image, GWN representation of a face, GWN features. Second row: sample images with detected features.

Log-Gabor wavelets are used in [67] to represent each facial point by embedding local surrounding features. Each facial feature has different response to a sample log Gabor, and this

response is used to locate the same feature in an unknown given face image. Geometric relationships between the facial points are used to remove the outliers and locate the exact facial points. Three sample face images are utilized to determine the log-Gabor representation of each facial point and in total seven facial features have been detected by this approach.

Automatic location of facial features, such as eyes, nose and the mouth are detected in [68] using a neural network. This approach mainly extracts the location of eyes and the so-called micro-features which are the points around the eyes. The features are modeled as a structural assembly of micro-features (or simply facial feature points). The micro-features are defined based on the magnitude and the orientation of the intensity gradient of image, using a neural network. Then, the neural network responses are post-processed by a probabilistic approach by taking into account the geometrical information of the micro-features.

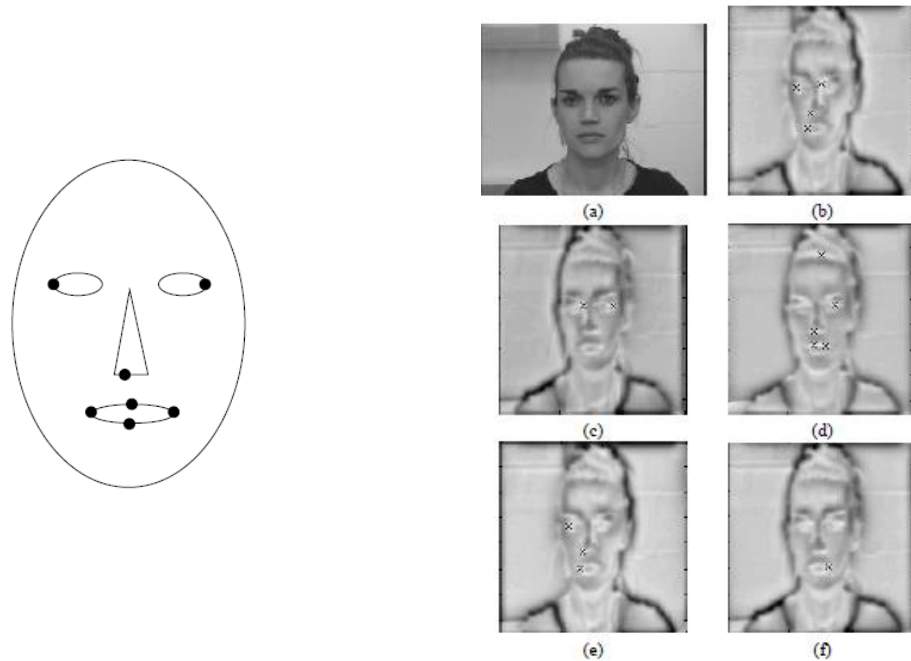


Figure 3-2. The approach proposed in [67] uses the mask and geometrical features. Left: the mask used for geometrical feature relations, Right: The given image (a) and all the feature candidates for eight desired features (b-f).

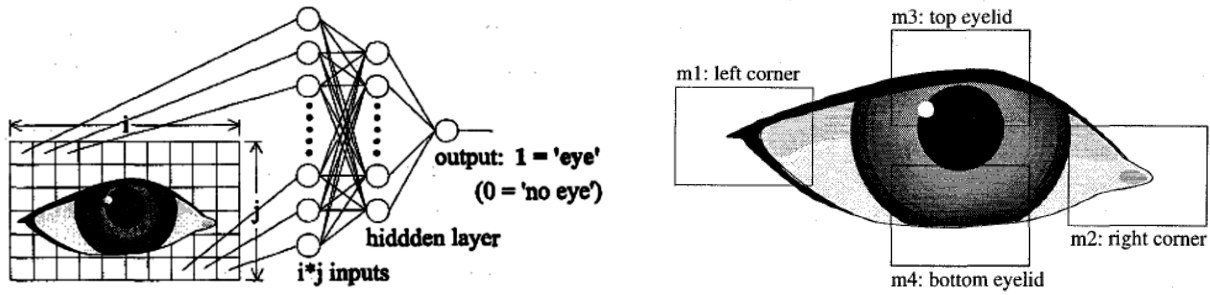


Figure 3-3. Micro-features detection proposed in [68]. Left: Eye detector using neural network and image gradient intensities. Right: The defined micro-features around eye.

Twenty facial features are chosen to be detected using Gabor-based boosted classifiers in [69]. First, face is detected using modified version of well-known Viola-Jones method called the Haar feature based GentleBoost classifiers [71]. The detected face is divided into 20 relevant regions, and so-called feature patches are used to detect the points within the regions. The patches are 13x13 pixel GentleBoost templates built from gray-level images and Gabor wavelet features. The improved version of [69] is presented in [70]. The presented method uses the combined Support Vector Regression and Markov Random Fields to extract 22 facial points. The search space for each facial points is constrained using Markov Random Fields, and the regressors (predictors) learn a mapping between the appearance of the neighbors around the points and their positions. The regression procedure is the process of deriving a function (like $f(x)$) so that it has the least deviation between predicted and experimentally observed responses for all training samples. The advantage of this approach is having a detector which is robust against pose and expressions.

The shape-based methods, such as Active Shape Models (ASM) [72] or Active Appearance Models (AAM) [73], are based on the shape-based facial feature point detection. ASM is a

statistical model of the object shape that is defined by a group of points a set of points, which are controlled by the shape model. The shape can deform iteratively until it fits the object on a given unknown image. The shapes are constrained by the PDM (point distribution model) to vary only in ways seen in a training set of labelled examples.

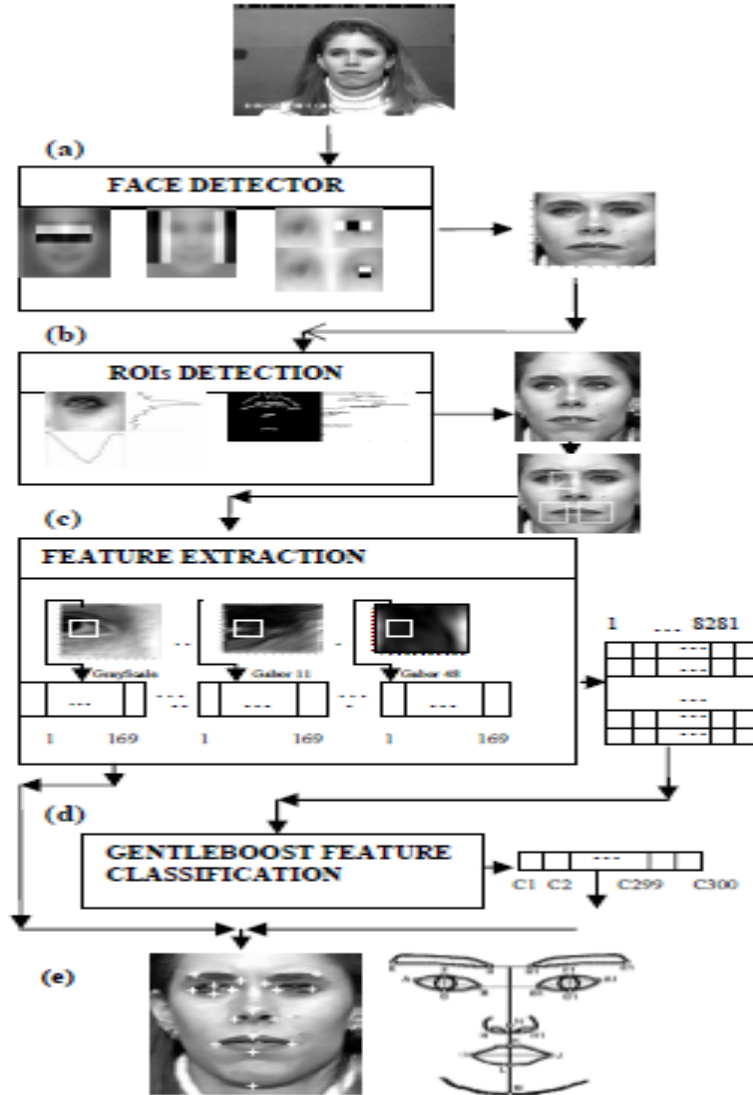


Figure 3-4. The outline of the method presented in [69]. (a) Modified face detector, (b) ROI extraction, (c) Gabor-based feature detection, (d) GentleBoost feature classification, (e) Detected features based on the proposed mask.

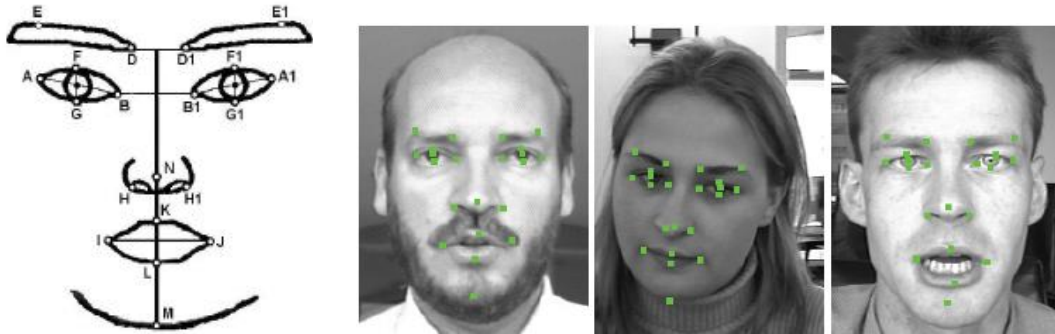


Figure 3-5. Left: Point model of 22 fiducial points suggested in [70]. Right: samples of detected features with different pose/expressions.

AAM uses the same information as ASM, in addition to the gray-level appearance of the model. To match the model with a new unknown image, it is required to find the model parameters, which minimize the difference between the image and a synthesized model example, projected into the image. To perform the task, it is needed, in the training phase, to manually annotate the landmarks on the different images. The variance in the type of pictures and the type of faces enable the AAM to successfully fit the model into the new images. Thus, having a highly diverse collection of images helps to build a stronger model.

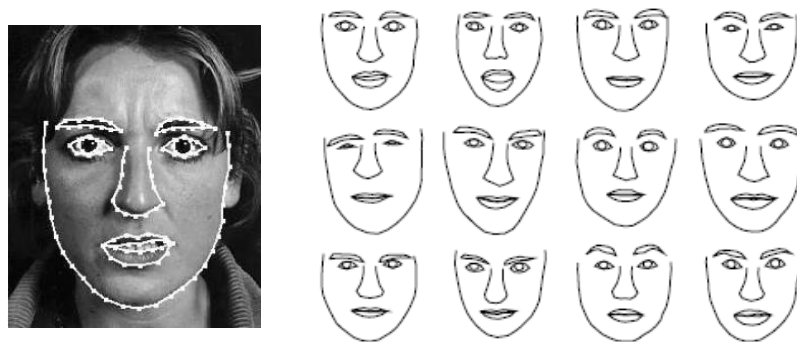


Figure 3-6. Left: Example of ASM. Right: The face image annotated with landmarks [74].

In [75], a two-dimensional facial feature point templates have been introduced, which uses two ASM in series to achieve a higher detection accuracy. Another example of using ASM was provided in [76] where an improved version of ASM first finds the pupil centers as a reference for an initial position of the mask. Then, a new texture model is applied to localize the facial features. This reduces the iteration numbers to fit and find the right shape. A hybrid AAM is introduced in [77] by combining the local skin similarity with the original local grey-level appearance model. Using pre-processing image enhancement, a hybrid AAM by combining the local skin similarity with the original local grey-level appearance model is applied. Next, the Gabor feature around the feature points was extracted and trained by linear discriminant analysis (LDA) and further classified by K-Nearest Neighbor in order to give the precise location of the feature points. The experimental results indicated that 60 facial feature points can be located, but the facial expression variation is not taken into account in this study.

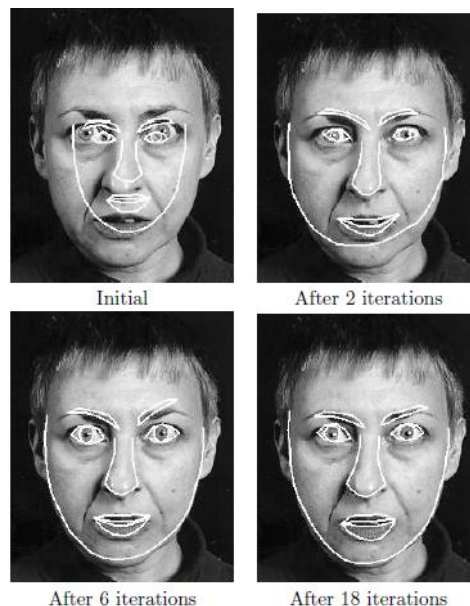


Figure 3-7. Search using Active Shape Model of a face [74].

There are other approaches that use the combined versions of aforementioned techniques. For example, Cristinacce and Cootes [78] applied Haar features with an AdaBoost classifier combined with the ASM in a probabilistic framework. Chen et al. [79] proposed a boosting technique to first extract the facial point candidates for each pixel and then apply a shape model to filter out the outliers.

Although some of the above described approaches reported very good results on various databases with different number of images, the main problem is the need in an “offline” training phase. The training phase requires a huge amount of time, and needs all the images being annotated with precision. Moreover, some of the approaches are not quite suitable in real-time applications due to their computational costs. Another problem of the available methods is that most of them can detect only few facial points with “high accuracy”, such as eyes corners, mouth or nose, but not all essential facial points as depicted in Figure 3-8.

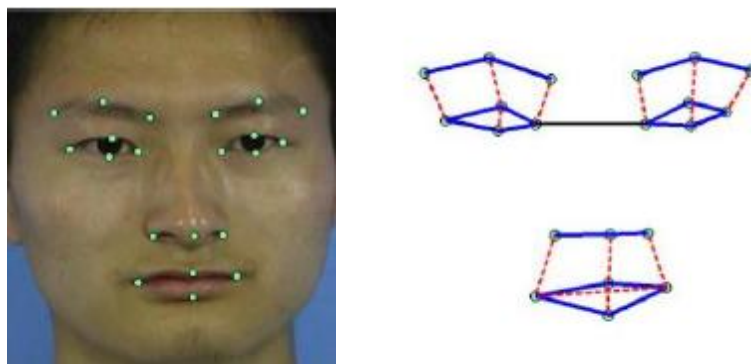


Figure 3-8. Left: Fiducial points overlaid on a neutral expression face from

In many approaches, the point detection is assumed to be successful if the distance between the automatically labelled point and the manually detected one is less than 30% of the true inter-ocular distance (the distance between the eyes), or roughly 30 pixels [80]. This is not an accurate detection as in the available databases like Cohn-Kanade [51], the width of the whole eye is

approximately 50 pixels. In facial expression recognition we are interested in small changes around facial points, and, therefore, having this amount of error can cause serious problems.

3.2 Face detection

In any application, which deals with facial biometrics, such as face recognition, face tracking, eye detection, or facial expression recognition, the first step is to detect the face. The key features of a good face detector are: robustness against orientation, position, and lighting conditions of faces. Nowadays, there are many face detection techniques; however, face detection is still a hot topic. A good survey of face detection techniques is presented in [81].

In this thesis, since our main focus is on facial expressions recognition, we use one of the available face detector [82] which is highly cited in literature because of its simplicity, speed, and accuracy. The mathematical description of this detector that is known as Viola-Jones will be explained along with experimental results.

The available methods can be categorized in four groups [83]: knowledge-based methods, feature-invariant methods, template matching methods, and appearance-based methods. In knowledge-based techniques, several rules are derived based on human faces which can describe the facial features or the relation between these features [84] [85]. This approach is simple but not accurate, as faces are easily deformed in real world because of expressions, illumination and pose. In feature-invariant methods, the facial features which are stable under different poses (eg. eyes), expressions or illumination are taken into account as references and then the face is detected based on these features [86] [87] [88] [89]. The drawbacks of this approach is its performance often undermined due to illumination, noise, and occlusion [83]. In addition, shadows can cause serious problems due to generating dummy edges on face images. In template matching techniques, a mask which contains facial features, is defined. The correlations between

this mask and input image are calculated to verify the existence of a human face [90] [91] [92] [93]. This technique is combined with other approaches to achieve better results. Finally, the appearance-based methods also use the same concept of mask (template), but the appearance characteristics of face are also included in the model. Like the deformable models, these techniques [94] [95] [96] need training phase. These techniques are complicated in nature, and their implementation is time-consuming. As mentioned before, the training phase also needs lots of time and effort.

3.2.1 Viola-Jones face detector

The well-known and fast cascade-based face detection algorithm proposed by Viola and Jones is used in our approach [82]. The publicly available implementation of this face detector, from the OpenCV library, has been utilized. The normalization is done by scaling each image to a fixed size for different databases.

Papageorgiou [97] introduced the rectangle features, which are known as Haar-like features for face detection. The practical implementation of this technique was introduced by Viola and Jones [82]. The proposed real-time object detection technique has been used in many applications due to its performance and simplicity. The algorithm detects face based on its geometric feature values of facial points like eyes. Different types of a rectangle feature can be easily defined, and the feature is called sub-window. Each of these sub-windows is constructed to detect specific facial features based on its geometry. Viola and Jones used four rectangles for face detection as shown in Figure 3-9. These rectangles (Figure 3-9 a, and b) are designed so that they can detect directional features (horizontal/vertical). The boundaries here is to separate the foreground (object) edges from background. For example the rectangle in Figure 3-9 c is used to detect objects which have “complex” texture in the middle and “smooth” texture on its neighbors.

Finally the rectangle in 9d is used for diagonal geometry features. The approach involves finding the sums of image pixels within rectangular areas. The value of any given feature is the sum of the pixels in white rectangles, subtracted from the sum of the pixels within black rectangles. For example the value of the rectangle feature shown in Figure 3-10 is calculated as [98]:

$$V = (A + D) - (B + C)$$

where V is the rectangle feature value, A, B, C and D are the sums of pixel values within four small rectangles. The size and location of this sub-window can be changed.

Using this simple calculation, Viola and Jones consider each of these rectangles with different size and location as a “weak classifier”. To simplify the system and reduce the computational cost, the weak classifiers should be filtered out, and the more “distinctive” shall be combined to form a stronger classifier for face detection. Generally, a learning method called “AdaBoost” is used to select the efficient features and train the classifiers [99]. An example of the face detection using built-in function in MatLab[®] software is shown in Figure 3-11.

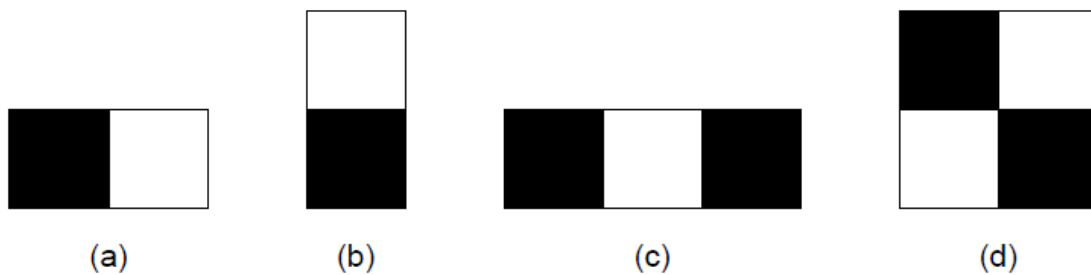


Figure 3-9. Rectangle features defined by Viola-Jones [82].

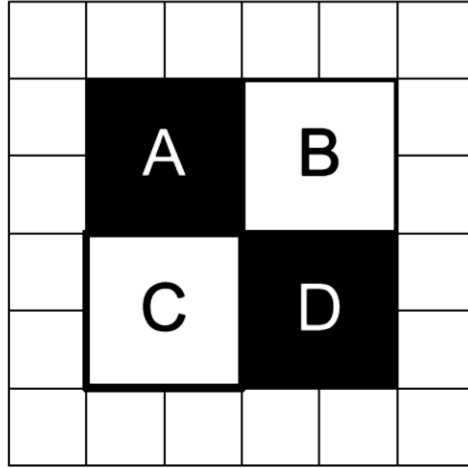


Figure 3-10. An example of rectangle feature where each small rectangle indicates a pixel value. The example is borrowed from [98].

3.3 The proposed method

In this thesis, the 21 facial points illustrated in Figure 3-8 are detected using a novel proposed approach. The whole face is first detected using Viola-Jones method. Next, the two-step approach, coarse and fine detection, is applied to detect pupil eyebrows, mouth and nose regions. Finally, the points are easily extracted from the detected regions in previous step. Although the algorithm can precisely detect all the fiducial points, the proposed system does not rely on the exact locations of these points for feature extraction. The proposed method is straightforward, fast and robust under different pose and illuminations. It has been tested on available face databases with/without expressions to test its performance and is reported in Section 3.5.

In this chapter the proposed facial point detection algorithm is described in details. Then, the experimental results on different databases in addition to the comparison with some available methods are presented.

The face detection example photo has been removed because of copyright restriction.

Figure 3-11. Example of face detection using Viola-Jones method [100].

3.3.1 Facial point detection

The proposed method for facial point detection uses a modified Frangi filter [63] which was proposed by Frangi et al. This original vesselness filter enhances the vessels in medical images due to background suppression properties and multiscale nature. This filter has been modified and extended in this paper, in order to be used as a facial feature detector. We derive an approach that retrieves facial features including: eyes, eyebrows, nose, and the mouth. Our algorithm does not make use of any learnt model, and works with gray-scale images. The method is robust against shape, and texture variations due to expression, and pose changes. This is because the filter is sensitive to the nature of facial features, especially eyes. The two-step eye detection scheme is applied to the localized face. By adjusting the scale factor of the filter, a coarse estimation of eye position is obtained. At the same time, eyebrows, nose, mouth and a few other objects are detected. The geometrical properties of eyes (distance and orientation), in addition to some constraints, are taken into account to filter out the outliers. The feature triplets (eyes/lips) are established, and eye and the mouth triplets, out of all objects that are far from the center of the face rectangle, are removed. The same filter, but with a different scale, is locally applied to the face image for precise detections.

The purpose of Frangi filter is to enhance vessel structures to be used for segmentation. This preprocessing enhancement is to maximize the intensity projection display, improve small vessel delineation and reduce organ over-projection. The multi-scale approaches normally assess the local orientation via eigenvalue analysis of the Hessian matrix [101] to locally determine the

likelihood that a vessel is present. The Hessian matrix is a square matrix of second-order partial derivatives of a function and it describes the local curvature of a function of many variables. Frangi considered vessel enhancement as a filtering process that searches for geometrical tubular structures. To consider different sizes of the vessels, a measurement scale is introduced which varies within a certain range.

The filter is based on an anisotropic diffusion (reducing image noise without removing significant parts of the image content) scheme guided by Vessel-likelihood at pixel level. Different smoothing filters, which strength and direction of diffusion is determined by a vesselness measure have been reported in literature. Frangi [63] introduced vesselness by calculating and analyzing the eigenvalues (λ) of the Hessian matrix. The eigenvalues of the Hessian matrix of a 2D image are sorted by increasing magnitude. The eigenvalues collected for the vertical and horizontal directions and then processed later.

$$|\lambda_1| \leq |\lambda_2| \quad \mathbf{3-1}$$

Frangi's vesselness function is composed of two components, formulated to discriminate tubular structures from other ones, as shown in Equation (3-1).

$$U(s) = \begin{cases} 0 & \text{if } \lambda_2 > 0 \\ \exp\left(-\frac{\mathcal{R}_\beta^2}{2\beta^2}\right) \left(1 - \exp\left(-\frac{S^2}{2c^2}\right)\right) & \end{cases} \quad \mathbf{3-2}$$

where $\mathcal{R}_\beta = \frac{\lambda_1}{\lambda_2}$ is the ‘‘Blobness’’ in the image, $S = \sqrt{\lambda_1^2 + \lambda_2^2}$ is called ‘‘second order structureness’’, and β and c are the thresholds which control the sensitivity of the vesselness measure. The vesselness measure in (3-2) is calculated at different scales (S), and the maximum response is selected. The features in \mathcal{R}_β and S are mapped into the probability-like estimates of

vesselness, according to different criteria. It was shown in [63] that the best results are achieved by setting β to 0.5, while the value of c depends on the grey-scale range of the image. Figure 3-12 shows example of the filter in the original paper.



Figure 3-12. Left to right: X-ray image of the peripheral vasculature, calculated vesselness of the image, calculated vesselness after inversion of the grey-scale map, image obtained by subtracting reference image from left image for visualization [63].

The performance of this filter highly relies on the choice of scale. Although the multiscale vesselness filter can automatically determine the scale of the detected structure (by maximum response across all possible scales), finding the “possible scales” is a big challenge. There are few papers on how to find these scales [102]. Equation (3-2) should be calculated for each different possible scale, and the maximum of the filter in each pixel is assigned to the final filter’s response. Thus, not only the computational cost is high, but finding automatically the appropriate scales for each image is a problem.

3.3.2 Filter modification

To overcome the problems mentioned in previous section (finding possible scales and computational costs), the following mathematical extension is proposed. The first modification is

limiting the filter to only one scale factor. Although at first glance it seems non-plausible to work with a single scale instead of a multiscale, the suggested method performed well on different image databases. This approach helps to reduce the response time by the factor, equal to the number of scales used in the original filter. In our algorithm, the filter is applied twice, once in coarse estimation of the eye location, and then in the precise detection of features. The scale factor for coarse estimation is set to the average value of histogram bins of image multiplied by 0.2. The scale factor for fine localization is set to 1/5 of the first scale. Assume the output of the original filter is J at one scale, the new filter is defined by the following equation:

$$Fr = S^{normal(Re(-\log(\lambda_1)^J))} \quad \mathbf{3-3}$$

Where Re is the real part of the filter and normal function normalizes the outputs to the interval $[0 \ 1]$. The Fr image is the new output of the filter which is sensitive to facial features, especially to the eyes. Figure 3-13 shows the output of Frangi filter applied to face images.

3.3.3 Eye detection

To localize the eyes, the following two steps are performed. First, by using the scale factor mentioned in previous section, the rough estimation of eye position is obtained. In the filtered image (called $F1$ -see Figure 3-13), two eyes, mouth and nose in addition to the few other objects, are detected.

Since the eyes are located in the upper-half of the face, the other features are assumed to be in the regions showed in Figure 3-14. The shape of the detected eyes is often ellipsoid and can be easily filtered out based on their shape and position. So, the output of the filter is binarized and then the eye location is easily extracted, based on the mentioned information. In order to obtain the rough estimation of the lip region, the triple of the eyes-mouth is established.

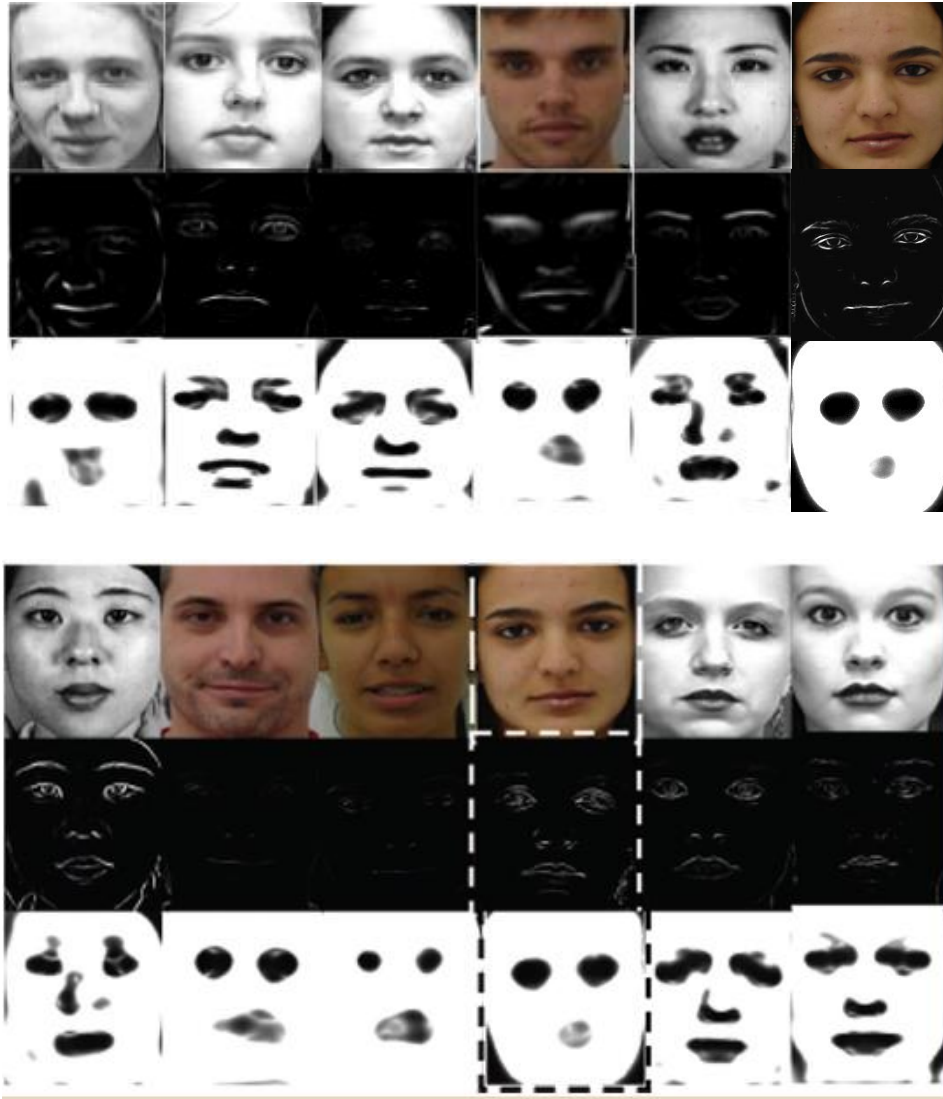


Figure 3-13. Top to bottom: Input face images, the output of the Frangi filter, the output of the proposed filter (F1).

Consider the face mask shown in Figure 3-14, which was created based on the common proportions of human body and adopted from [103]. The distance between two eyes is $P/2$, and the distance between the mouth and the line connecting the eyes is $Q/3$. These assumptions help to remove outliers. These distances are just used as an estimated location of facial features. In this approach we just talk about the front face images which can have head rotation but not

much. For the rotated images, the same approach can be used but after eye detection, the image is aligned. These types of images are not discussed in this thesis. The exact location of the eyes is obtained by using the same filter with a different pre-determined scale factor. The estimated location of the eyes is searched in a new filtered image (called $F2$ -see Figure 3-16). The position of pupil is found precisely by searching for low intensity pixels. The average coordinates of these pixels determine the exact location of the eyes, and the smallest convex polygon that contains these pixels is kept. Figure 3-15 shows few examples of eye detection.

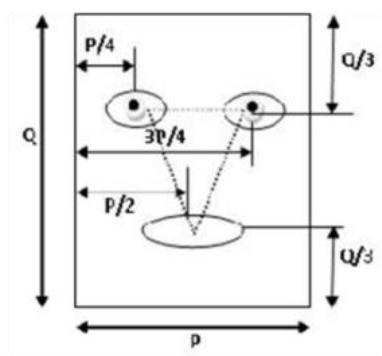


Figure 3-14. Face mask used for localization of other features based on eye position.



Figure 3-15. Fine estimation of eyes position based on binary mask obtained from $F1$ image.

3.3.4 Detection of other facial features

The same approach is used to detect the remaining facial features. Following the detection of the eyes, the locations of other features are easy to find. To detect eyebrows, the following facts are used to limit the search area:

- Eyebrows are always at the top of eyes and exact position of eyes is already known.
- The eyebrows are darker or lighter than skin.

In image $F2$, eyebrows are always darker than the background and can be easily detected by searching a certain distance from top of eye position, and converting this area into a binary image. In Figure 3-16, the procedure is demonstrated. For lips detection, based on information provided in image $F1$, the procedure similar to eyebrow detection is applied.

In image $F2$, the lips are darker than the background, and by locally converting the lips area to a binary image, the lips are detected. The position of the nose is the easiest one, because, unlike the eyes, brows and mouth, the nose cannot undergo large deformations. The only problem is that the nose is highly affected by shadows or unclear texture. To extract the nose contour, we assume that the nose is within the two “x” and “y” coordinates of the eyes, and on top of the mouth. This region is searched for dark pixels versus background in image $F2$. The output of the proposed filter always generates darker pixels for nostril and surrounding pixels, thus, extracting the nose is a simple task. Figure 3-16 shows an example of nose detection.

Using $F1$ and $F2$, which contain coarse and fine spatial information of facial features, the eye positions are used as a reference for extraction of other facial features by using simple constraints. There is no need to do any processing except the manipulation of geometric constraints. The precise location of facial features to support registration can also be found based on the detected areas. For instance, in several papers, the four corners of the lips are detected as

fiducial points. Here, based on a rectangle around the lips, the four corners can be found rather easily. This is also applicable to the nose and eyebrows, if their corners or the point between the corners are of interest.



Figure 3-16. Other facial feature detection based on geometric information of eyes (F2). Left to right: coarse and fine estimation of eyebrows, lips detection, nostrils detection.

3.4 Experiments and results

The face detection is performed using the Viola-Jones face detection technique which has been previously explained. The performance of the proposed filter has been tested on four different databases: JAFFE and Extended Cohn-Kanade (CK+) as the controlled face databases, and FEI and ORL as the uncontrolled ones. The Japanese Female Facial Expression (JAFFE) database [104] contains 213 images with the resolution of 256×256 pixels. Various facial expressions are available in this database. The images were taken from 10 Japanese female models, and each emotion was subjectively tested on 60 Japanese volunteers. The extended Cohn-Kanade database [51] contains 593 sequences from 123 subjects. The image sequences vary in duration, and incorporate the onset to peak formation of the facial expressions. In total, there are 6917 images in the database with the resolution of 640×490 . The FEI face database [105] contains a set of face images taken at the Artificial Intelligence Laboratory of FEI in Brazil. There are 14 images

for each of the 200 individuals, a total of 2800 images. The face pose of each subject varies from left to right. For each subject 4 images have been selected that have the least optimal front view of the face. In total, 800 images are selected. The resolution of face images is 360×260 . The ORL database [106] contains 400 gray-scale images of 40 persons (each 10 images). For some subjects, the images have been taken by varying the lighting, facial expressions and facial details (glasses / no glasses). 4000 images in total have been selected from all subjects in the CK+ database randomly. All images from the other three databases (213, 800, and 400) are used in the experiment as well. The total number of images from all databases was 5413. Each face is enclosed in a 160×120 window for JAFFE database, and 256×256 for CK+ and FEI, and 92×112 for ORL. Once face detection is done, images are checked manually. If the face detection task was not performed well, the procedure was repeated manually. To test the performance of the system, the obtained coordinates of all facial features were compared against manual landmarks.

In our experiments, a small rectangle is used to label the detected features. The distance between the center of the left and right rectangles for each feature was compared with a ground truth value. For the eyes and eyebrows, it is the left and right ones, for the nose it is two nostrils, and for the mouth it is the left and right corner of the lips.

To evaluate the performance of the proposed method, the relative error defined in [107] was utilized. This error depends on the distances between the expected and the estimated feature positions. Let d_1 be the Euclidean distance between the left detected features, d_2 be the same for the right detected feature, and d_3 be the distance between the left and right features for ground truth. The relative error (RE) is defined as:

$$error = \frac{\max(d_1, d_2)}{d_3} < T \quad 3-4$$

If the relative error is less than 0.25, the detection is considered to be correct [108]. Table 3-1 shows the average detection rate for all features, in addition to the detection rate for relative error ($T = 0.25$). It is difficult to determine from the references, if a particular method works better, because of the differences in training data or test procedure. In Table 3-2, the approaches, which used the same databases, have been reported for fair comparison. Most of them just reported the accuracy of eye detection. The numbers are the average detection rate for the eyes. The detection rate of the proposed method is in the range of the reported ones.

Table 3-1. Average and separate detection rate

Database	Avg. detection	Eyes	Eyebrows	Mouth	Nose
CK+	98.20	98.87	97.62	98.2	98.08
JAFPE	99.53	100	98.12	100	100
FEI	96.68	96.62	94.25	98.5	97.37
ORL	96.73	97.25	95	98.44	96.25

Table 3-2. Comparison with some of the existing methods

Face database	Eye detection method	Eye detection %
JAFPE	Kim et al. [109]	100
JAFPE	Ma et al. [108]	98.6
CK+	Leite et al. [110]	95.58
CK+	Rowley-Baluja-Kanade [96]	96.32
ORL	Li et al. [111]	93.8
ORL	Kim and Kim [112]	91.9
Proposed	JAFPE, CK+, FEI, ORL	100, 98.08, 97.37, 96.25

The recognition rate for FEI and ORL image databases are lower than the others, since these are uncontrolled databases, but the detection rate is comparable with other approaches. Some examples are shown in Figure 3-17.



Figure 3-17. Examples of automatic detection of facial features.

3.5 Conclusions

In this chapter, we described a new filter based on Frangi's vesselness filter modified for facial feature extraction. The extracted facial features can be used later for facial recognition or expression recognition. The eyebrows, eyes, nostrils, and mouth were considered as facial features in this thesis. The eyes were segmented in the coarse estimation step, and the estimated position of the mouth was also obtained. The fine detection was performed with the same filter, but using a different scale. The mouth, nose and eyebrows were precisely localized by using

geometric information of facial features. The system has been tested on four different databases with different expressions and poses. The average detection rate of 97.92% for eye detection (100% for JAFFE, 98.08 for CK+, 97.37 for FEI, and 96.25 for ORL databases), and 97.78% for all other detected features have been achieved.

Chapter Four: **FEATURE EXTRACTION FOR FACIAL EXPRESSION RECOGNITION**

Research on automatic facial expression recognition is divided into two categories: appearance-based and geometry-based methods [113]. The appearance-based techniques utilize color information of each pixels in the image in order to recognize the facial expression, whereas the geometry-based approaches consider the geometric relationship between facial points for recognition. The literature review here is borrowed from [113] which provided comprehensive study of the available techniques in each class.

4.1 Appearance-based modelling

The primary approach to automatic FER is the appearance-based approach. It involves methods that classify facial expressions based on the color of the face pixels. Appearance-based algorithms include optic flow, dimensionality reduction techniques, image filters, and wavelets. A brief description of each technique is described in below.

4.1.1 Optical flow

Optical flow analysis is one of the early techniques which has been used as an appearance-based method for FER. It analyzes the changes in pixel intensity (x,y) in an image from several consecutive frames in order to track object movements. The magnitude of the image velocities in X and Y directions (v_x, v_y) , are associated with each pixel. The feature vector is the velocity vectors over the studied frames and is used for FER. Since different expressions result in displacement of facial points, using optical flow, the image velocities of these facial points for different expression would be different.

Mase [114] was among the first ones who used optical flow for the first time for FER and he proposed two different approaches called “top-down” and “bottom-up”. The former uses optical flow to study the deformation of muscles that are engaged in different expressions while in the

latter the facial expression is classified directly from the optical flow's output over several small patches in a face image. The mean and variance of the optical flow in each patch along X and Y direction are used as feature and a KNN classifier performs emotion labeling. The recognition rate of 80% has been reported in [114].

Yacoob and Davis [115] tracked several rectangles around the facial points and used their motion as features. The rectangles are drawn around the mouth and eyebrows and optic flow fields are calculated for eight directions. A rule-based classifier has been applied for recognition and an accuracy of 86% has been obtained. Black and Yacoob [116] applied optical flow analysis and defined parametric motion models for the eyes and mouth. The recognition rate of 86% over 40 subjects was reported.

4.1.2 Pixel intensity values

This simple approach uses the color of face pixels features in the appearance-based FER category. Many of existing approaches so far do not use color information of the face images but just the intensities. Around important facial points like the mouth, eyes, and eyebrows, a set of pixel values are extracted in multiple frames. Although this technique does not provide high recognition rate due to its “raw information” without any pre or post-processing, but it is quite simple and fast. For example, Littlewort, et al. [117] used multiple SVMs for the classification of different expressions by using pixel intensity values. The accuracy is only around 73%.

4.1.3 Dimensionality reduction

Similar to many other approaches, the pixel value is the key point for feature extraction. These values can be used directly or be processed to be further used as features. However, the main problem is the huge number of pixels in each captured image regardless of its resolution. In addition, there are many pixels in face image which do not contain valuable information for

FER. If the pixel's intensity (value) changes significantly from one expression to another, that pixel is a good candidate to be selected as a feature. There are also many pixels which are correlated to each other in facial muscles deformation. Thus, removing these types of redundancy is important. The well-known techniques such as PCA and ICA are normally used along with the pixel-analysis approaches. Bartlett, et al. [118] used PCA to classify upper and lower facial AUs, six each. The first 30 principle components are used as features and 79.3% average accuracy has been achieved. They outperformed their approach reported in [119] by using 50 principle components and a neural network classifier to obtain 88.6% recognition rate. Donato, et al. [44] used ICA instead of PCA with the same AUs as used in [119] and reported 96% accuracy.

4.1.4 Gabor filters

In the appearance-based category, using different textural filters for feature extraction demonstrated high performance and reliability [113]. Among the different textural filters, Gabor filters [120] showed very good results on facial biometrics. The problem with dimension reduction techniques like ICA is the efforts needed for training and calculation of the independent components [113]. The filtering technique can intensify the deformations on the face texture. Gabor functions are Gaussians filters modulated by sinusoidal functions and it has been proved that Gabor filtering is an effective scheme for image representation [121]. A two-dimensional (2D) Gabor filter can be represented by the following equation in the spatial domain:

$$G(x, y; \theta, f) = \exp \left\{ -\frac{1}{2} \left(\frac{\hat{x}^2}{\delta_x^2} + \frac{\hat{y}^2}{\delta_y^2} \right) \right\} \cos(2\pi f \hat{x})$$

$$\begin{aligned} \hat{x} &= x \cos \theta + y \sin \theta \\ \hat{y} &= y \cos \theta - x \sin \theta \end{aligned}$$
4-1

where f is the frequency of the sinusoidal plane wave along the direction θ from the x-axis, $\delta_{\hat{x}}$ and $\delta_{\hat{y}}$ are the space constants of the Gaussian envelope along \hat{x} and \hat{y} axes, respectively.

The frequency parameter f is often chosen to be a number of power 2.

An example of a Gabor filter is given in Figure 4-1, which shows the absolute value (left), real component (middle), and imaginary component (right) of the filter in the space domain. The real and imaginary components accentuate, respectively, the symmetric and asymmetric responses of the image to the filter's characteristic frequency and orientation. The filter can then be applied to an input image $I \in \mathbb{R}^2$ using a two-dimensional convolution. A filter bank of several Gabor filters with different scales and orientations, is used for feature extraction for FER. The combined response is called a jet. The number of filter differs in every paper but typically 6 to 8 different orientations, and 4 to 6 different frequencies (scales) are used [113]. Since the output of Gabor filter is complex-valued, the magnitude of filtered image is used as features. The Gabor filters can be applied locally or globally onto face image. Few of Gabor-based FER with good results can be found in [122] [123]. The drawbacks of Gabor filtering approach although they have high recognition rate, are the extensive computational costs and amount of memory needed for calculation of different filters. For example, having 8 orientations and 5 frequencies in filter banks needs $8 \times 5 = 40$ times more memory and takes 40 times longer than using only one filter. The convolution of a Gabor filter with an input image requires a transformation from time to frequency domain in order to perform FFT. The image is multiplied by the filter in frequency domain and finally the inverse FFT should be calculated. This procedure should be repeated 40 times.

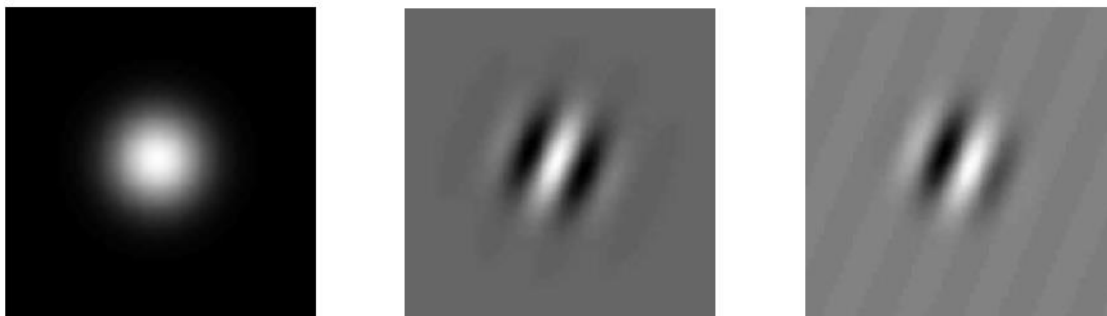


Figure 4-1. Gabor filter response: left to right are the absolute, real and imaginary parts of a Gabor filter [113].

4.1.5 Wavelets

One of the effective solutions to reduce the cost is using wavelets and especially Haar-wavelet. The binary nature of the Haar wavelet reduces the computational cost of the filtering procedure. Instead of multiplying the wavelet (filter) by an image, the Haar coefficient is computed as the difference in average pixel values between the black and white regions. Figure 4-2 shows four types of rectangular Haar wavelet. Thus, there is no need to perform FFT and using the “integral image” technique demonstrated by Viola and Jones [82], the system response is way faster than Gabor filter [113].

4.2 Geometry-based modeling

Using geometry of face in FER is quite common. The geometric positions of fiducial points (important facial points), along with their relative locations respect to each other, construct the geometric features. Different approaches in this category are available that use various facial points and relative distances to represent geometric features. Usually the fiducial points are located around the eyes, eyebrows nose and the mouth.

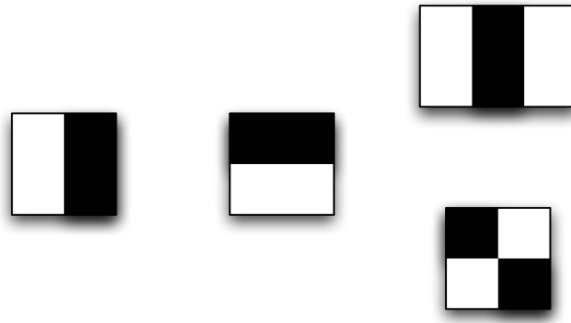


Figure 4-2. Four types of rectangular Haar wavelet-like features.

Different expressions result in movement of fiducial points which differ from expression to expression. This simple fact motivated many researches to come up with the best geometry mask (mesh) that can represent the total face deformation. The key point in this kind of techniques is to localize the fiducial points correctly as any error in the detected locations cause misclassification. In video processing, the tracking of these points is another problem. As discussed in previous chapter, there are various methods for facial features detection. Some of the famous techniques include optical flow, elastic graph matching, active appearance models, and deformable models [73].

The differences between published methods using geometric features mainly depend on:

- The number and location of selected points on the face
- Using 3-dimensional information of the face or not
- Constructing distances based on the detected points.

The main difference between the available techniques is how a set of selected facial points is converted into feature vector. As mentioned earlier, one solution is to use the relative positions of these points and construct several distances. In video processing, the displacements of the corresponding distance in different frames over time is also used. The normalization of these

distances over different faces is performed to enhance the accuracy of the system. Few of the reported techniques in [113] are reviewed here. Figure 4-3 shows an example of this approach

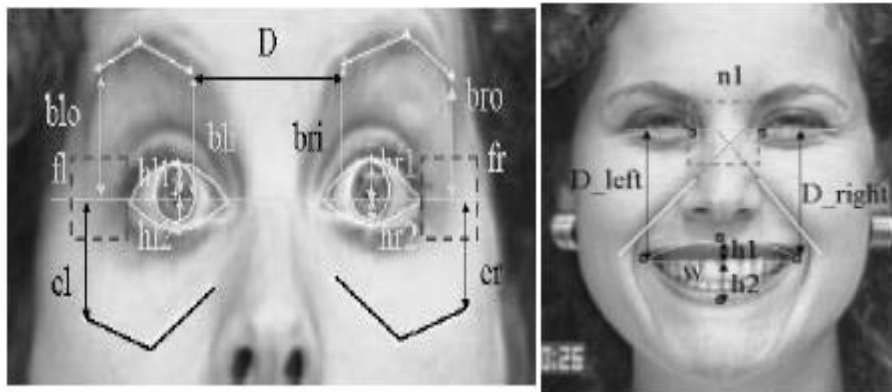


Figure 4-3. Example of geometry-based approach [62]: lower and upper face geometry features.

Other approach is to construct distances and curves out of fiducial points. For example in [124], the height and width of facial components like the eyebrows and mouth in addition to the shape of these components like the curvature of the lips are used as features. These feature then are tracked over time for FER. Figure 4-4 shows an example of this approach.

The other way to encode facial points is to construct a model which represents facial deformation. Cohen [125] used a 3D mesh model to track the displacement of facial muscles over time. For any expression, the classifier learns the deformation of the mesh and the relative distances in a video. Other approach is to use fiducial points' movement to analyze the corresponding face muscles where the muscles movement over time that constructs an expression is learnt. For example, Essa proposed a system that analyzes facial expressions by observing expressive articulations of a subject's face in video sequences using optical flow [43]. The motions are then coupled to a physical model that describes the skin and muscle structure.

Testing over many images, a minimal parametric representation of a face with different expressions is constructed and used for FER. Figure 4-6 shows an example of this approach.

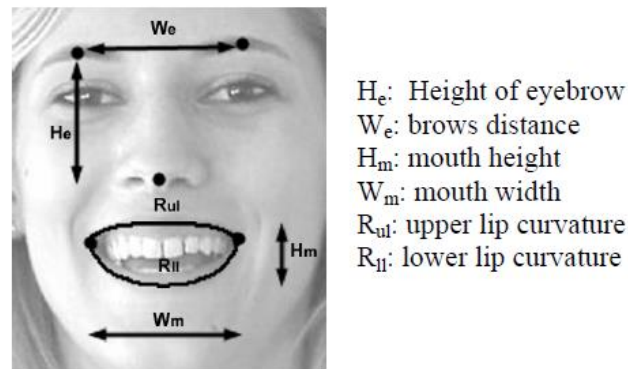


Figure 4-4. Example of using locations and relative Distances approach [124]: Geometrical parameters of the face, forming the feature vector.

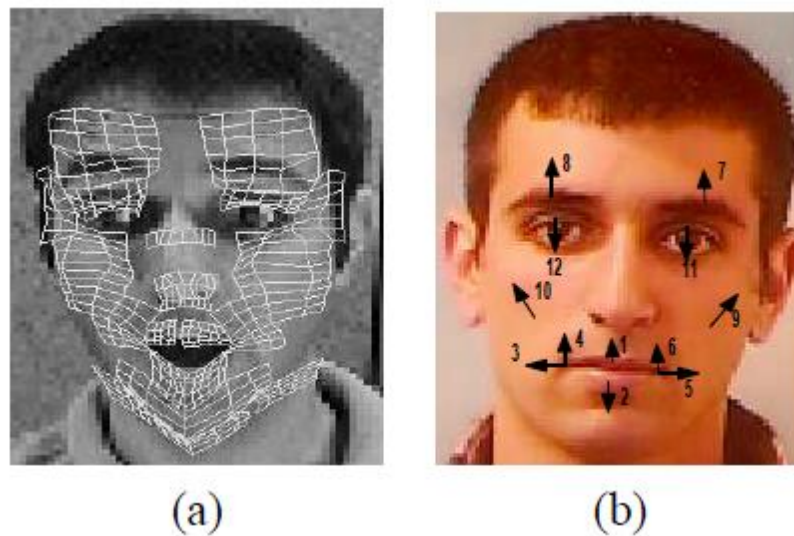


Figure 4-5. Example of parameter estimation approach [125]: The deformable model, (b) the facial motion measurements.

4.3 Gauss-Laguerre wavelets

The circular harmonic wavelets (CHWs) are polar-separable wavelets, with harmonic angular shape. They are steerable in any desired direction by simple multiplication with a complex

steering factor, thus they are referred to as self-steerable wavelets. The CHWs were first introduced in [126] and utilize the concepts from circular harmonic functions (CHF) employed in optical correlations for rotation invariant pattern recognition. A CHF is represented in polar coordinates [127] as

$$f_k^n(r, \theta) = V_k^n(r) e^{in\theta} \quad \mathbf{4-2}$$

where n is the order, k is the degree of the CHF, and $V_k^n()$ is the radial profile. The same functions also appear in harmonic tomographic decomposition, and have been considered for the analysis of local image symmetry. CHFs have been employed for defining rotation-invariant pattern signatures [128]. A family of orthogonal CHWs, forming a multi-resolution pyramid, referred to as the circular harmonic pyramid (CHP), is utilized for coefficient generation and coding. Each CHW, pertaining to the pyramid, represents the image by translated, dilated and rotated versions of a CHF. At the same time, for a fixed resolution, the CHP orthogonal system provides a local representation of the given image around a point in terms of CHFs. The self-steerability of each component of the CHP can be exploited for pattern analysis in the presence of rotation (other than translation and dilation), in particular, for pattern recognition, irrespective of orientation.

CHFs are complex, polar separable filters, characterized by harmonic angular shape, which allows building rotationally invariant descriptors. A scale parameter is also introduced to perform a multiresolution analysis. The Gauss-Laguerre filters from the family of orthogonal functions, satisfying the wavelet admissibility condition required for multiresolution wavelet pyramid analysis.

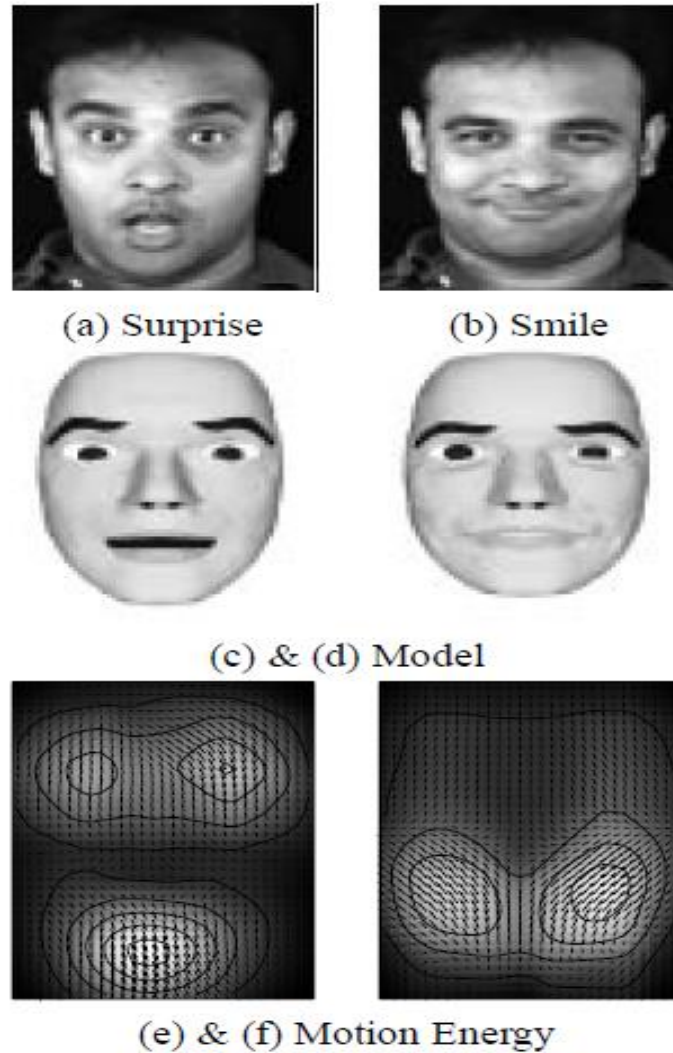


Figure 4-6. Example of using Models of face musculature approach [43]: Determining of expressions from video sequences. (a) and (b) show expressions of smile and surprise, (c) and (d) show a 3-D model with surprise and smile expressions, and (e) and (f) show the spatio-temporal motion energy representation of facial motion for these expressions.

Similar to Gabor wavelets, any image may be represented by translated, dilated and rotated replicas of the Gauss-Laguerre wavelet. For a fixed resolution, the Gauss-Laguerre CHF's provide a local representation of the image in the polar coordinate system centered at a given point, named the pivot point. This representation is called the Gauss-Laguerre transform [129]. It

is characterized by a circular harmonic function, which is a complex polar separable filter with a harmonic angular shape, represented in polar coordinates. For a given image $I(x, y) \in L^2$, the expression $I_p(r, \theta) = I(\tilde{x} + r \sin \theta, \tilde{y} + r \cos \theta)$ is the representation in the polar coordinate space centered at the pivot (\tilde{x}, \tilde{y}) [104]. $I_p(\cdot)$ can be decomposed in terms of CHF, based on its periodic characteristic with respect to θ :

$$I_p(r, \theta) = \sum_n \sum_k V_k^n(r) e^{in\theta} \quad 4-3$$

where radial profile $V_k^n(\rho)$ is given by the Fourier integral:

$$V_k^n(r) = \frac{1}{2\pi} \int_0^{2\pi} I_p(r, \theta) e^{-jn\theta} d\theta \quad 4-4$$

The expansion of radial profiles is represented by the series of weighted orthogonal functions, which are Gauss Laguerre CHF:

$$\mathfrak{S}_k^n(r, \theta) = (-1)^K 2^{\frac{(|n|+1)}{2}} \pi^{\frac{|n|}{2}} \left[\frac{K!}{(|n|+K)!} \right]^{\frac{1}{2}} r^{|n|} L_K^n(2\pi r^2) e^{-\pi r^2} e^{jn\theta} \quad 4-5$$

where $L_K^n(r)$ is the generalized Laguerre polynomial defined by

$$L_K^n(r) = \sum_{h=0}^K (-1)^K \binom{n+K}{K-h} \frac{r^h}{h!} \quad 4-6$$

As any CHF, GL functions are self-steering, i.e. they are rotated by angle φ when multiplied by factor $e^{jn\varphi}$. In particular, the real and imaginary parts of each GL function form a geometrical pair in phase quadrature. Moreover, GL functions are isomorphic to their Fourier transform. It is

shown in [129] that each GL function defines an admissible dyadic wavelet. Thus, the redundant set of wavelets; corresponding to different GL functions, represent a self-steering pyramid, utilized for local and multiscale image analysis. The real part of the GL function is depicted in Figure 4-7(a).

An important feature, applicable to facial expression recognition, is that GL function with various degrees of freedom can be tuned to significant visual features. For example, for $n = 1$, GLs are tuned to edges, for $n = 2$ to ridges, for $n = 3$ to equiangular forks, for $n = 4$ to orthogonal crosses, irrespective of their actual orientation [130]. Given an image $I(x, y)$, for every site of the plane, it is possible to perform the GL analysis by convolving it with each properly scaled GL function as follows:

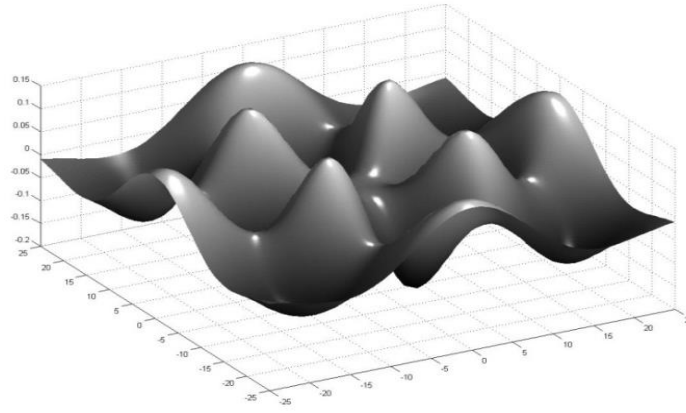
$$g_{jnK}(x, y) = \frac{1}{a2^j} L_K^n\left(\frac{rcos\theta}{a2^j}, \frac{rsin\theta}{a2^j}\right) \quad 4-7$$

where $a2^j$ are the dyadic scale factors. Figure 4-7(b) shows a plot of GL for a fixed dyadic scale factor and variable n and k .

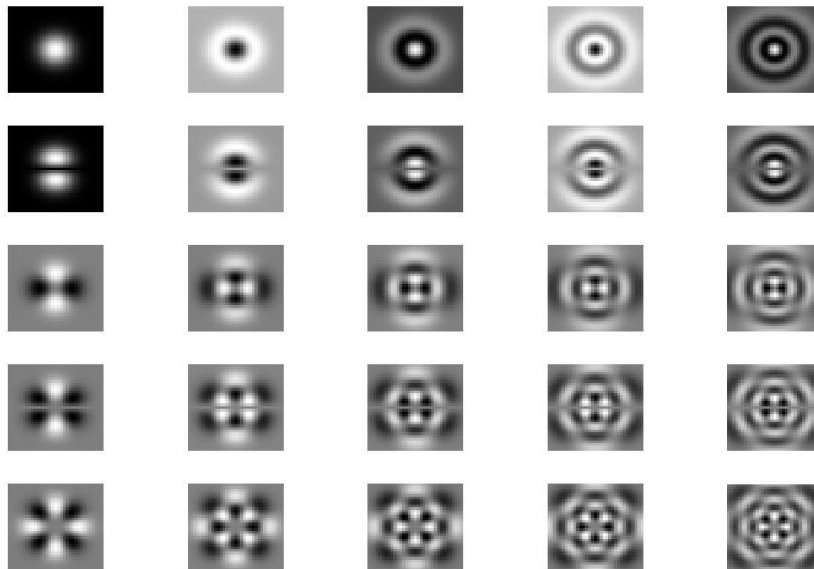
4.4 Implementation and comparison

In this thesis several experiments have been conducted. In early stage of this research [8] only GL textural features have been used by applying on whole face image. In this case, the facial points detection has been removed from the process, and the global facial features have been used. Later on, in [11], we proposed that the features have been selected based on the combination of textural and geometric features as explained above, while the facial points have been extracted using Active Appearance Model (AAM). The details of the AAM were given in

Chapter 3. At that stage of the research, the proposed facial feature detection based on modified Frangi filter was evaluated, and AAM was selected for facial point detection.



(a)



(b)

Figure 4-7. (a) Real part of GL function; $n = 4$, $K = 1$, $j = 2$. (b) Real part of GL CHF; with the variation of filter in spatial domain using the fixed scale, $k = 0, 1, \dots, 4$, and $n = 1, 2, \dots, 5$.

In [10], we proposed a new facial point detection technique, and the FER recognition system was completed, as introduced in this thesis. Each step of this progress with the experimental results will be described in this section, and the improvements made at each step will be discussed.

4.4.1 Global GL feature selection

In order to perform the feature selection, the GL filter is used, followed by PCA for dimension reduction. The parameters of the GL filter are selected by conducting several simulations with different filter parameters and observing the classification rate. The best parameters of the filter, that have been selected based on the trials, are $n = 2, k = 1, j = 1$.

The filter is tested on the JAFFE database and Figure 4-8 shows examples of each expression. Preprocessing is normally performed before feature extraction for FER. The aim of this step includes scaling, intensity normalization, and size equalization is to have images, which only contain a face expressing certain emotion. Sometimes, histogram equalization is also used to adjust image brightness and contrast. For face localization, we used the well-known algorithm by Viola-Jones. The cropped face image is then resized to 128×96 . By applying GL filter to the cropped image, the size of output image is the same size of the input image with complex-valued pixels. The real part of each pixel is used as feature and all pixels ($128 \times 96 = 12288$). As explained, the size of the textural feature vector is quite large for classification. If the dimension of the input vector is large, and the data is highly correlated, there are several methods to remove redundancy, such as principal component analysis (PCA). During PCA, the components of input vectors are orthogonalized, which means they are no more correlated with each other. The components are put in order, so that ones with largest variation come first and those with low variation are eliminated. The data is usually normalized before performing PCA to have zero mean and unity variance. In our case, the size of feature vector, after down sampling, is 3072 (by

factor 4), which is further reduced to 384 samples per image using the PCA. Figure 4-9 shows the output of GL filter with different parameters. In order to perform the feature selection, the GL filter is used, followed by PCA for dimension reduction. The parameters of the GL filter are selected by conducting several simulations with different filter parameters and observing the classification rate. The best parameters of the filter, that have been selected based on the trials, are $n = 2, k = 1, j = 1$.



Figure 4-8. Different expressions from JAFFE database (left to right): anger, disgust, fear, happiness, sadness, surprise, and neutral.

Finally, a two layer perceptron neural network is used as classifier. Extracted features by GL are fed in the input units of neural network. Seven outputs are considered for different expressions, and each of them gives the probability of the input image belonging to the associated facial expression [131]. Number of neurons in hidden layer is obtained through multiple tests, by adding neurons one-by-one and monitoring the recognition rate of the network. The best result was obtained with 12 neurons. Figure 4-10 shows the recognition rate versus the number of neurons in the hidden layer, and Figure 4-11 shows the evaluation of error during training.

In JAFFE database, each subject posed 3 or 4 examples for each expression. In training phase, according to [30] the following procedure is used:

1. The database is partitioned to 10 distinct segments randomly.

2. The neural network is trained by 9 partitions and the performance is tested by evaluating the error (Figure 4-10). The other partition is used for testing.
3. The procedure is repeated for all possible choices and the average recognition rate is reported.

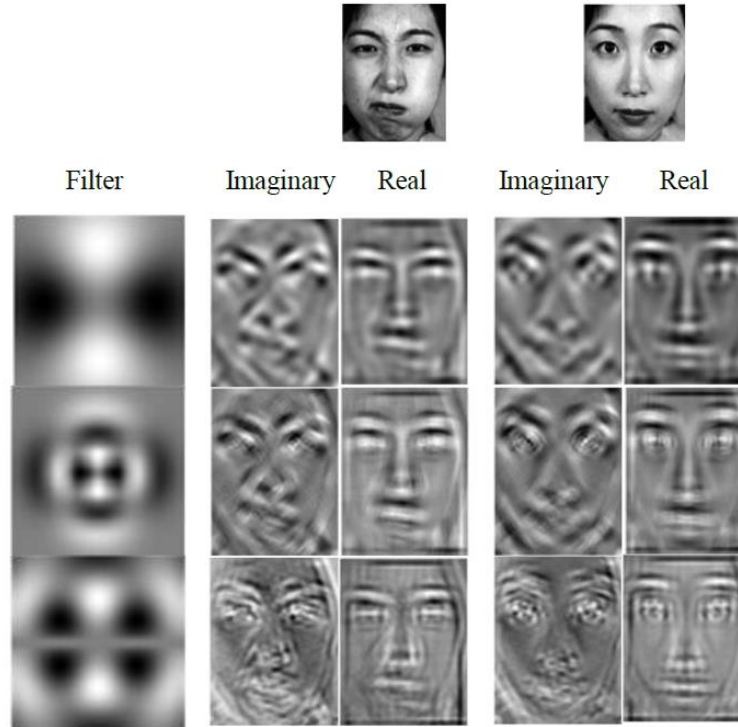


Figure 4-9. The GL wavelet response for three different filters on neutral and face with expression.

Since the training highly depends on initial guess of the weights of the perceptrons, the training is also repeated 5 times with different initial weights. In our experiment, $5 \times 10 = 50$ training sets have been formed, and the average of recognition has been reported. Table 4-1 shows the confusion matrix, which represents the percentage of matching the facial expression (matching rate), thus showing if the algorithm is confusing two classes. The data in the Table 4-1 is the rounded average rate in three trials. The total recognition rate is 94.7%; the best rate is for

surprise and neutral expressions, and the lowest ones are for fear and sadness. This approach only used the “global” textural feature and has been tested on a simple database. We did not use any geometric information of the face and textural information around important facial points like eyes, brows, nose, and mouth have not been used. The system fails on more complex database, and main reason was using a “weak” feature selection.

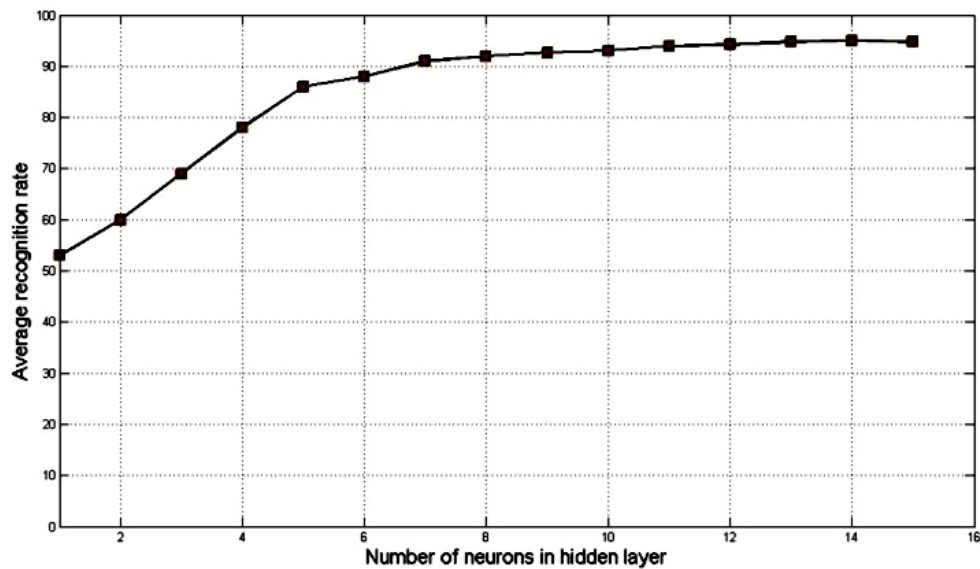


Figure 4-10. Recognition rate with respect to number of neurons in hidden layer.

Table 4-1. Confusion matrix for JAFFE database

	AN	FE	SU	DI	HA	SA	NE	Total
AN	94%	0	0	4%	0	2%	0	30
FE	0	91%	4%	3%	0	0	2%	32
SU	0	0	98%	0	2%	0	0	30
DI	3%	0	0	94%	0	3%	0	29
HA	0	0	0	0	97%	0	3%	31
SA	6%	0	0	3%	0	91%	0	31
NE	0	0	0	0	0	2%	98%	30

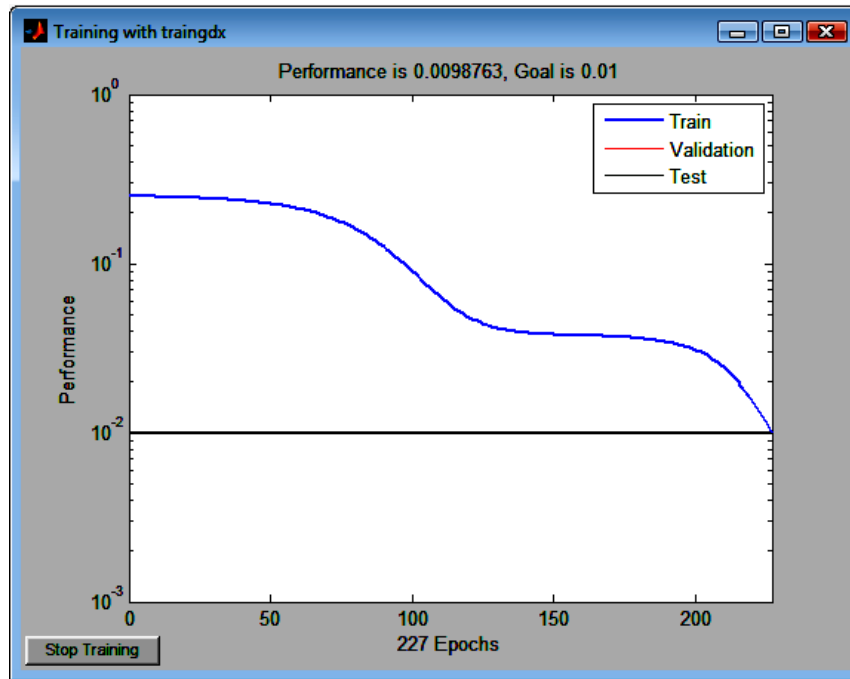


Figure 4-11. Evolution of error in neural network during training.

4.4.2 Global textural and geometric feature selection

The second phase of the research was to use global textural features in addition to geometric features. The face is preprocessed, similarly to the previous approach. The image with neutral expression is scaled, so that it has fixed distance between the eyes. No intensity normalization has been considered, since the GL filters can extract an abundance of features without any preprocessing.

The eye detection is implemented based on the AAM, which is widely used in face recognition and expression classification, due to its remarkable performance to extract face shape and texture information. The AAM contains both statistical model and texture information of the face, and performs matching via finding the model parameters, which minimizes the difference between the image and the synthesized model. We used 18 fiducial points to model the face, and

distinguish facial expressions. In our experiment, the AAM model has been created using 120 images from JAFFE image database. After creating the AAM, the position of eyes in each image is automatically extracted, and the line, which connects the inner corner of eyes, is used for normalization.

For face localization, we applied the Viola-Jones algorithm. The localized face image is cropped automatically and resized to 128×96. The next step is to normalize the geometry and Figure 4-12 shows an example of this procedure. This step is important since if the face is not detected well, the AAM cannot find fiducial points.

The feature vector consists of two different types of feature: textural features which are extracted globally by applying GL filter, and geometric information of the local fiducial points. The global textural features have been selected using the same procedure explained in previous section with the size of 384.

For geometric features, 18 fiducial points are put together to construct the model (Figure 4-13). These points are extracted automatically based on AAM model. The coordinates of these fiducial points are used to calculate 15 Euclidean distances. Different expressions result in different deformation of the corresponding facial components, especially near eyes and mouth. The selected geometric features extracting is performed as follows:

- AAM is applied to extract the 18 points. The distances are labeled by d's as shown in Figure 4-13.
- For upper face, 10 distances are calculated, according to Table 4-2.
- For lower face, 5 distances are calculated, according to Table 4-2.

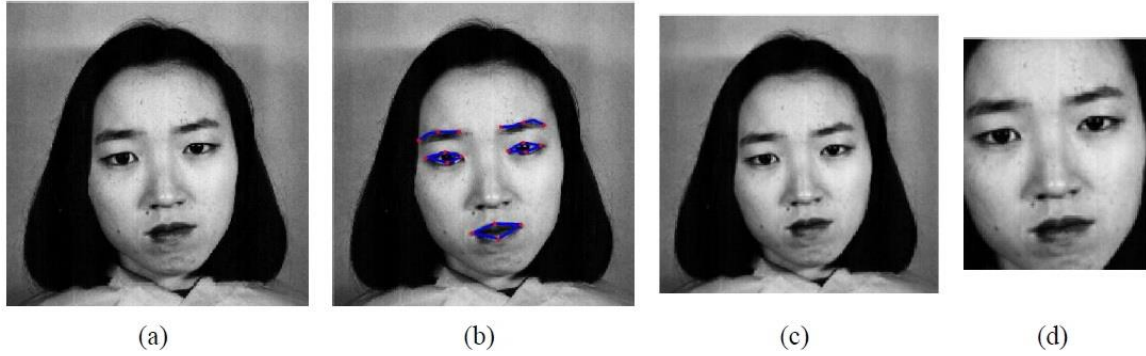


Figure 4-12. Normalization procedure (left to right): (a) input image (from the JAFFE database), (b) the extracted AAM fiducial points, (c) normalized image to have fixed distance between eyes (d) the localized and resized face.

The final feature vector is the combination of two types of features. Since the dimension of texture feature vector is 384, and the dimension of the geometric feature vector is 15, the total size is 399. During the simulations, it was observed that the geometric features are more important than texture ones. To find the appropriate weight coefficients for both types of features, the average recognition rate versus different weight coefficients for geometric feature have been monitored. The average recognition rate has been obtained via three trials and leave-one-out (LOO) approach. In each trial and for each database, randomly one image is used for testing and the rest utilized for training. In this case the three trials, generally speaking, have different sets for train/test schemes. The coefficient for geometric features varied from 0.5 by step-size 0.01. Figure 4-14 shows the average recognition rate versus geometric coefficient weight. The best average recognition rate within coefficients was 0.69 for geometric features and 0.31 for texture features.

We then used the KNN classifier, which is a well-known instance-based classification algorithm. It does not make any assumptions on the underlying data distribution. The similarity between the test sample and the other samples, used in training, is calculated, and k most similar set of

samples are determined. The class of the test sample is then found based on the classes of its K-Nearest Neighbors.

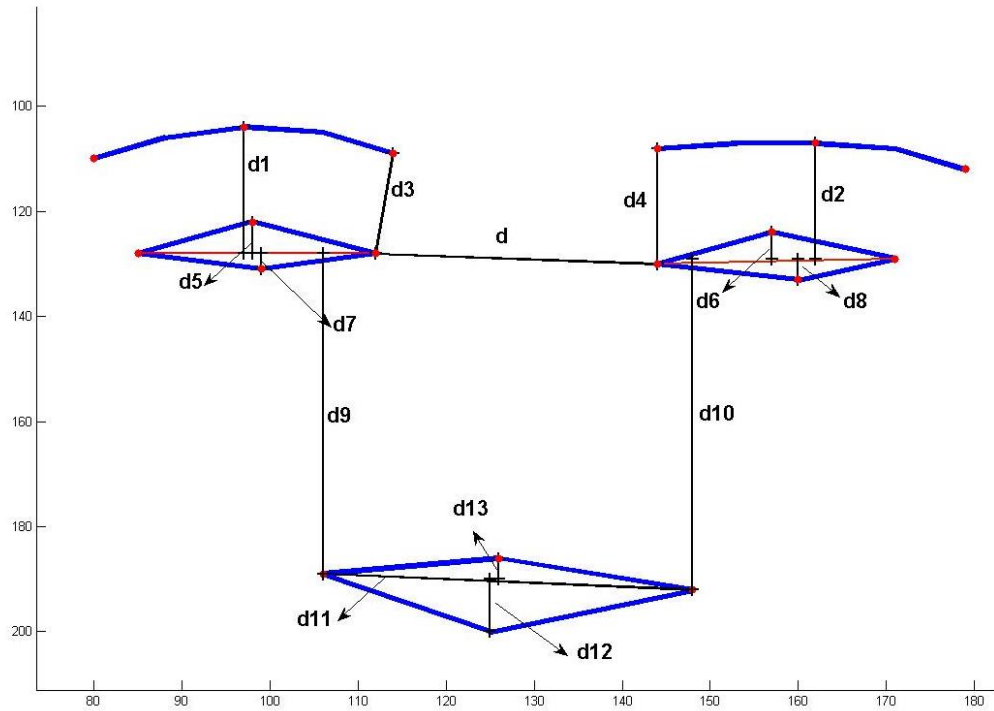


Figure 4-13. Geometric feature selection based on fiducial points.

This classification suits a multi-class classification, in which the decision is based on a small neighborhood of similar objects. In the classification procedure, the training data is first plotted in n dimensional space, where n is the number of features. Each of this data consists of a set of vectors labeled with their associated class (arbitrary number of classes). The number k defines how many neighbors influence the classification. Based on the suggestion made in [129], the better classification is obtained when $k = 3$. This suggestion was based on different experiments and observing classification rate on JAFFE database. The same classifier is used for Cohn-Kanade and MMI database as well.

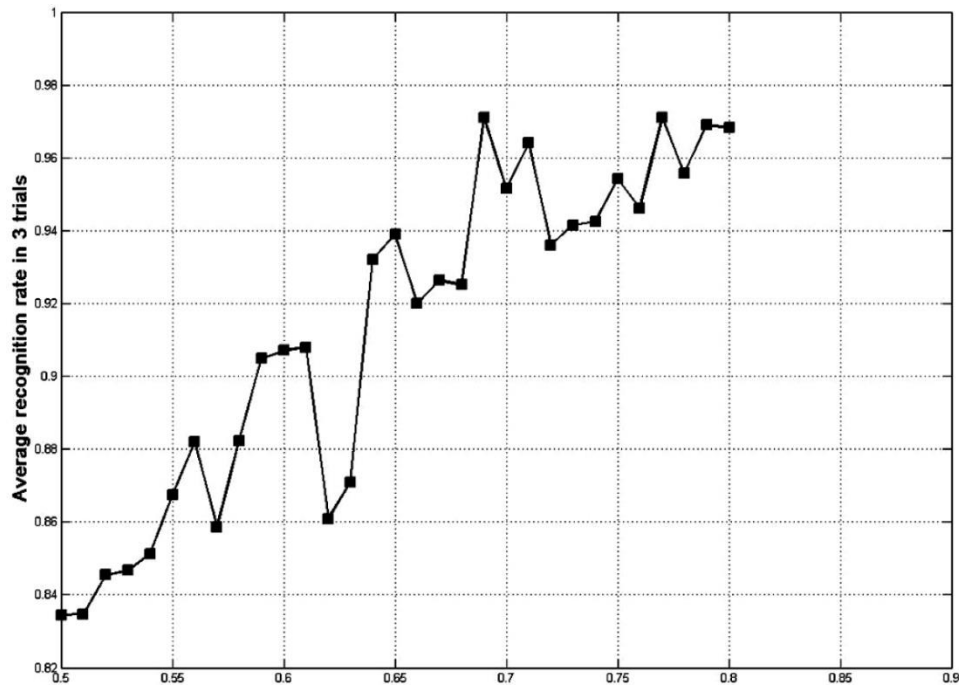


Figure 4-14. Variation of coefficient for geometric features versus average recognition rate.

The Cohn-Kanade database [130], which is widely used in literature for facial expression analysis, consists of approximately 500 image sequences from 100 subjects with the resolution of 490×640 pixels. The subjects range in age from 18 to 30 years where 65% are female, 15% are African-American, and 3% are Asian or Latino. The images contain six different facial expressions: anger, disgust, fear, happiness, sadness and anger. Each sequence contains 12-16 frames which expressing different stages of an expression development (from a low arousal stage, reaching a peak of arousal and then declining). Figure 4-15 shows examples of some expression.



Figure 4-15. Examples of various expressions from the Cohn-Kanade database.

The MMI Facial Expression Database [132] was created by the Man-machine Interaction Group, Delft University of Technology, Netherlands. This database is initially established for research on machine analysis of facial expressions. The database consists of over 2,900 videos and high-resolution still images of 75 subjects of both genders, who range in age from 19 to 62 years and have either a European, Asian, or South American ethnic background. These samples show both non-occluded and partially occluded faces, with or without facial hair and glasses. In our experiments, 96 image sequences were selected from the MMI database. The only selection criterion is that a sequence can be labeled as one of the six basic emotions [133]. The sequences come from 20 subjects, with 1–6 emotions per subject. The neutral face and three peak frames of each sequence (hence, 384 images in total) were used for 6-class expression recognition. Some sample images from the MMI database are shown in Figure 4-16.



Figure 4-16. The sample face expression images from the MMI database.

Three different methods were selected to verify the accuracy of this system:

1. Leave-one-out cross validation: For each expression from each subject, one image was left out, and the rest were used for training [19].
2. Cross validation: the database was randomly partitioned to 10 distinct segments, and 9 partitions were used for training and the remaining partition was used to test

performance. The procedure was repeated so that every equal-sized set was used once as the test set. Finally, the average of 10 experiments has been reported [134].

3. Expresser-based: the database was divided into several segments; each of them corresponded to a subject. For JAFFE database, 213 expression images posed by 10 subjects were partitioned into 10 segments, each corresponding to one subject [135]. For Cohn-Kanade database, 375 video sequences have been used which is over 4000 images. Nine out of ten segments were used for training and the tenth for testing. It was repeated, so each of ten segments was used in testing. The average results for those 10 experiments have been reported.

JAFFE database: Table 4-3 shows the average success rate for different approaches. Moreover, the confusion matrix for leave-one-out method is presented in Table 4-4. For the average recognition rate every time expression images of 9 out of 10 classes are used for training and the last one is the testing set. This procedure is repeated for each subject.

The confusion matrix is a 7×7 matrix containing the information of actual class label in both rows and columns. The diagonal entries are the rounded average successful recognition rates in 10 trials, while the off-diagonal entries correspond to misclassifications. The total recognition rate is 96.71%; the best rate is for surprise and happiness expressions, and the lowest one is for anger. The performance of proposed method has been compared against some published methods in Table 4-5.

Cohn-Kanade database: Table 4-6 shows the average success rate for different approaches. The confusion matrix for the leave-one-out method is also presented in Table 4-7. Different number of images have been selected for experiments in literature, and images are selected based on the different criteria (see Table 4-8).

Table 4-2. Upper (distance 1-10) and lower (distance 11-15) face geometric distance

Meaning	Distance	Meaning	Distance
left inner brow-left inner eye corner	$P_1 = \frac{d_3E - d_3N}{d_3N}$	right eye height	$P_6 = \frac{(d_6E + d_8E) - (d_6N + d_8N)}{(d_6N + d_8N)}$
right inner brow-right inner eye corner	$P_2 = \frac{d_4E - d_4N}{d_4N}$	left top eye point-line connecting left eye corners	$P_7 = \frac{d_5E - d_5N}{d_5N}$
left top brow-line connecting left eye corners	$P_3 = \frac{d_1E - d_1N}{d_1N}$	right top eye point-line connecting right eye corners	$P_8 = \frac{d_6E - d_6N}{d_6N}$
right top brow-line connecting right eye corners	$P_4 = \frac{d_2E - d_2N}{d_2N}$	left bottom eye point-line connecting left eye corners	$P_9 = \frac{d_7E - d_7N}{d_7N}$
left eye height	$P_5 = \frac{(d_5E + d_7E) - (d_5N + d_7N)}{(d_5N + d_7N)}$	right bottom eye point-line connecting right eye corners	$P_{10} = \frac{d_8E - d_8N}{d_8N}$
Mouth height	$P_{11} = \frac{(d_{12}E + d_{13}E) - (d_{12}N + d_{13}N)}{(d_{12}N + d_{13}N)}$	Mouth width	$P_{12} = \frac{d_{11}E - d_{11}N}{d_{11}N}$
left lip corner-line connecting left eye corners	$P_{13} = \frac{d_9E - d_9N}{d_9N}$	right lip corner-line connecting right eye corners	$P_{14} = \frac{d_{10}E - d_{10}N}{d_{10}N}$
top lip-line connecting lip corners	$P_{15} = \frac{d_{13}E - d_{13}N}{d_{13}N}$		

In this experiment, 375 image sequences have been selected from 97 subjects so that the criterion was to be that of a sequence labeled as one of the six basic emotions, with the video clip being longer than ten frames. The total recognition rate is 92.2%; the best rate is for the happiness expression, and the lowest one for sadness. The performance of the proposed method has been compared against some published methods in Table 4-8. Although several frames from each video sequence are used, we consider them as “static” images without using any temporal information.

MMI database: Table 4-9 shows the average success rate for different approaches. The total recognition rate is 87.6%; the best rate is for the happiness expression, and the lowest one being for sadness. The experimental results show that the proposed method meets the criteria of accuracy and efficiency for facial expression classification. It outperforms, in terms of accuracy, some other existing approaches that used the same database. The average recognition rate of the proposed approach is 96.71%, when using leave-one-out method, and 95.04% when using cross validation for estimating its accuracy on the JAFFE database. For the Cohn-Kanade database, the average recognition rate of the proposed approach is 92.20%, when using leave-one-out method, and 90.37% when using the cross validation for estimating its accuracy. For the MMI database, the average recognition rate of the proposed approach is 87.66%, when using the leave-one-out method, and 85.97% when using cross validation for estimating its accuracy. Few articles reported the accuracy on emotion recognition on the MMI. Most of them reported the recognition rate on the AU. Sánchez et al. [136] achieved 92.42% but it is not clear how many video sequences were used. Cerezo et al. [137] reported 92.9% average recognition rate on 1,500 still images of mixed MMI and CK databases. Shan et al. [133] used 384 images from the MMI, and the average recognition rate of 86.9% was reported.

Table 4-3. Recognition accuracy (%) on the JAFFE database for different approaches.

Expression	Leave-one-out	Cross validation	Expresser-based
Anger	94.3	92.8	89.3
Disgust	95.7	94.6	90.7
Fear	96.0	95.0	91.1
Happiness	98.2	96.9	92.6
Sad	96.7	94.2	90.2
Surprise	98.2	96.3	92.3
Neutral	97.9	95.5	91.7
Average	96.71	95.04	91.12

Table 4-4. Confusion matrix for the leave-one-out method (the JAFFE database)

	AN	FE	SU	DI	HA	SA	NE	Total
AN	94.3%	0	0.7	3.2%	0	1.8%	0	30
FE	1%	96%	2%	1%	0	0	0	32
SU	0	0.5%	98.2%	0	1.3%	0		30
DI	1.1	0	0.5%	95.7%	0	2.7%	0	29
HA	0	0	0	0	98.2%	0	1.8%	31
SA	2%	0	0.3%	1 %	0	96.7%	0	31
NE	0	0	0	0	0	2.1%	97.9%	30

For the leave-one-out procedure included in Table 4-6, all image sequences are divided into six classes, each corresponding to one of the six expressions. Four sets, each containing 20% of the data for each class, chosen randomly, were created to be used as training sets, while the other 20% were used as the test set. The procedure of classification is repeated five times. In each cycle, the samples in the testing set are included into the current training set. The new set of samples (20% of the samples for each class) is again formed to have a new test set, and the remaining ones are the new training set. Finally, the average classification rate is the mean of the success rate in classification.

Table 4-5. Comparison with other methods on the JAFFE database

Method	Recognition rate (Ave. %)
Lyons et al. [134]	92.00%
Zhi and Ruan [138]	95.91%
Zhang et al. [30]	90.34%
Liejun et al. [139]	95.7%
Shin et al. [35]	95.71%
Zhao et al. [140]	93.72%
Guo and Dyer [141]	91.00%
Proposed	96.71%

Table 4-6. Recognition accuracy (%) on the Cohn-Kanade database for different approaches

Expression	Leave-one-out	Cross validation	Expresser-based
Anger	88.25	87.03	82.35
Disgust	94.9	91.58	88.65
Fear	92.28	90.98	87.12
Happiness	97.82	96.92	91.05
Sad	87.87	84.58	81.11
Surprise	92.08	91.23	86.32
Average	92.2	90.37	86.1

Table 4-7. Confusion matrix for the leave-one-out method (the Cohn-Kanade database)

	AN	FE	SU	DI	HA	SA
AN	88.25%	1.21%	0.89%	1.08%	0.00	8.57%
FE	0.87%	92.28%	1.98%	2.17%	2.21%	0.49%
SU	1.87%	1.43%	92.08%	1.35%	1.31%	1.96%
DI	1.65%	0.67%	0.00	94.9%	0.00	2.78%
HA	0.00	0.00	2.18%	0.00	97.82%	0.00
SA	6.54%	4.63%	0.00	0.96%	0.00	87.87%

Table 4-8. Comparison of facial expression recognition for the Cohn-Kanade database

Method	# of selected video sequences	Recognition rate (Ave. %)
Zhan et al. [142]	300	90.4
Shan et al. [133]	320	92.1
Bartlett et al. [28]	313	86.9
Littlewort et al. [143]	313	93.8
Yang et al. [144]	352	92.23
Tian [123]	375	93.8
Aleksic and Katsaggelos [60]	284	93.66
Zafeiriou and Pitas [48]	374	97.1
Irene Kotsia and Ioannis Pitas [145]	-	99.7%
Proposed	374	92.20

Table 4-9. Recognition accuracy (%) on the MMI database for different approaches

Expression	Leave-one-out	Cross validation	Expresser-based
Anger	86.14	85.08	80.11
Disgust	85.22	86.4	78.22
Fear	89.91	84.42	81.32
Happiness	91.1	88.81	83.21
Sad	83.44	81.79	77.11
Surprise	90.2	89.34	81.02
Average	87.66	85.97	80.16

4.5 Conclusions

This chapter has investigated the proposed feature extraction for FER using GL wavelet. This filter is capable of providing efficient features that represent unique characteristics of facial images for expression recognition. Local and global feature extraction scheme have been

performed on different databases in this thesis to come up with the best solution in terms of accuracy. We also studied different geometric features in addition to the textural features to enhance the feature extraction technique. The experimental results led us to conclusion that the highest performance, in terms of recognition accuracy, was achieved by using Frangi filter for facial points detection, along with the global textural features in addition to 15 geometric distances. To show the efficiency of the proposed approach, three different scenarios (leave-one-out, cross validation, and expresser-based) have been applied on three different databases (JAFPE, CK, and MMI). The average recognition rate of 94.29% was achieved for JAFPE, 89.55% for CK, and 84.60% for MMI database. It should be noted that the lower recognition rate on MMI and CK compared to JAFPE does not mean the weak performance; this is rather due to the fact that those databases are more complex in terms of image quality, occlusion, and the number of image per subject showing the exact expression.

Chapter Five: **EXPRESSION-INVARIANT FACE RECOGNITION**

The aim of this chapter is to propose a solution for facial biometrics that can handle both facial expression and face recognition at the same time in a real-time application.

The goal of this chapter is to investigate how the feature extraction method proposed in this thesis for FER will perform in the task of face recognition. The proposed approach is as follows: first, Gauss-Laguerre filter is utilized for feature extraction; next, the same classifier that is used for FER (K-Nearest Neighbor), performs the classification of faces. In this chapter, we report on testing the proposed algorithm on various datasets of faces with expressions. The results showed that this approach is a robust form of feature extraction, which does not require intensive pre-processing, while maintaining relative recognition accuracy. The GL filters can be tuned specifically for each individual facial feature, or for the entire face, in order to extract useful information. Unlike traditional methods of filter-based feature extraction, that require a bank of filters to handle minor rotation and occlusion due to facial hair, an appropriately tuned GL filter is applied only once.

5.1 Introduction

Face recognition is one of the most actively researched areas in signal processing, because of the wide range of applications, including biometrics and security. For robust facial recognition, a large-scale database of biometric records and real-time processing requires efficient facial recognition algorithms, as the number of identification requests could be very high. On the other hand, as the size of the biometric database increases, the system tends to have an increase in the false acceptance rate. Most facial recognition algorithms are applied on extensively pre-processed images, allowing the feature extraction methods to operate under near-ideal conditions. An example of dealing with large databases and/or large volumes of people, seeking

for an access authorization, is the customer identification in an airport access point. The higher the number of people enrolled in the database, the higher the number of comparisons required to identify a person. The size of the feature vector has an impact on comparison time, which includes the computational time of the distance among all samples. Therefore, it is crucial for an identification system to have a low response time, while maintaining a reasonable accuracy. Moreover, facial expressions add more complexity and challenges for face recognition algorithms, although many of available face recognition techniques are tested mainly on expression-free (neutral expression) images.

There are some effective algorithms that can handle the facial expression in the task of face recognition [6] [146] [147] [148] [149], but there are only a few works that jointly address both FR and FER. These algorithms usually encode the expression and identity variability of faces in two different (independent) methods which are then used for recognition [4] [5]. Tenenbaum et al. [7] introduced tensor decomposition that offers a solution for modeling the multi-factor interactions (pose, expressions, etc. in face). There are few other papers which use face decomposition [150] [2] [3] [1] in order to separate the identity and expression components of a face. Taheri et al. [6] proposed facial component separation algorithm based on the principle of sparse representation. A face image with expression is modelled as a neutral face, superimposed by a sparse image of deformations, corresponding to the expression on the face. On the other hand, an expressive face is decomposed into a sum of neutral and expression elements, using dictionaries specifically suited to sparsify them. The data-driven dictionaries are generated in training phase. By having several samples of expressive faces per subject, the neutral and expressive parts are decomposed. Figure 5-1 shows an example of this approach where the face image with expression (X) is decomposed to the neutral face (X_n) and the pure expression (X_e).

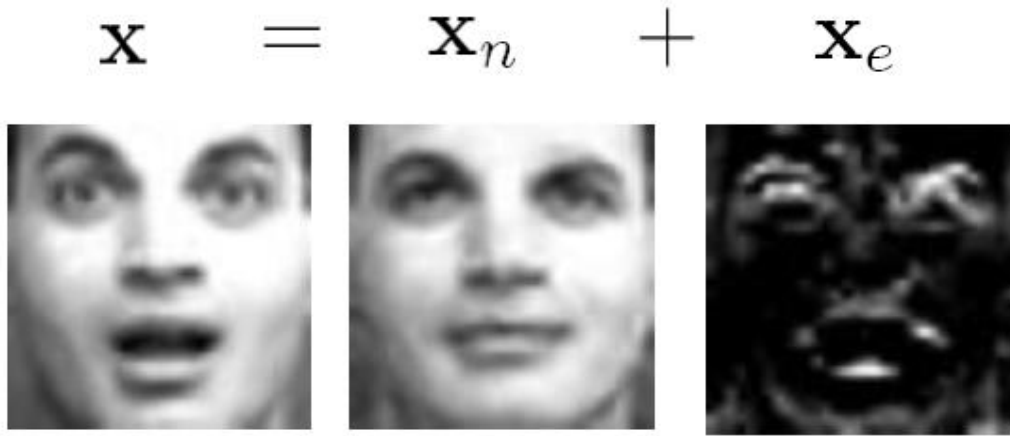


Figure 5-1. Facial component separation. (a) Original face image, (b) the superposition of a neutral component, (c) the expression component [6].

It is stated in [151] that facial expression usually affects the performance of face recognition algorithms. Some expression-invariant techniques use video frames to handle different expressions, since video contains more information than a single image [152] [153].

Other papers use still images instead of video. There are mainly two approaches in this class: subspace-based and optical flow-based. Tsai et al. [149] introduced subspace-based analysis for face recognition with expression. Subspace analysis methods are the processes of projecting high dimensional data to a lower dimension. Naseem et al. [147] formulated the pattern recognition problem in terms of linear regression: a linear model was developed, representing a probe image as a linear combination of class-specific galleries. Amberg et al. [154] suggested an expression-invariant algorithm for face recognition using an identity/expression separated 3D model fitted to shape data. Optical flow is mainly used to establish a correspondence between pairs of faces and compute the face warping transformation [155]. A weighting measure rewards the facial regions that are more similar between training and testing images, and penalizes the less importance areas. Jorstad et al. [146] proposed an insensitive lighting deformation metric to compare images

with different expressions. Nagesh et al. [148] introduced an expression-invariant face recognition method using compressive sensing: each training face image is represented by a feature that captures the common features of the faces in addition to another feature that captures the different expressions in all training images.

Colmenarez et al. [4] introduced a Bayesian framework for joint FR and FER. This method utilizes a face and facial expression that maximizes the likelihood for any given face image. Separated geometry and texture information of a face image for joint FER and FR was proposed by Li et al. [5]. These two types of information are projected onto different PCA spaces, and these spaces are structured to capture the characteristic features of different individuals; the separation of textural and geometric features was done by fitting a face mask and warping the texture on it. N-mode Singular Value Decomposition (N-SVD) was introduced by Vasilescu et al. [150] to separate the identity (for FR), and pose and expression (for FER). A modified version of this approach was used by Wang et al. [2]. The Higher-Order Singular Value Decomposition (HO-SVD) models the mapping between identity and expressions. The simultaneous face and expression recognition is performed as a result of facial expression decomposition.

5.2 Proposed method

In this thesis, we suggest a solution for expression-invariant face recognition which is consistent with the approach applied to expression recognition. This is different from the referred above approaches that have two independent solutions for both tasks. Such approaches are characterized by high complexity of the system and the response time, and therefore, are disadvantageous for real-time applications.

In our approach, we perform facial point detection, feature extraction and classification for both tasks in parallel. Figure 5-2 shows the flowchart of the system. The slight difference between FR

and FER is in the feature extraction block. They both use a GL filter but with different parameters. Since the nature of expression and identity recognition is distinct, the features extracted for each task are different. Thus, we use different parameters for GL filter (n, j, k), and build the filter “off-line”. The parameters for the FER are as follows: $n = 2, k = 1, j = 1$, while the parameters for the FR are $n = 2, k = 4, j = 2$. Figure 5-3 shows the components of the filter used for the FR, and Figure 5-4 shows sample of output of the filtered image from CK database. For each test image, the filter is convolved with pre-processed image and passed through the same classifier. The same normalization as discussed in 4.4.2 is used here.

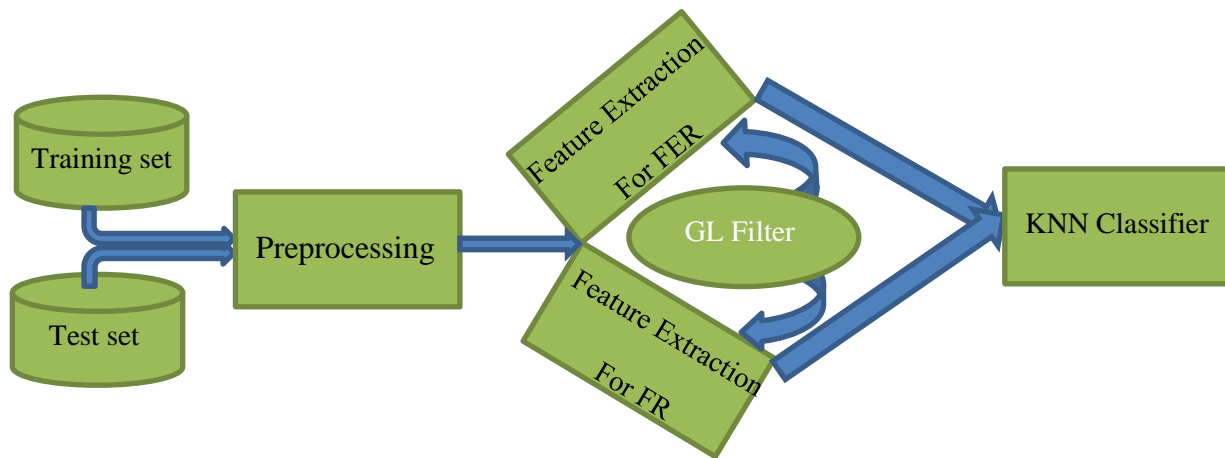


Figure 5-2. Flowchart of the joint FER and FR system. Feature extraction for FER and FR use the GL filter but with different parameter.

Similar approaches for textural feature extraction for FR normally use Gabor or Log-Gabor filters [156] [157]. The Gabor filter is a powerful tool for texture encoding, but in order to have rich feature vector we have to use bank of Gabor filter. Different Gabor filters with different orientations construct a bank of filters, which is used for feature extraction although in our case

we only use “one” filter. Normally the number of filters in Gabor bank changes from 15 to 40 [95]. Figure 5-5 shows bank of Gabor filters used in [120] for face recognition.

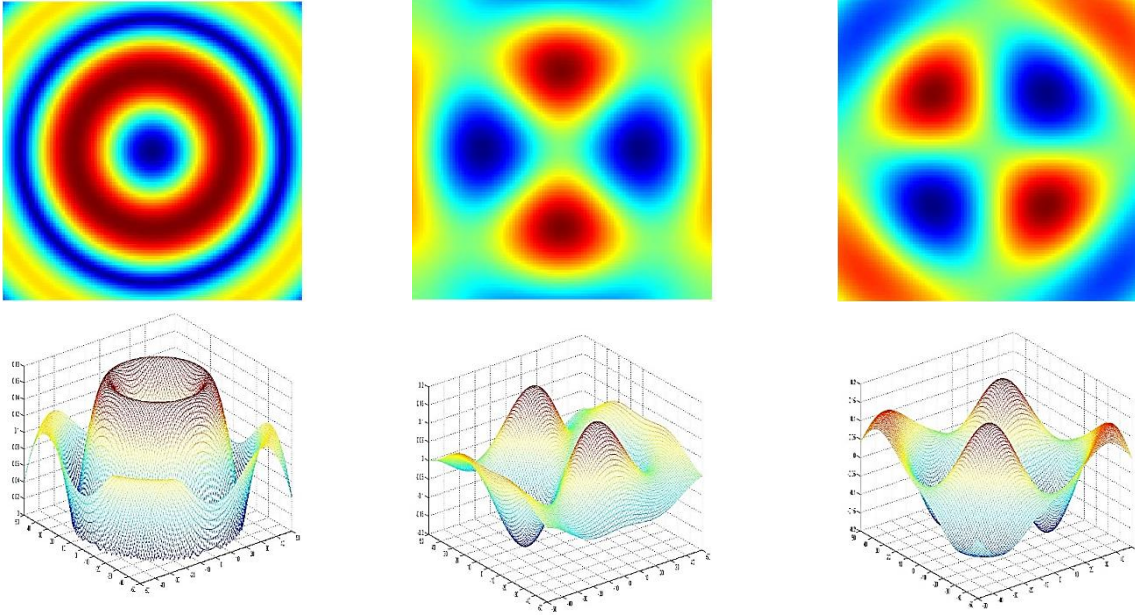


Figure 5-3. Gauss-Laguerre filter used for FR: Left-to-Right, Top-to-Bottom: Absolute, real, and imaginary part of the filter. The 3D representation of the absolute real and imaginary filter.

5.3 Experimental results

In our experiment, the face recognition has been tested on datasets of face images with expressions. The goal of the experiment is to compare the face recognition rate under varying expression from different types of classifiers.

The parameters of the GL filter should be properly tuned in order to achieve the best face recognition rate. Similarly to the approach used in Chapter 4 for FER, the recognition rate on training set has been monitored while changing the GL filter’s parameters.

The two databases, JAFFE and CK, which have been used in the previous chapter, were used here for the face recognition. Sample of images from the two databases are shown in Figure 5-6.

The number of images in each database is the same we used in FER (JAFPE: 213 images, and CK: 4074 images).

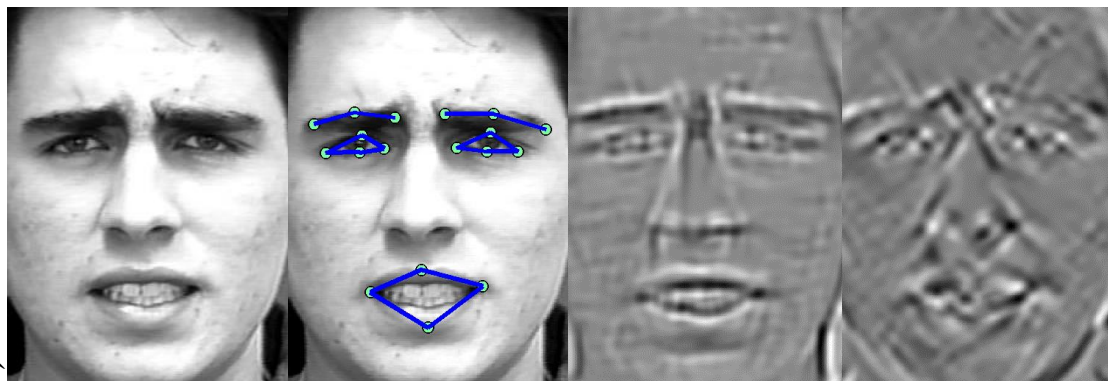


Figure 5-4. The textural and geometric features: Left-to-Right, original image, geometrics distances, real part of filtered image, imaginary part of fileted image (©Jeffrey Cohn [51]).

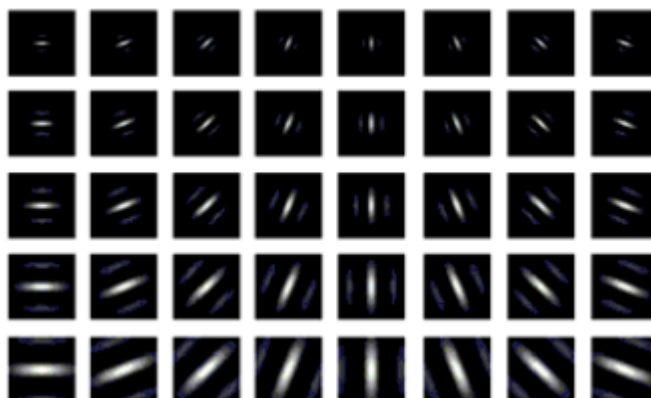


Figure 5-5. The real part of the Gabor filters with five frequencies and eight orientations [120].

The normalization process, explained in Chapter 4, was applied on all images. On the other hands, the face images are well-cropped, normalized to have fixed distances between eyes, resized to 128×96 pixels and aligned based on the line that connects two eyes together. This procedure is shown in Figure 5-7.



Figure 5-6. Sample images in JAFFE, CK and MMI databases used in the experiments.

The same feature vector we used for the FER (384 textural feature and 15 Euclidean distances, in total 399) is used for the FR. The appropriate coefficients for textural and geometric features are 0.58 and 0.42 respectively (Chapter 4, section 4.2).

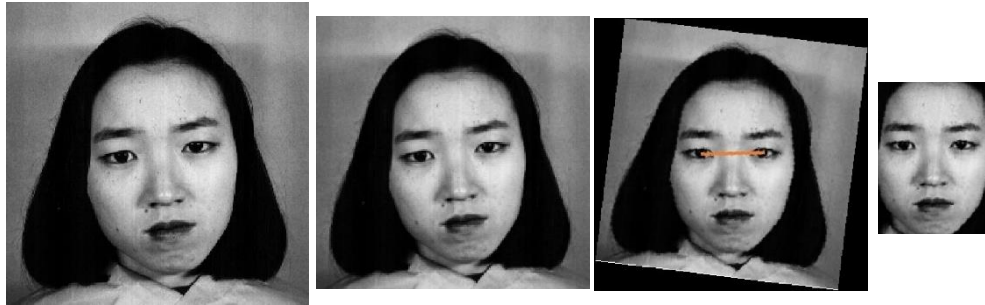


Figure 5-7. The normalization process in face recognition: (left to right): input image (from the JAFFE database), normalized image to have fixed distance between eyes, aligned image to have horizontal line between two eyes, resized image to 128×96 pixels.

5.3.1 Experiments on JAFFE database

For this dataset, we followed the set-up used in [6] [148] for face recognition. N images per subject have been randomly selected in order to construct the training set. The remaining images per subjects are used as test set. Ten female subjects who have participated in this database posed

3 or 4 examples of each of the six basic expressions (happiness, sadness, surprise, anger, disgust, fear), in addition to a neutral face for a total of 213 images of facial expressions. The three scenarios, used for expression recognition in Chapter 4, were repeated for face recognition as well. Two new scenarios, as suggested in paper [6], were also added to the experiment. In the first one, only the neutral image was used as training image, and all other images with expressions are used for testing. In the second one, the number of images for training varies from 2 to 5, regardless of expressions. The other three scenarios (leave-one-out, cross validation and expresser-based) are similar to the previous chapter.

The obvious observation is that using a bigger number of images per subject in the training phase improves the recognition rate (See Table 5-3). We compared the face recognition results demonstrated by our algorithm on JAFFE database, against several other approaches reported in literature (Table 5-1). It should be noted that no results were published on JAFFE database for both Face recognition and facial expression recognition. The results on JAFFE database for FER has been reported in Chapter 4. The papers mentioned in Table 5-1, only investigated the FR on face images (videos) with expressions. Again, the main focus of this thesis is to perform both FR and FER at the same time with minimum cost in terms of time and complexity. Therefore, having a fair comparison is not possible, as some approaches may have better results on FR that was their main purpose. For example, the perfect face recognition rate of 100% was achieved in [158], since they use 45-50% of the faces for the database for training.

The results of FR experiment using our approach, are shown in Table 5-2. We conducted more comprehensive experiment than those reported in Table 5-1. The leave-one-out approach shows the perfect recognition rate 100%, since the number of training images was 143, and number of

testing images was 70 (one image for seven expressions per subject); such training set is quite rich to handle different expression for FR.

For cross validation, likewise leave-one-out, the recognition rate is very close to 100%, since the training dataset contains mostly all expression from all subjects. The only reason to not having 100% is that the partitioning is a random process here, and there could be case that none of the expressions from a subject was included in the training set (only neutral expressions could be chosen).

The Expresser-based corresponds to a “real situation” scenario, since the test image from a subject is not included in the training set. On the other hands, the system is trained on one set, and is being tested on a different set. Although the left out partition is not technically from other set (or database), it is totally new to the system. The lower recognition rate, compared to the two aforementioned approaches, represents this situation accordingly.

As mentioned earlier, no paper was found that reports the same approach as proposed in this thesis, for classification-rate comparison. We have found that the following two approaches are the closest to our approach, therefore, we compare them against our technique in detail (Table 5-3). In the first approach, neutral images of subjects were used for training; this is the same approach as adopted by many others [158] [159] [160]. This kind of training shows the ability (robustness) of the system to handle different kind of expressions, while the system was trained on expression-free images. In practice, many recognition systems use images in different pose and illumination conditions for the training, and do not consider expressions. Compared to the papers that use such approach, our system demonstrates the similar FR rate, but it also accomplishes FER.

Table 5-1. Comparison with different approaches on the JAFFE database

Paper	Technique	Average FR rate	Average FER rate	Database Setup
[158]	Polyharmonic Extreme Learning Machine	100%	N/A	45-50% for training
[161]	Curvelet Transform	98.53%	N/A	2-5 images per subject for training
[148]	Compressive Sensing	96.12%	N/A	2-5 images per subject for training
[162]	Zernike Moment Features	97.6%	N/A	7 images per subject each per expression
[163]	Probabilistic Linear Discriminant Analysis	96.6%	N/A	4 images per subject in test and the rest for training
[164]	PCA + Fisher Linear Analysis	98%	N/A	66% of all images for training
[159]	SIFT-Based Feature Extraction	96.55%	N/A	1 image per subject (Neutral images)
[160]	Elastic Local Reconstruction	97.5%	N/A	1 image per subject (Neutral images)
[160]	Gabor Filter	97%	N/A	1 image per subject (Neutral images)

5.3.2 Experiments on CK database

For this database, there are some published results for joint face recognition and facial expression recognition and, therefore, we are able to fairly compare our results. Again, the main advantage of our algorithm is that the expression analysis can be done together with face recognition at almost no additional cost. To show this advantage, CK dataset is chosen, which is normally used for expression recognition. 375 image sequences have been selected from 97 subjects, such that each video sequence contains the frames labeled as one of the six basic

emotions, with the video clip being longer than ten frames. The best paper to compare our approach against is [6]. It used only 25 subjects for 4 expressions out of 6. Again, in the results published on CK database, it was not mentioned which subjects (or image sequences) have been selected. To be consistent with the approach proposed in [6], we performed the same three different set-ups they used.

The last method is using different number of images (2 to 5) per subject for training. Since it is not clear whether the other papers used neutral images as one of the training images or not, we randomly selected the images for training and repeated the experiments 40 times. The results are shown separately in Table 5-3. The recognition rate of the proposed approach are in the same range of the published paper. It also shows a minor improvement when the number of training images is higher.

Table 5-2. Face recognition accuracy (%) on the JAFFE database for various approaches

Approach	Average FR Rate (%)	Average FER Rate (%)
Leave-one-out	100	96.51
Cross validation	99.89	94.96
Expresser-based	90.66	91.03
Neutral images for training	96.83	N/A

Table 5-3. Face recognition accuracy (%) on the JAFFE database for the approaches in [111] [148] and the proposed one

# training	S2			S3			Proposed		
	High	Low	Average	High	Low	Average	High	Low	Average
2	97.22	86.11	92.84	95.56	83.89	90.94	95.92	88.43	91.74
3	98.82	94.12	96.71	97.65	91.76	95.88	98.98	93.07	96.56
4	99.38	95.63	96.94	91.12	93.75	96.62	100	95.44	97.04
5	100	96.67	98.53	98.67	96.67	97.80	100	97.20	99.11

In the first set-up, three sequences per subject were randomly selected for training, and the rest of sequences were used for testing. This process was repeated 10 times and the average recognition rate for face and expression was reported. In the second set-up, expression-based approach were used. One subject was removed from the training set, and the rest of subjects were used for training. In the third set-up, the neutral images were used for training, and images with expressions were used for testing. The database reported in [6] is CK+, which is slightly larger than CK database, but the authors only used a very small part of the database (25 out of 106 subject). Therefore, our experiments is more comprehensive. The results from the first and the third set-ups are compared against the ones reported in [6] which include the following approaches: Component-based Dictionary Learning (DCS), B-JSM (B-Joint Sparsity Model) [148], KSVD (K-Singular Value Decomposition) [165] and FDDL (Fisher Discrimination Dictionary Learning) algorithms [166].

The component-based learning uses a dictionary-based component separation algorithm. Each image with expression is decomposed into a neutral and an expression part, and a dictionary-based learning is used for classification. Joint Sparsity Model (JSM) is normally used for efficient coding of multiple inter-correlated signals. B-JSM is a modified JSM to be applied to 2D images. It finds the components of all training images of any specific class. K-SVD uses singular value decomposition approach to create a dictionary for sparse representations for dictionary learning. Fisher Discrimination Dictionary Learning (FDDL) is a dictionary learning approach where a structured dictionary is learned so that after sparse coding, the reconstruction error is used for pattern classification.

Table 5-4 shows the comparison between the face recognition results using the above approaches. Our algorithm does not demonstrate the best recognition rate but it has the second rank. It should be noted that the database that we used for training and testing, is 3 times bigger than the others.

Table 5-4. Face recognition rates (%) on the CK and CK+ dataset

Set-ups	Proposed	DCS	B-JSM	KSVD	FDDL
First	96.76	99.14	85.2	85.6	98.8
Third	93.17	95.1	81.5	91.2	95.3

In the next comparison experiment, we implemented the approaches reported in literature, and our approach, on the same testing platform, - MatLab. The source codes for FDDL [167], KSVD [168], and JSM [169] were obtained from the available online repositories. The DCS approach was implemented in this work. The images we used in the experiments for all expressions have been utilized to test all the 4 other approaches. The results are reported in Table 5-5 and are different from the ones shown in Table 5-4. The performance of our approach is almost similar to the best one (DCS) in the first set-up, and it is also better than DCS in the third.

Table 5-5. Face recognition rates (%) on the CK dataset

Set-ups	Proposed	DCS	B-JSM	KSVD	FDDL
First	96.76	97.14	81.43	82.37	94.52
Third	93.17	92.33	76.54	89.55	91.24

The results of FER, face expression recognition, are also compared against the four aforementioned approaches, on the CK and CK+ databases. The results from both first and the second set-ups are reported in Table 5-6. As shown in Table 5-6, the best face expression recognition belongs to [144] which is only 0.1% better than ours; note that [144] can only

perform face recognition. Our results are better than many of the reported ones and very close to the best one [6], while we tested on a larger database. Many of these techniques used several features for FER, and the trained classifier like SVM, Adaboost and Neural Networks [6]. Our technique only uses a single GL filter with a simple KNN classifier, which saves memory and decreases response time of the implementation.

The final experiment regarding the comparison of our approach against the best reported one [6] was focusing on the effect of different expressions on face recognition rate using CK/CK+ database. The FR rate was calculated by keeping out each expression from the training set, and it was used only for testing. In Taheri's approach [6], the "angry" and "sad" faces are the easiest expressive faces to recognize, while "surprise" is the most challenging one. In our approach, "happy" and "surprise" are the easiest one while "fear" is a challenging one for FR (Figure 5-8).

5.4 Conclusions

This chapter has investigated the proposed method for face recognition using GL wavelet. The FR requires the same approaches as FER at all steps (preprocessing, feature extraction and classification). The slight difference between FR and FER lies in feature extraction. However, it is the GL filter that was used, albeit with different parameters, to provide textural features for both tasks. Compared to a few available methods that can handle both FR and FER at the same time, our system provides a low cost, one-step feature extraction technique for both classifications. Extensive comparison with other approaches on two different databases showed the efficiency of the proposed system in various classification scenarios. The average recognition rate of 96.76% was achieved for CK, and 96.61% for JAFFE.

Table 5-6. Comparison of various approaches on the CK/CK+ databases for face expression recognition

Paper	Technique	Average FER rate	Average FR rate	Database Setup
[170]	Gabor filter + SVM+ Multinomial Ridge Logistic Regression	91.5%	N/A	313 sequences (614 frames) of CK
[171]	Ratio-image based appearance feature + GMM	87.6%	N/A	47 subjects (2981 frames) of CK
[51]	AAM + Similarity-Normalized Shape and Canonical Appearance Features +SVM	83.15%	N/A	On 5 expressions, 327 sequences of CK+
[147]	RegRankBoost	88%	N/A	96 subjects of CK
[144]	Adaboost + Local Patches, Combined features	92.3%	N/A	352 sequences from 96 subjects of CK
[1]	Nonlinear Decomposable Generative Model	70.85	N/A	16 subjects of CK
KSVD-First set-up [165]	KSVD Dictionary Learning	49.2	85.6	25 subjects of CK+
KSVD-Second set-up [165]	KSVD Dictionary Learning	64.6	-	25 subjects of CK+
FDDL-First set-up [166]	FDDL	73.7	98.8	25 subjects of CK+
FDDL-Second set-up [166]	FDDL	73.6	-	25 subjects of CK+
DCS-First set-up [6]	Component-based Dictionary Learning	81.64	99.14	25 subjects of CK+
DCS-Second set-up [6]	Component-based Dictionary Learning	86.8	-	25 subjects of CK+
DCS-Second set-up [6]	Component-based Dictionary Learning	89.21	-	106 subjects of CK+
Proposed-First set up	GL Filter + KNN	92.2	96.76	375 sequences from 96 subjects of CK
Proposed-Second set up	GL Filter + KNN	91.28	93.17	375 sequences from 96 subjects of CK

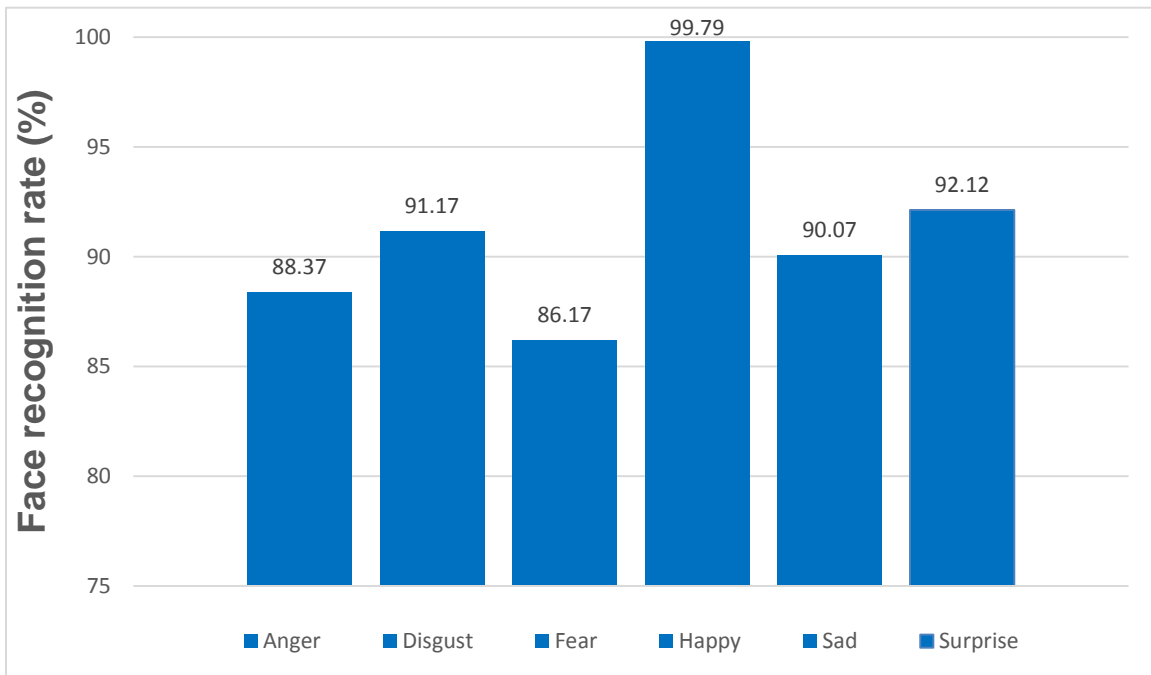
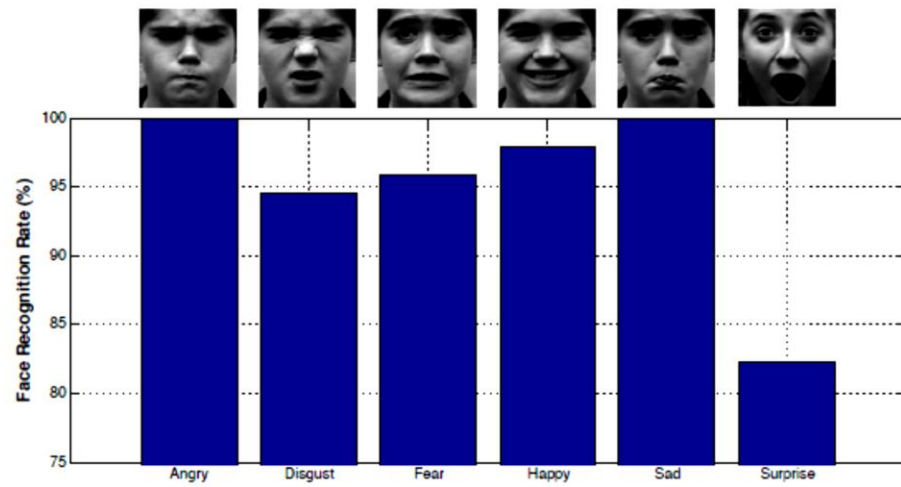


Figure 5-8. Face recognition variation by keeping out different expressions. The top image is from [6] and the bottom image is from our approach.

Chapter Six: **FACIAL EXPRESSION RECOGNITION FROM INFRARED IMAGES**

In this chapter, we focus on using thermal images for FER, based on the approach we proposed for visible images. The combination of infrared and visible FER, using a common feature extraction approach, has also been studied, and the results show certain improvement in the recognition rate. Two common databases (that have both visual and infrared images of the same person) have been used in the experiments.

Many of the available methods utilize visual information for facial expression. However, the main problem is illumination variation in uncontrolled environments. Thermal infrared imaging is a suitable solution, as it is not sensitive to illumination variations, since only the emitted radiations from the skin surface are recorded by a thermal camera. Compared to visible images, infrared images reflect temperature distributions of face, while the details on face are blurry. Hence, the texture can be retrieved, using these distribution changes, indicating positions of facial muscles and blood vessels. Such changes occur mainly around forehead, nose, mouth and cheeks.

Using thermal images for facial analysis has two drawbacks. First, not all facial features and texture information are available in infrared images, especially around eyes and mouth. Second, wearing eyeglasses results in dark pixels in thermal image. However, the visible images supply suitable properties, which are indispensable in any automated FER system. The best case scenario is using the “good” attributes of both spectra with the minimal added complexity. This approach is studied in this thesis

6.1 Previous works

Several papers have addressed FER using infrared thermal images. Khan et al. [172] used statistical features with Linear Discriminant Analysis (LDA) classifier for both posed and

spontaneous facial expressions. The face thermograms were used as a reference for comparison with different expressions. Thermographic cameras detect radiation in the infrared range of the electromagnetic spectrum (9–14 μm) and result in images of that radiation, called thermograms. Each thermogram was divided into grids of 128 squares, and for each square the highest temperature was recorded. Then 75 out of 128 recorded measurements around specific points on face used as features for classification.

Trujillo [173] used an unsupervised local and global feature extraction technique. Four regions including the left and the right eye, the nose, and the mouth were selected out of infrared image and features from Eigenimages were extracted to be classified by several Support Vector Machines (SVMs). Eigenimages are based on Principal Component Analysis (PCA), where images are projected onto a lower dimensional space that spans the significant variations (Eigenimages) among known face images.

Shen et al. [174] performed infrared video analysis. Statistical features were extracted from the horizontal and vertical temperature difference sequences of different facial sub-regions and the Adaboost algorithm, with the weak classifiers of KNN, was used for classification. The face was divided into four regions of interest, including the forehead, the nose, the left/right cheek and the mouth. Next, the difference between each two pixels was calculated on each frame in horizontal and vertical direction. Each of these four region data was used to find four static statistic parameters (minimum, maximum, standard deviation, and mean). Correspondingly, these statistic feature sets were calculated for the whole sequence. The optimal feature set was constructed using F-value feature selection method. Finally, the Adaboost algorithm, with the weak classifiers of KNN, was utilized for classification.

Koda et al. [175] used speech recognition to localize specific frames from thermal video at the three timing positions before speaking, and speaking phonemes of the first and last vowels. 2-DCT was utilized to extract features from the specific regions of face.

Wang et al. [176] used wavelets for feature extraction. A 2-level Haar-4 wavelet was applied to the image, which decomposed it into 8 frequency bands. The mean image of different expressions from training set was constructed and the Euclidian distance of each single image from the mean image was calculated. A neural network was used for classification of different expressions based on the obtained distances, in order to maximize the distance between different expressions.

6.2 Proposed Method

A new fast and accurate feature extraction and facial feature detection method for FER in visible imagery system was proposed in Chapter 3. The system uses both textural and geometrical features, derived based on Gauss–Laguerre wavelets, and the positions of 18 selected points of the eyes, eyebrows and mouth for visible images in Chapter 4. The KNN was used as a classifier. The response time and accuracy are the two main factors. Thus, the ideal case is to have the same feature extraction and classification approach for both visible and infrared images. As the nature of captured image in both spectra is totally different, we cannot use the exact same features for both types. The optimal case is having the same source for feature extraction (GL filter), but with different parameters. Thus, the computational time can be reduced significantly in real-time analysis.

In this Chapter, we propose a facial expression recognition method using infrared thermal images, according to the scenario given in Chapter 4, but with different parameters. To demonstrate the feature extraction with GL filter in infrared images, the facial features are

detected manually, but in real time application since both the cameras are synchronized together, and the facial points are detected based on the method proposed for visible face images. Textural features are extracted using GL filter from different regions on the face. A KNN classifier similar to the one, applied to visible images, is utilized as well. Many of the published methods are based on a small set of infrared database, ranging from 4 to 30 subjects, while few of them use a larger database [174]. In this research we use two different databases with large number of subjects.

6.2.1 Database

The OTCBVS dataset [177] has images of 30 subjects taken in the UT/IRIS Lab. Three different expressions, including surprise, happiness and anger in 11 different viewing angles, are captured from subjects. The resolution of each color images in the database is 320x240 pixels. The database also contains images with the subjects wearing glasses. Likewise most other published approaches tested on this database, we selected to use the best frontal views images in the experiments. Eight subjects that have eyeglasses were excluded manually from the database, along with a subject showing only two expressions. 195 frontal view images including 66 “surprise”, 66 “anger” and 63 “happiness” expressions, in addition to two near-frontal view, have been selected out of the database. In total, 585 images are used in the experiments. Figure 6-1 shows examples of this database.

The USTC-NVIE database [178] has images of 215 subjects (157 males and 58 females), ranging from 17 to 31 years old, collected at the University of Science and Technology of China. The spontaneous expressions consist of image sequences from onset to apex, collected by a visible and an infrared thermal camera at the same time under front, left, and right illumination. The available database contains images of 103 subjects under front, 99 subjects under left, and

103 subjects under right illumination, for spontaneous expressions and 107 subjects for posed expressions. Figure 6-2 shows few samples of these two databases.



Figure 6-1. Samples of OTCBVS database. From top to bottom: surprise happy, and angry expression.

6.2.2 Feature extraction

Figure 6-3 illustrates the block diagram of the approach, which includes the following steps: preprocessing, feature extraction, and classification. In order to be consistent with the visible images, we want to use the same feature extraction and classifier with the minimum modifications. As the nature of visible and thermal images are totally different, it is very difficult to use the “exact” features for both. In all published papers, mainly the statistical features were used for infrared images which are different from the techniques used for visible images (see Chapter 4). One of the novelties of this thesis is to introduce a “unique” technique, which can be applied on both imaging spectra with slight changes.

Prior to finding facial features on both visible and infrared images, we need to “calibrate” both imaging systems. Stereo camera calibration is well-studied and implemented in many literatures. The translation and rotation between two cameras can be found, so that the corresponding pixel

of the visible image in infrared image is determined. By performing the calibration of both infrared and visible cameras, the extracted points from visible image can be translated onto the infrared images automatically, such as shown in [179].

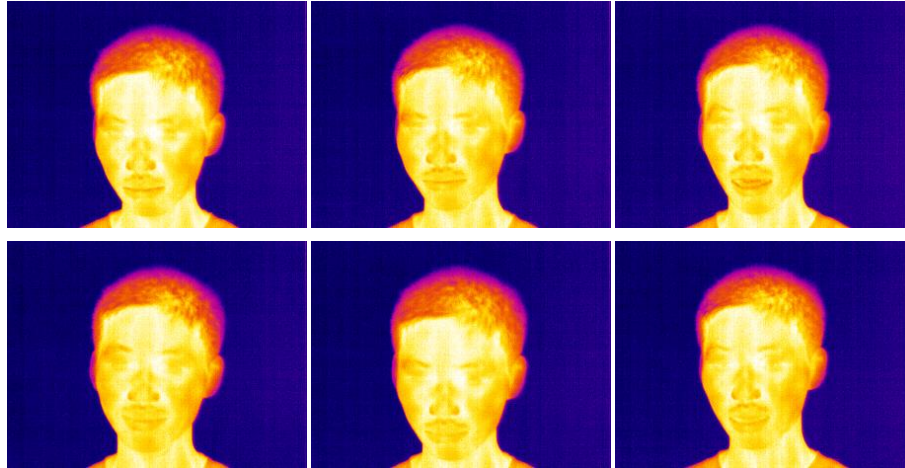


Figure 6-2. Samples of USTC-NVIE database. From top to bottom, left to right: anger, disgust, fear, happy, sad, and surprise expression.

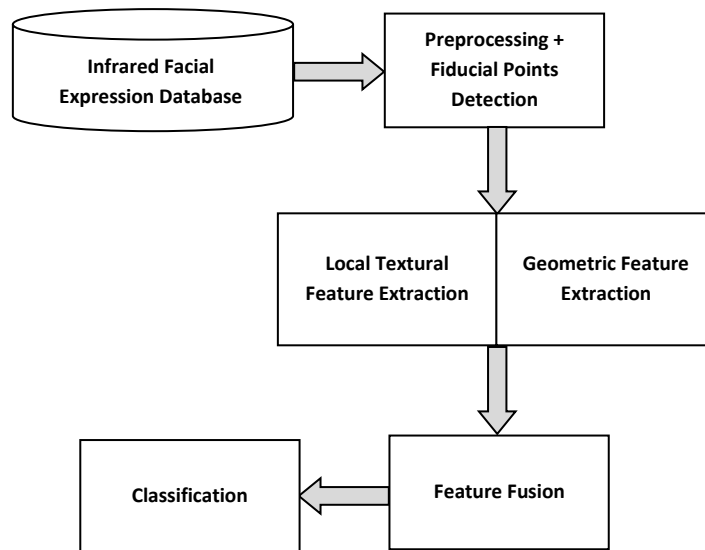


Figure 6-3. Flowchart of the proposed method.

There are several algorithms to localize facial features from thermal images [180]. In order to demonstrate the ability of GL filters for efficient feature extraction, the face image is first cropped, and then fiducial points are extracted manually. However, in real world, as we have multiple cameras (visible and thermal), the facial points can be extracted from infrared images by using the same approach applied for visible images.

In our approach, twenty one points are extracted from regions of interest as shown in Figure 6-4(a-b). These points are imposed here on a visible face image for better visualization. The texture information around these points is coded by a GL filter. The relation between these points constructs the geometric features. The final feature vector is the combination of these two types. The details of GL filter has been already discussed in Chapter 4.

We used 21 points of interest on each face image. A 3x3 mask around each point is filtered by the designed GL filter. Therefore, the total length of textural feature vector is $21 \times 3 \times 3 = 189$. The parameters of the filter (n, j, k, a) are selected based on the best recognition rate, obtained from several experiments with different parameters.

The coordinates of these 21 points are used to construct 13 Euclidean distances, as shown in Figure 6-4(c). Different expressions result in different deformations of the corresponding facial components, especially near eyes and mouth. The final feature vector is a combination of the two feature sets with the size of 202 (189+13). However, the importance of textural and geometric features is different. Thus, these two types of feature sets are multiplied by two different weights, where the appropriate weights are obtained through trial and error. The weighting procedure is similar to the one explained in Chapter 4. The average recognition rate versus different weight coefficients for geometric features is monitored during the experiments using the leave-one-out

approach in three trials. The best average recognition rate, within coefficients, were determined to be 0.72 for geometric and 0.28 for textural features.

6.2.3 Classification

In order to integrate the proposed module with the module for visible facial expression recognition, we used the same classifier of expressions, K-Nearest Neighbor. This classifier does not make any assumptions regarding the underlying data distribution. The number K defines how many neighbors influence the classification. This number is set to 3, according to our previous experiments on visible images. In our experiments, we excluded all subjects wearing eyeglasses. Only posed expression images have been used.

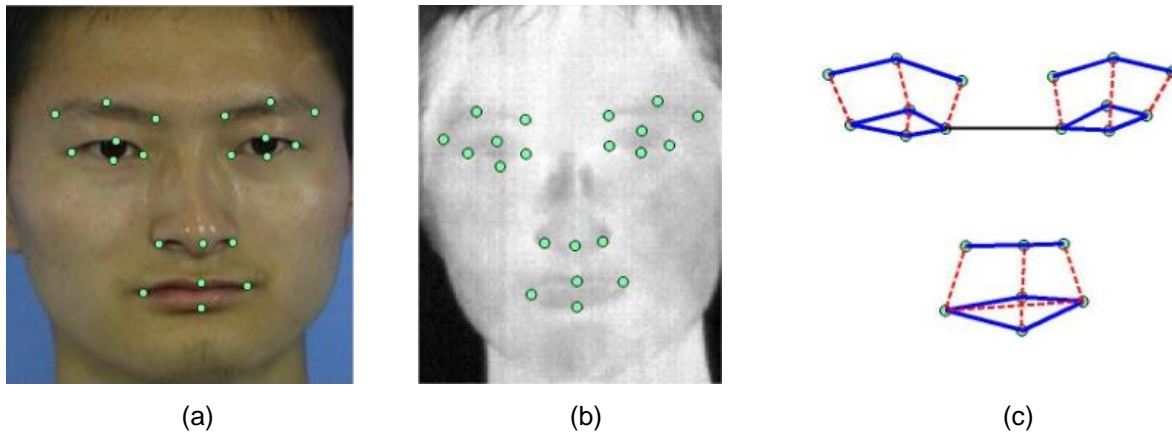


Figure 6-4. (a-b) Selected points of interested on face, (c) The distances used as geometric features from the defined mask.

6.3 Fusion

To benefit from both visible and infrared images, we use fusion at feature level for FER. At the same time, we keep the complexity of the system to be the same level as the one that uses only visible images. This is because the feature sets of both spectra come from a “single” feature extractor. It allows reducing the complexity of the system and speed up the system response. To the best of our knowledge, the available papers in fusing visible and infrared images use different

techniques for feature extraction, and this needs more memory, time and overall is more complex in implementation. In addition, facial feature detection and classification are performed simultaneously for both visible and infrared images. The combined geometric features, applied on a face mask and the textural features based on Gauss-Laguerre filter from both spectra are fused, and a single KNN classifier is used for classification. The fusion enhances the recognition rate with a minor increase in complexity, compared to the results that use only visible images. The proposed approach is tested on USTC-NVIE database that has both visible and infrared images for each subject. Figure 6-5 shows the flowchart of our approach, which consists of image acquisition and calibration, facial points detection, feature extraction, and classification.

After camera calibration, the facial features are extracted using the same approach as for visible images. Total of 21 points of interest (Figure 6-6) are extracted from face image. For both infrared and visible image, a 9x9 mask around each point is filtered by a tuned GL filter. We use the geometric features only once from visible image and then fuse it with the textural features from both spectra. In this case, the length of textural features is $21 \times 3 \times 3 = 189$ for each image. The parameters of the filter (n, j, k, a) are selected based on the best recognition rate obtained from several experiments with different parameters [174]. These parameters are different for visible and infrared, but allow for “one-time-job”. On the other hand, based on the parameters of the filter, the filter is generated off-line and can be stored as a complex-valued matrix. For each given image, the feature vector is a convolution of the matrix with the image. For both infrared and visible image, the feature extraction comes from a single GL filter, and not from two different systems.

The coordinates of the 21 points of interest are used to construct 20 Euclidean distances, as shown in Figure 6-7 with small arrows. These distances represent deformation around eyebrows,

eyes, nose and lips. This information helps to build an extended feature vector. In total, there are 20 geometric distances, as these distances are identical in both spectra. The approach is exactly similar to the case when using only infrared images but with different mask.

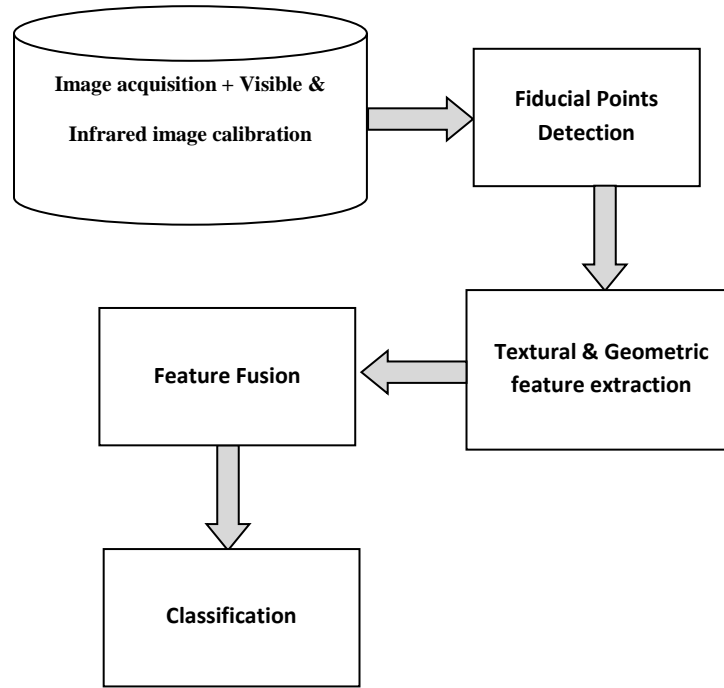


Figure 6-5. Flowchart of the proposed method for fusion at feature level.

The final feature vector for each image is a combination of the two feature sets with the size of $189+20 = 209$. However, the “significance” of textural and geometric features is different. Based on the conducted experiment, these two types of feature sets are multiplied by two different weights, and the values for the weights are obtained through trial-and-error approach. The average recognition rate versus different weight coefficients, for geometric features, has been monitored during the experiments using the leave-one-out approach in 10 trials for both infrared and visible. The best average recognition rate, within coefficients, was 0.715 for geometric and 0.285 for textural features in infrared images. For visible images, these coefficients are .707 for

geometric, and 0.293 for textural features. Figure 6-9 shows examples of applied GL filters with different parameters.

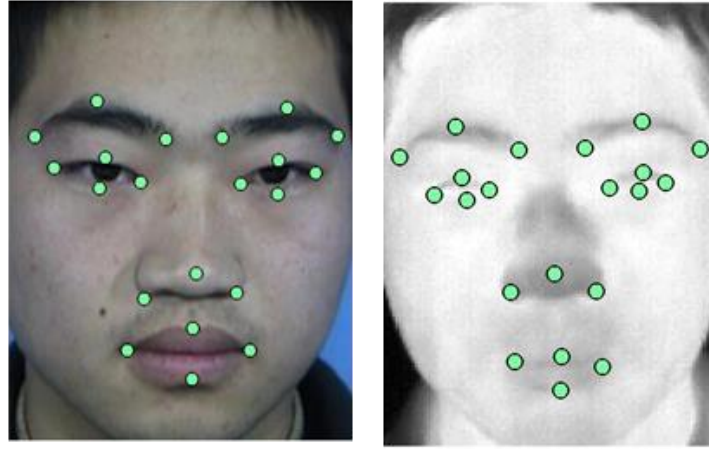


Figure 6-6. Facial points extraction for both spectra.

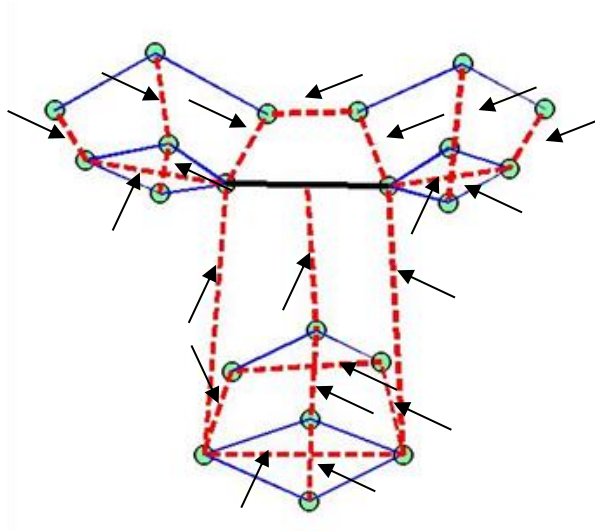


Figure 6-7. Geometric mask used for feature extraction.

The feature fusion in this thesis is pixel-based. We keep all the geometric facial features (20 features). Out of $2 \times 189 = 378$ textural features (which are complex-valued number), we compare the phase of each feature in both spectra. On the other hands, let us assume $\alpha_i \in$

{textural features from the infrared}, and this set has 189 elements. Also, $\beta_i \in \{\text{textural feature from the visible image}\}$. The algorithm applied for feature fusion is given below:

```

for any  $i = 1, 2, \dots, 189$  ,
    if phase of  $\alpha_i > \text{phase of } \beta_i$ 
        Select  $\alpha_i$ 
    else
        Select  $\beta_i$ 
    end
end

```

The final fused feature vector consists of 20 geometric features and 189 textural features, 209 in total. This feature vector is used in the classification step. Figure 6-8 shows the procedure. The same K-Nearest Neighbor classifier is used again here. This classifier is widely used in facial expression classification [174] [13] [181] [47]. In the classification procedure, the training data is first plotted in n-dimensional space, where n is the number of features. Each of these consists of a set of vectors labeled with their associated class (arbitrary number of classes). The number K is set to 3 according to our previous experiments on visible and infrared images [12] [10].

6.4 Implementation and results

There are two sets of experiments. First, the two databases (OTCBVS and USTC-NVIE) that have infrared images are used to show the performance of the proposed method for infrared images. Then, USTC-NVIE database is used in fusion experiments, as this is the only available database which has both visible and infrared images with all six expressions.

6.4.1 Experiment on IR images

OTCBVS: As the size of this database is small, in order to evaluate the performance of the proposed algorithm, the cross validation (CV) approach is selected to avoid overfitting. In the CV approach, all the images are used to train and test the classifier, while the validation is performed by using subset of the images which is not used in the training process.

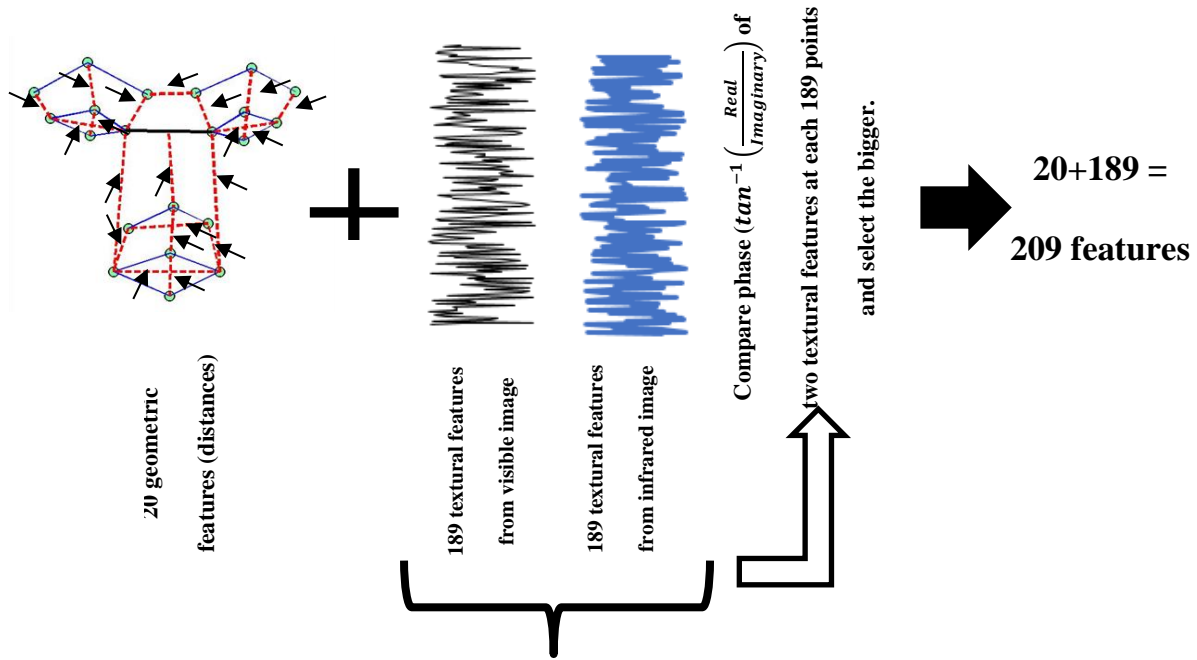


Figure 6-8. Flowchart of fusion technique.

The database is randomly partitioned into ten distinct segments, and nine partitions are used for training, while the remaining one is used to test the performance. The procedure is repeated, so that every equal-sized set is used once as the test set. Finally, an average of ten experiments is been reported. The average accuracy of 95.33% was achieved for this database. The confusion matrix is shown in Table 6-1. The previously reported recognition rate, when using different subsets of this database, is 77% [173] and 96% [182], correspondingly. However, since the number of used images is not equal, it is impossible to compare the methods in terms of accuracy.

USTC-NVIE: In the experiment we used only the frontal posed expression images of 107 subjects. Some of the subjects do not have all six expressions. The distribution of six different expressions is as follows:

“Anger” = 90, “Disgust” = 88, “Fear” = 90, “Happy” = 90, “Sad” = 89, and “Surprise” = 86.

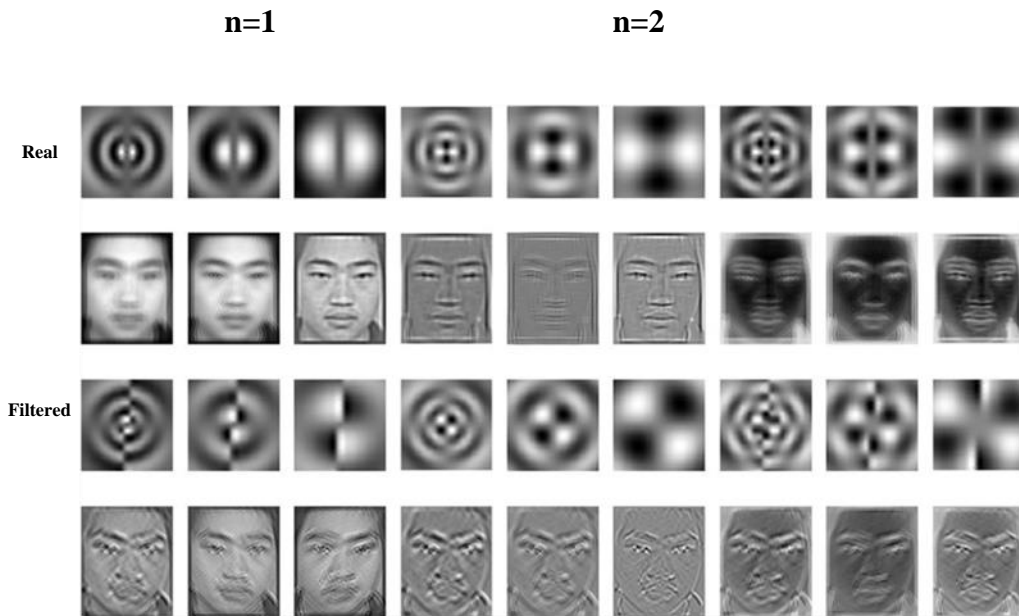


Figure 6-9. Samples of filtered face image with real and imaginary parts of GL filter with different values for n and j , $k = 4$, and $a = 2$.

Table 6-1. The confusion matrix – cross validation – OTCBVS

Expression	Happy	Surprise	Angry
Happy	98.43%	1.57%	0%
Surprise	4.36%	95.64%	0%
Angry	3.43%	4.65%	91.92%

In total, 533 images are used for the experiments. Two different methods are used for classification: leave-one-out and cross validation. In leave-one-out, for each expression from each subject, one image is left out, and the rest are used for training. In the CV, the same method as was used for OTCBVS database is applied. Table 6-2 and Table 6-3 show the confusion matrix for both approaches. To the best of our knowledge, there are no published studies

regarding using the infrared images from this database. The average recognition rate we achieved is 85.9% for LOO and 81.3% for CV. All numbers are shown using one-digit precision.

Table 6-2. The confusion matrix (%) – LOO–NVIE

EXP	AN	DI	FE	HA	SA	SU
AN	88.2	1.3	2.4	0	7	1.1
DI	2.5	85.3	4.4	0	5.1	2.7
FE	3.1	4.4	82.9	4.1	2.2	3.3
HA	0	1.6	0	91.3	1.5	5.6
SA	8.2	1.3	6.5	0.7	83.3	0
SU	3.1	3.5	2.1	2.8	4.2	84.3

Table 6-3. The confusion matrix (%) –CV–NVIE

EXP	AN	DI	FE	HA	SA	SU
AN	82.4	1.1	3.9	0.2	8.4	4.0
DI	2.4	81.8	5.2	1.1	6.8	2.7
FE	4.0	5.2	79.4	5.7	3.1	2.6
HA	0	2.4	1.1	87.1	2.3	7.1
SA	8.4	1.9	8.2	1.3	78.2	2.0
SU	4.3	3.7	3.3	4.2	5.4	79.1

6.4.2 Fusion experiment

Similarly to the described above infrared experiment, in the fusion study, we used the posed expression and only the frontal view images of 107 subjects. Some of the subjects do not have all six expressions. All subjects in the NVIE database, who wore eyeglasses, have been excluded from the experiment, since in infrared images the area of the eyeglasses is completely dark.

Table 6-4 and Table 6-5 show the confusion matrices for both approaches. To the best of our knowledge, no study was published on the posed facial expression recognition using the fusion

of visible and infrared images on this database for all six different expressions. The only published research is on spontaneous expression recognition, and only on three expressions (happiness, fear and disgust) [183]. The average recognition rate using our approach is 88.6% for LOO, and 84.2% for CV. All numbers are shown with one decimal digit precision. The previously published results on NVIE database did not use the exact number of images, and it is not clear which images have been exactly selected in the experiments. Thus, having a fair comparison with previous methods is impossible.

He et al. [184] used 532 samples of 123 subjects of three facial expressions (disgust, fear, and happiness) under different illuminations and just for infrared images. The average recognition rate of 51.3% has been reported. Wang et al. [176] utilized 236 apex images of 84 subjects for visible and thermal images, including 83 images for disgust, 62 for fear, and 91 for happiness expressions. Some non-frontal facial images were discarded in the experiments. The average recognition rate for visible images, using AMM and KNN, was reported to be 67.8%. This rate was reported as 41.49% for thermal images. Wang and Wang [185] used 535 images for spontaneous expression recognition under different illuminations for three expressions. The feature-level fusion algorithms showed 60.57% recognition rate. Liu and Wang [186] used 176 thermal sequences for three expressions. Statistical features along with a Hidden-Markov-Model classifier resulted in average recognition rate of 59% just for thermal images.

Compared to the expression recognition using the same infrared images from this database [12], the recognition rate was in your work 85.9% for LOO, and 81.3% for CV. The geometric mask used in using only infrared images is different from fusion approach. To have a fair comparison, the recognition rate with exact feature vector, while having only visible images, was calculated. The average recognition rate was obtained to be 86.3% for LOO and 82.8% for CV. The

confusion matrices with and without fusion using LOO are shown in Table 6-5 and Table 6-4, respectively.

Table 6-4. The confusion matrix (%) – LOO– without fusion on visible images

EXP	AN	DI	FE	HA	SA	SU
AN	89.8	1	2.2	0	6.3	0.7
DI	2.7	86.2	5.3	0.6	3.2	2
FE	2.8	3.7	83.7	3.2	3	3.6
HA	0	3.3	0	90.8	1.1	4.8
SA	6.7	4	4.1	2.3	82.9	0
SU	2.1	3.9	1.4	3.1	5.3	84.2

Our approach that uses fusion outperforms every single expression classification with the highest increase (3.1%) for “sadness” expression, and the lowest increase (1.9%) for “angry” expression. Figure 6-10 shows the comparison of each expression, using visible, infrared [12] and fusion of both types. For happiness and sadness, the recognition rate for infrared images, using the old geometric mask [12], is slightly better than visible images when using the proposed geometric mask. The reason is that the feature vector, derived in [12] in terms of geometric and textural features, is different from this paper and, thus, it is not a completely fair comparison. While performing the comparison between visible and infrared recognition rate, using the method proposed in this approach, the rate is always better for visible images.

Table 6-5. The confusion matrix (%) –LOO– with fusion

EXP	AN	DI	FE	HA	SA	SU
AN	91.3	1.1	1.9	0	5.4	0.3
DI	2.2	88.1	4.9	0.4	3.2	1.2

FE	2.4	2.9	85.8	3.5	2.5	2.9
HA	0	2.1	0	93.3	0.5	4.1
SA	6.4	3.1	3.4	1.1	86	0
SU	2.2	4.1	1.1	2	3.6	87

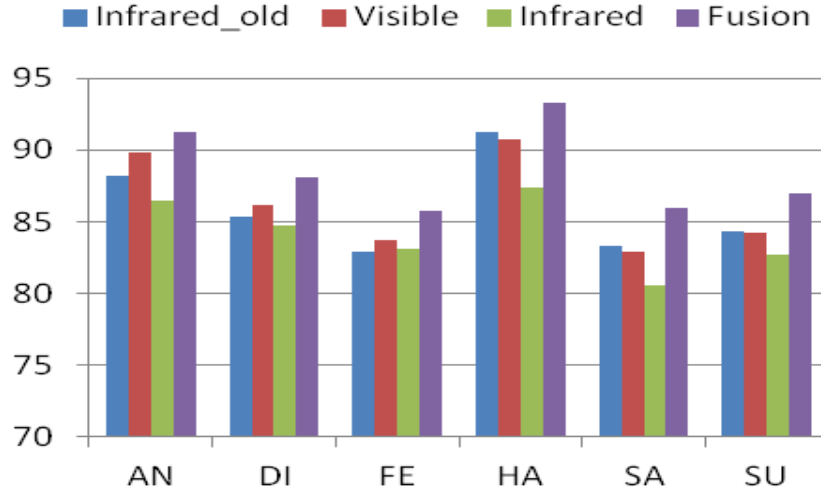


Figure 6-10. Recognition rate for different expressions with and without fusion.

6.5 Conclusions

In this chapter, we proposed a novel approach for facial expression recognition in infrared images. The aim of this study was to use the same approach that was previously applied on visible images for facial expression recognition, in order to demonstrate the benefits of illumination-invariance of thermal images, in the task of FER, with minimum modifications. The outcome of this method is decreasing the computational cost and increasing the speed of the system equipped with multi-band cameras for behavior analysis in real-time. The advantage of the proposed GL filter to extract features from infrared images of face is its rich frequency extraction capability for texture analysis, as well as being a rotation-invariant.

In addition, fusion of visible and infrared images for expression recognition has been studied. Most systems, that utilize fusion at the feature level, use partially and sometimes completely different algorithms to extract features from infrared and visible images. This increases the computational complexity, since it should perform two different feature extraction algorithms. In the proposed approach, the feature extraction for both infrared and visible image is performed, using a single GL filter. This is due to applying a uniform feature extraction technique for both spectra. Application of fusion results in outperforming the previously proposed approach, based on the same GL based feature extraction: the performance, in terms of FER rate, is better for infrared images by 2.7%, and by 2.3% for visible images.

Chapter Seven: **APPLICATIONS**

In this Chapter, we consider several applications where the behavioral biometrics such facial expression finds its use. Analysis of expression relates to understanding human behavioral pattern, mood and intentions, and is important in any system that is based on human-machine interaction. In particular, facial expression analysis is instrumental in monitoring and surveillance for security, for enhancing multimedia interfaces (such as monitoring of facial expression of users of computers during their interaction), as well as for biomedical applications (such as analysis of facial expressions in health care patients for the purpose of detection of pain, or analysis of facial expressions of bipolar patients).

7.1 Situational Awareness System (SAS)

Situational awareness refers to the perception and understanding of what is happening in a complex environment. Originally a military concept, it's now a design goal for many next-generation security systems, including ones that use human biometrics. For example, traditional biometric-based physical access security systems verify a person's identity based on commonly used physiological traits, such as a handprint, fingerprint, or iris scan, as well as some behavioral traits, such as voice. However, researchers are now working on authentication systems that can capture other physiological biometrics, such as skin temperature, as well as nonverbal behavioral biometrics, such as facial expression and gait. During access authorization, the system can compare some types of data to that of pre-screened individuals to identify criminal suspects or those on watch or no-entry lists. It can analyze other types of data to flag potentially suspicious behavior.

A situational awareness paradigm is spreading towards a vast variety of systems, including those that perform the analysis of appearance, physiological data and behavioral patterns of humans,

that is, biometrics. The DARPA research program, HumanID, aimed at the detection, recognition and identification of humans at a distance, in early warning support systems for force protection and homeland defense [187]. In Haifa airport (Israel), passengers are asked questions, while an officer runs speech analysis software to detect high agitation and nervousness. These examples are the first signs of the coming of a new generation called Physical Access Security System (PASS). New design concepts are required for the next generation of such systems to provide reliable authorization in a short amount of time [188]. In this approach, pre-screening is aimed at situational awareness, in addition to the primary application of biometric data, that is, “identification”. There are similar systems like walkthrough iris identification “Iris on the Move” [189] and Morphotrak [190] which mainly focus on security and identification using different types of biometrics. One of the first concepts of a fully functional PASS, to support the user by providing early detection of certain physiological, and psycho-emotional patterns, based on visual, infrared (IR) and other data collected during the pre-screening, has been proposed in the Biometric Technologies Laboratory, at the University of Calgary, in 2006 [191] [192]. The concept emphasized on semantic representation of information, derived using advanced decision-making approaches. In 2007, it was reported that Homeland Security Advanced Research Agency, and Science and Technology Human Factors Behavior Science Division of Department of Homeland Security (DHS), USA, have been developing a system called Future Access Screening Technology (FAST) [193]. The system measures pulse rate, skin temperature, breathing, facial expression, body movement, pupil dilation and other physiological and behavioral patterns. The technologies involved: high-resolution video for looking at facial expressions and body movements, a remote eye tracker, thermal cameras that provide information on skin temperature of the face, a remote cardiovascular and respiratory sensor, to

measure heart rate and respiration, and an audio system, for analyzing changes in voice pitch. In 2008, preliminary testing of this system had been reported as 78% accurate, on malintent detection and 80% on deception. The developers also address privacy protection issues in the system [194].

A situational awareness system, shown in Figure 7-1 uses multiple sensors and various software components to analyze biometric data and provide dialogue-based decision-making support to aid security personnel. Sensors measure a subject's temperature and blood flow paths, blood pressure, and pulse, from which the system can detect nervousness or estimate the presence and level of drug or alcohol intoxication. Other sensors include a microphone to capture voice, and a depth sensor, which the system uses in combination with video cameras to track body movements and capture gestures and facial expressions.

The system extracts features from biometric signals and uses this data to compare the subject against a database of prescreened individuals as well as to evaluate the subject's physiological and psycho-emotional states. It then notifies the system operator if the results indicate a match with a known person or signify potentially suspicious behavior.

In our approach of designing a PASS, we focus on the situational awareness paradigm, and artificial intelligence (AI) modules or assistants [192] [17]. More specifically, our primary goal is to set up a way of extracting biometric information - helpful in the context of situational awareness – for discrimination of person physical, and psycho-emotional states, as well as for the detection of transitions between those states. This concept departs from biometric-based security system, where main task is the authentication of customers. Most such systems utilize the visual appearance of customers, to compare against “lookout checklists” or suspected activity, and do not make effective use of other biometrics (e.g., skin temperature, voice, non-verbal behavior

characteristics) that can be acquired before (screening), or during the access authorization (interviewing) of individuals. On the other hand, the known approaches to the design of biometric-based physical access systems [188] utilize biometric devices as separate entities, and, sometimes, multimodal biometrics [195] with some degree of information fusion at the decision-making level, for human authentication.

However, using biometric devices solely for authentication do not exploit the full potential of acquired data, specifically, the data that can be collected during the interview, at access point. Hence, a significant improvement can be achieved by employing diverse biometric devices integrated in a framework with multiple artificial intelligent “biometric assistants” [191] [195] [17].

The structure of the proposed PASS involves the artificial intelligent assistants, such as the facial biometric assistant, infrared, voice, gait, and dialogue support and others (Figure 7-1). In this thesis, we focus on the facial biometric assistant, as a part of PASS. This assistant provides information on recognized facial expressions, recognized faces, and additional information derived from IR data. We also demonstrate how this information can be integrated for the decision-making support, via providing examples of semantic data, generated from facial biometrics. The semantic data is required for dialogue support in PASS.

The dialogue support assistant follows the concept of “interview support” systems. For example, if the related assistants detect elevated body temperature, or other features of particular interest, the dialogue support assistant automatically generates questions, and suggests them to the security personnel; for example: ‘Are you experiencing fever?’ or ‘Warning: Possible intention to change appearance; features of disguise are detected.’

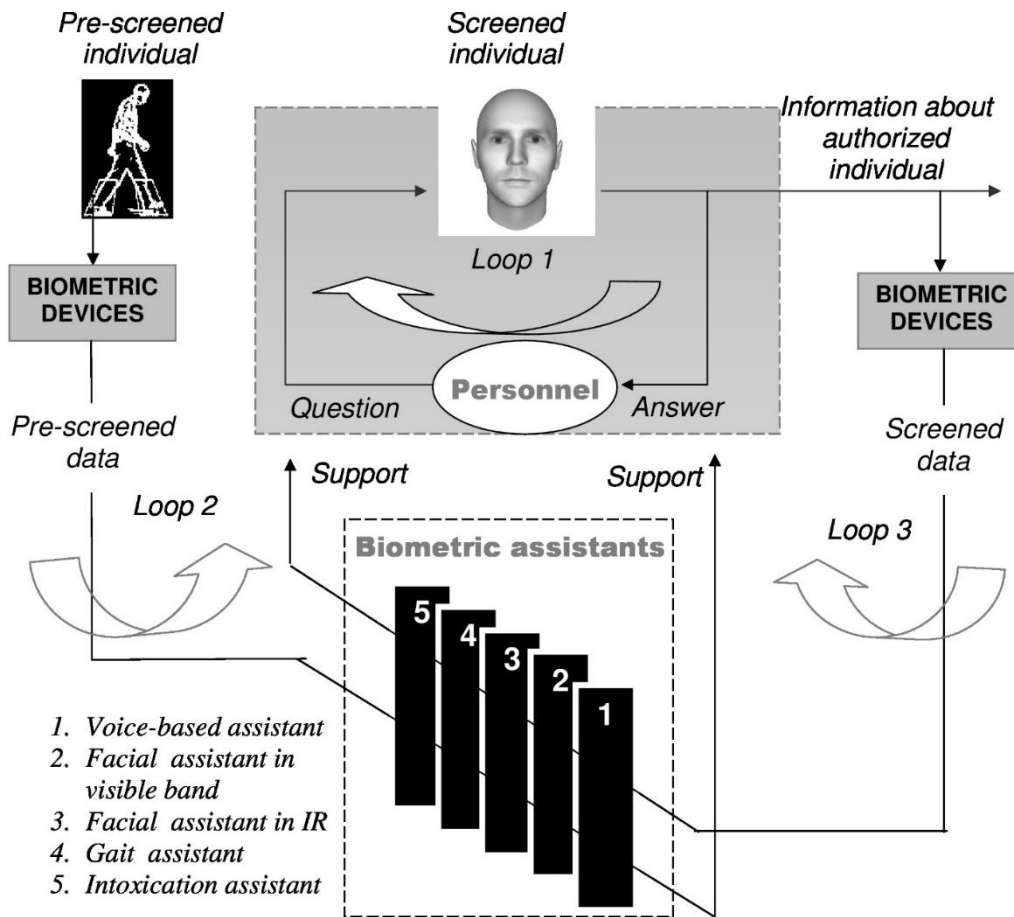


Figure 7-1. A proposed situational awareness system uses multiple sensors and various software components to analyze biometric data and provide dialogue-based decision-making support to aid security personnel.

It generates a response in “semantic” form, thus assisting the personnel in the decision-making, regarding authorization. Another aspect is that recommendations, generated by PASS, should be understood; therefore security personnel must be trained to use the full range of system functions. The training system must be designed to provide a real operating environment to the trainees. The concept of training was first proposed in [192] [195].

7.1.1 Facial biometrics for situational awareness

A concept of a facial biometric component, and its relationship to decision-making components, such as the dialogue support assistant, is presented in Figure 7-2. This component includes the

following assistants: IR; visual/depth; visual-behavioral and decision-making assistant. Biometrics represented by feature vectors are considered in a multimodal biometric in the form of a concatenation into a single feature vector. Feature-level fusion is performed using matching score and corresponding rules. However, the problem becomes more complicated if discriminative features must be detected and assigned to an emotional state, psychological condition or physiological properties. One of the approaches is to transform biometric features into a semantic form (knowledge domain) and integrate the relationship among various biometrics into a decision-level fusion. In the proposed design of biometric assistants, we utilize a Bayesian (probabilistic) interpretation of uncertainty that provides an acceptable reliability for decision-making; for example, in situations where conditional probabilities can be evaluated by experiment (preliminary knowledge about uncertainty). The Bayesian approach is one of the possible interpretations of uncertainty; combination with other decision profiles and approaches such as fuzzy estimation, entropy measures and Dempster–Shafer evidence methods can be applied as well. Biometric data in semantic form supports an expert (operator) in dialogue with a customer.

7.1.2 Decision making support

Knowledge domain coordination is performed, using results from fusion. The decision-making is based on the Bayesian model of belief.

In Bayesian belief estimations, input data are the classification results, as well as quantitative data (such as temperature measurements) in the hyper-spectral band. This structure is specific to each biometric assistant. For example, the IR face assistant uses temperature and blood pressure data to detect the alcohol and drug intoxication. These data are then passed to the dialogue support assistant. Consider a Bayesian network shown in Figure 7-3. The following states are

used to design the Bayesian network: five facial expressions (happy, sad, surprised, fear, disgust), three levels of temperature (low, medium, high), exposure to the virus (yes, no), and receiving medical attention (yes, no).

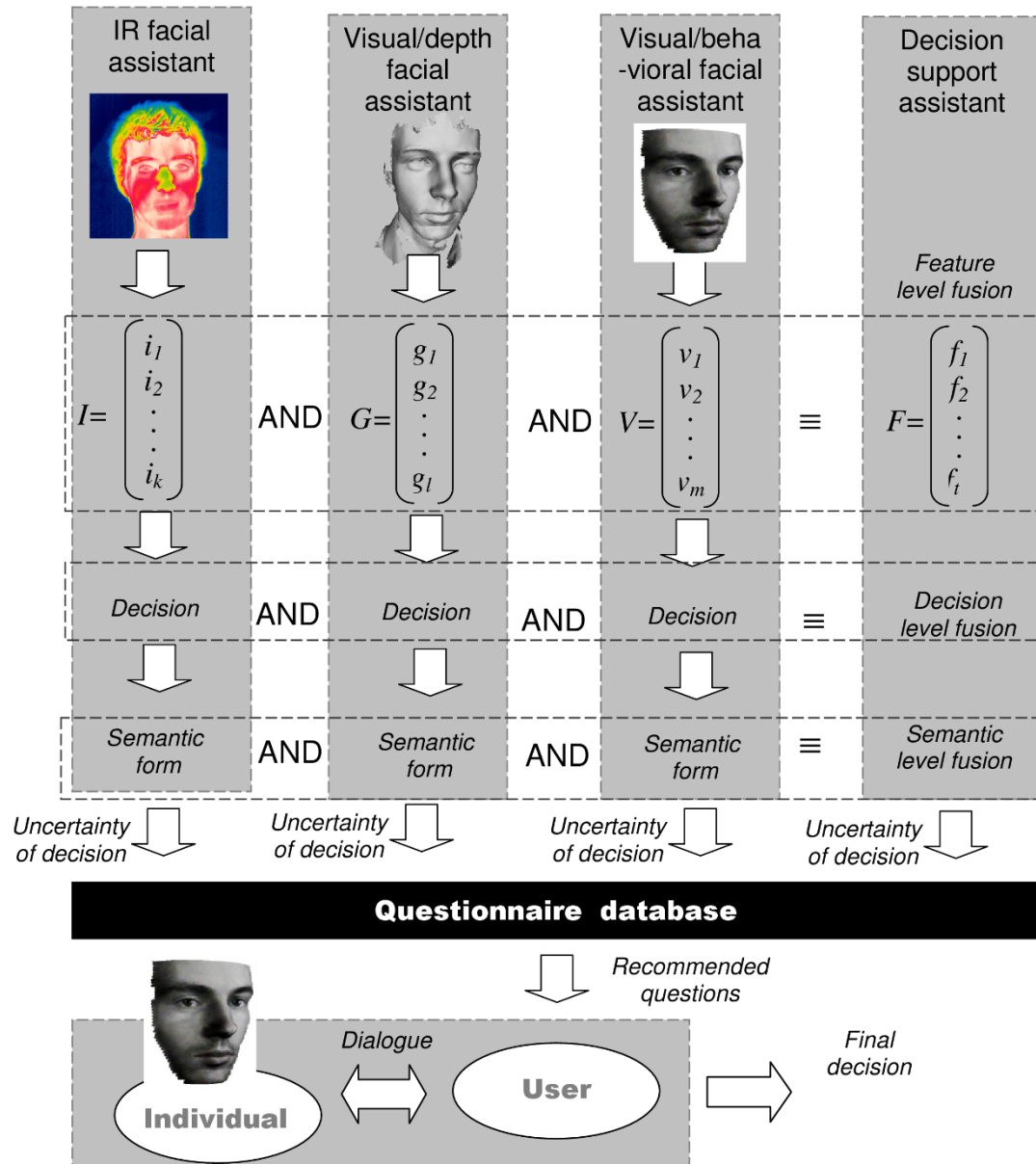


Figure 7-2. Structure of the facial biometric assistant.

For example, given an evidence that the high fever was detected (using the IR biometric assistant), and the customer has received medical attention, the probabilities of various facial expressions (disgust, happy, sad, surprise and fear) are evaluated as follows:

$$P(E|T = H, MA = Yes) = [0.1219, 0.1586, 0.3250, 0.3211, 0.0734]$$

Another example addresses finding joint probability on a set of nodes. For example, the joint probability of the event when expression “Sad” is recorded, temperature is medium and medical attention will be received is $P(E, T = M, MA = Yes) = 0.0470$; the joint probability of the event such that the facial expression “Fear” is registered, temperature is medium and medical attention will not receive is $P(E = F, T = M, MA = No) = 0.0280$.

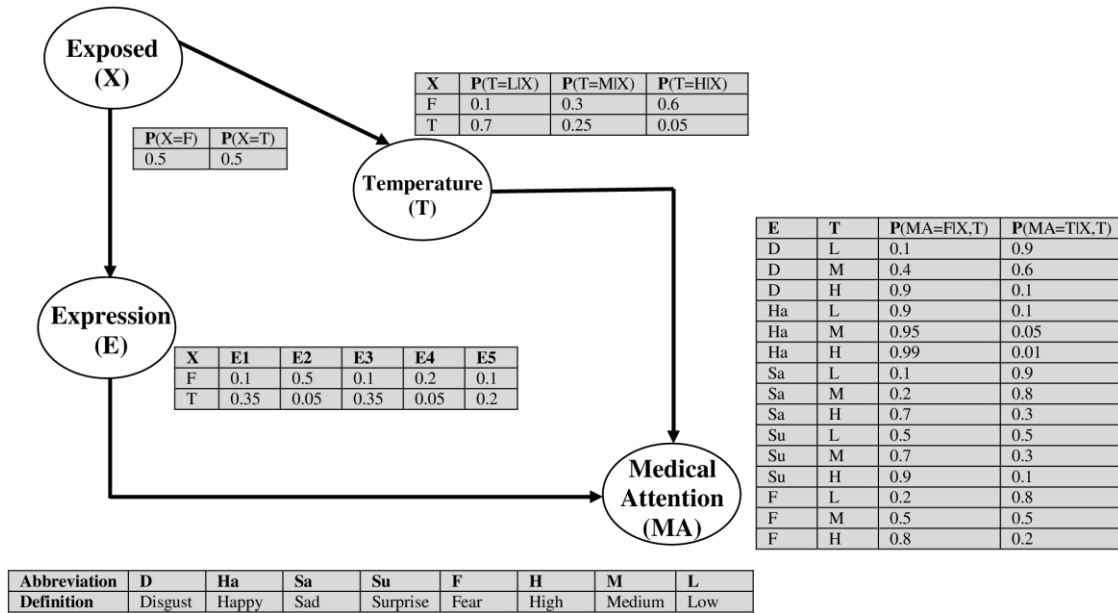


Figure 7-3. Bayesian network implemented in the experiment.

7.1.3 Dialogue support

This artificial intelligent assistant aimed to assist human-machine interaction in protocol form (recommendations) using a linguistic format. These protocols generate data such as level of

warning or alarm, the reason why this data was recognized as a warning or alarm, and recommendations to security personnel on possible actions. Security personnel use this information for further interviewing, as well as for making the final decision. As a rule, most of emotional responses of customers during a dialogue are hard to register by the user. This fact significantly contributes in uncertainty of human decision-making. Our goal is to minimize this uncertainty by providing support for the user. In particular, some hidden behavioral responses can be detected using an IR camera. Acquired IR data, along with video recording, and sometimes audio recording, can reveal some physiological and behavioral features, similar to the ones, estimated by the lie detector. However, in contrast to the latter, surveillance-based techniques are non-invasive, which are dictated by application-specific environment.

Let us consider an example of a scenario, in which a system generates data about a screened person.

It follows from this protocol (Figure 7-4), that this system evaluates a third level of warning, using automatically measured temperature, for the screened customer. The AI assistant evaluates any risks, and generates two possible solutions. The user can, in addition to the automated analysis, evaluate the images acquired in the visible and IR spectra. We distinguish scenarios that correspond to the results of the matching of the customer's data with data in local and global databases.

The following example introduces a scenario based on analysis of behavioral biometric data. The results of such analysis are presented to the user. Let us assume that there are three classes of samples assigned to "Disability", "Alcohol intoxication" and "Normal". The following linguistic constructions can be generated by such a system: 'Not enough data, but abnormality is detected' or 'Possible alcohol intoxication' or 'An individual with a disability'.

In the proposed system, human-human interaction is supported by the dialogue support assistant module. Knowledge, generated by this assistant, includes updated biometric parameters, and additional information, such as voice stress features and gait pattern.

Protocol for person #36 screening

Time: 12.30.07:

Warning: level 04

Specification: Drug or alcohol intoxication,
Level 03

Possible action:

1. Inquire database
2. Clarify in the dialogue

Protocol for person #36 interviewed

Time: 12.40.20:

Warning: level 04

Specification: Drug or alcohol intoxication,
Level 03

Local database matching: positive

Possible action:

1. Continue dialogue
2. Direct to an inspection

Figure 7-4. Sample protocols for a person.

We distinguished two types of uncertainty about the customer: the uncertainty that can be minimized by using customer responses, documents, and information from databases; and the uncertainty of appearance (physiological and behavioral) information, such as specific features in an IR facial image, gait, and voice. The last type of uncertainty can be minimized by specifically

oriented questionnaire techniques. These techniques have been used in criminology, in particular, for interviewing and interrogation. Deception can be defined as a semantic attack that is directed against the assistant. Technologies for preventing, detecting and prosecuting semantic attacks are still in their infancy. Some techniques of forensic interviewing and interrogation formalism with elements of detecting this semantic attack are useful in dialogue development [196]. In particular, the assistant's task is to alert of inconsistencies in the customer's replies. A deceptive person generally finds that more and more lies are necessary, as additional details are required, and the person either forgets what he has previously asserted or fabricates details that are not compatible with previous statements.

In assistant prototyping, we use Bayesian belief networks [191]. This is supported by the fact that the Bayesian (probabilistic) interpretation of uncertainty provides an acceptable reliability for decision-making in adaptable systems, that is, the systems that can learn in the process and adapt themselves as more information becomes available. In our experimental setup, we used:

1. Two JAI CV-M9 CL $3 \times 1/3''$ progressive scan RGB color cameras with 1034×779 4.65 μm effective square pixels for each CCD. These cameras can acquire full resolution images at a rate of 30 frames per second and output 24 bit RGB images via a Camera Link base configuration. The cameras are equipped with 16 mm lenses that allow them to capture face images from a distance of about 2-3 m.
2. A Thermoteknix MIRICLE 307 K uncooled micro-bolometer infrared camera, with a focal plane array of 640×480 pixel size and a dynamic range of 14 bits. The spectral band of the camera is 7-14 μm and the standard frame rate is 25-30 frames per second.

The camera is equipped with a 50 mm lens that allows it to capture faces from the same distance (2-3 m).

Protocol of the person #36 under screening

Time 00.00.00:

Warning, level 04

Specification: Drug or alcohol consumption,

Level 03

Local database matching: positive

Proposed dialogue questions:

Question 1: Are you in need of medical assistance?

Question 2: Have you had any problems during the flight?

Question 3: Do you plan to rent a car?

Question 4: Does anyone is going to meet/pick you up?

Question 5: Have you consumed wine or liquor aboard?

Question 6: Do you have anything to declare?

Protocol of the person #36 under screening (continuation)

Level of trustworthiness of Question 1 is 02:

Level of trustworthiness of Question 2 is 02:

Level of trustworthiness of Question 3 is 03:

Level of trustworthiness of Question 4 is 01:

Level of trustworthiness of Question 5 is 03:

Level of trustworthiness of Question 6 is 03:

Possible action:

1. Direct to special inspection
2. Further inquiry using dialogue

Figure 7-5. Sample scenario based on analysis of behavioral biometric data.

3. A PC station with acquisition boards (Euresys GRABLINK Expert 2 for the video cameras and Picolo Pro 2 for the thermal camera). For flicker-free illumination, two continuous light sources are used.

In addition, preliminary experiments have been conducted using the Microsoft Kinect sensor for tracking and face detection, as well as using depth information for multi-person scenes.

7.2 Face biometrics in human-machine interfaces

Facial biometric is a source of information, important for social communication. Humans use facial expressions to communicate information about emotional state but sometimes it is difficult to “read” it from the face. Although cultures may differ in social rules, certain facial expressions of emotion are universally recognized. In general, the information is not limited to a single region in face and/or exact mimics and also the degree of expressiveness varies between individuals. For unfamiliar faces, information such as age, gender and ethnicity can be extracted more or less reliable. Moreover, this information can also be used to remember and/or match unfamiliar faces, for example, by means of an attribute [197] [198] [199]. It is debatable that in order to achieve the accurate effective human-computer interaction (HCI), a computer is needed such that it can communicate with human like human-human interaction takes place. In applications such as computer-aided tutoring/learning, it is required that the computer’s response considers the emotional/cognitive state of the human user [200].

The authors of [201] study the application of recognition of facial expressions, such as anger, on recognizing emotion of a vehicle driver. They argue that it is important to recognize emotions, as the drivers affected by anger tend to drive faster and more aggressive. Their adaptive advanced driver assistance system detects and classify the participants’ emotions according to the FACS.

Having an interface with the capability of recognizing facial expression is an important aspect of social robots [202]. The robots that can display emotions (in the form of on-screen avatars, or robot-head) is an inverse problem to emotion recognition via analysis of facial expression.

Some research groups propose multi-modular systems that combine visual emotion detection (using facial expression) with other media channels, such as voice [203]. In [204], a general interaction system of a social robot is reported, based on emotion recognition from both facial expressions and voice. For the facial expression recognition module, they use a library for high-speed object recognition engine (SHORE) that detects faces, tracks the position and orientation, and detects emotion [205] [206]. The related area of application in human-machine interaction is gaming. In [207], emotion detection was used to adapt the dialogue of a computer game to the user state. The most recent version of the Microsoft Kinect was advertised to be able to “detect” facial expression to improve the adaptation of the game to the player [208].

7.3 Potential biomedical applications of face biometrics

Different neurological disorders can impair the ability to recognize facial emotions which are an important aspect of interpersonal communication. Some of these well-known disorders are:

1. Aspergers syndrome: Impairment in the use of multiple nonverbal behaviors such as eye-to-eye gaze, facial expressions, body postures, and gestures to regulate social interactions, a lack of spontaneous seeking to share enjoyment [209].
2. Autism: Impairment in the use of multiple nonverbal behaviors, such as eye-to-eye gaze, facial expressions [210].
3. Schizoid Personality: Usually in the Schizoid Personality Disorder is displayed a “bland” exterior without visible emotional reactivity and rarely reciprocate gestures or facial

expressions, such as smiles or nods. They rarely experience strong emotions such as anger and joy [211].

4. Bipolar Disorder: The subjects affected of Bipolar Disorder suffer from specific deficits of facial emotion perception. The patients present impaired recognition of disgust, fear and sadness [212].
5. Alzheimer's disease: In patients with Alzheimer disease may be a deficit in processing some or all facial expressions of emotions [213].
6. Prosopagnosia: The disorder of face perception where the ability to recognize faces is impaired, while the ability to recognize other objects may be relatively intact [214].

For example, emotions on faces of autistic patients were monitored in [215]. Among these disorders, the development of an application and/or tools for rehabilitation of a prosopagnosia and bipolar disorder could be a good application of facial biometrics.

The ability to recognize and remember familiar individuals by their faces enables directed communication over longer distances and enables social cooperation. The lack of this skill is the key deficit in prosopagnosia [214], the inability to recognize familiar faces. Prosopagnosia (also known as face-blindness) can be present from birth as a developmental and sometimes inherited disorder [216] or can be acquired later in life from a variety of cerebral lesions [217] in the fusiform, lingual, temporal or occipital cortex [218]. In both congenital and acquired clinical conditions, patients with prosopagnosia suffer from cognitive defects that may include perceptual or memory processes [219]. In apperceptive prosopagnosia, patients are unable to form an accurate perceptual representation of the structure of a viewed face; in associative prosopagnosia, formation of the facial percept is intact, but this information cannot be matched to facial memories in order to recognize a face that the individual has encountered before. The

dysfunction of either apperceptive or associative components leads to the same end result of failure to recognize a familiar face. To date, a rehabilitation treatment enabling patients with prosopagnosia is not available. Since prosopagnosia is suggested to affect 2% of the population [220], the development of a face recognition system tailored to patients with prosopagnosia is a priority in the field of clinical neuroscience.

Bipolar disorder is an example of facial expression recognition application, which is yet to be investigated. A further work for this thesis could be an implementation of both FR and FER on a wearable device (for example, eyeglasses equipped with camera and a headphone). The camera records live video from a subject or multiple subjects and the system can perform FR and FER at the same time. The results can be translated into voice (eg. name of the person or the emotion state) to help the patient to recognize other people (in prosopagnosia) or the emotion (in bipolar disorder). A similar project has been attempted at the University of Huston [221]. This paper proposed a system consisting of an eyewear, connected to a Smartphone that runs face recognition in real-time for prosopagnosic patients. The Smartphone displayed the tagged identity information using smartphone controlled eyewear.

7.4 Conclusion

In this chapter potential applications of the facial biometrics have been mentioned. These applications include biomedical, human robotic interaction, decision making and dialogue support. A numerical example of decision making support system using Bayesian network was presented. Implementation of facial expression recognition for biomedical application, such as wearable face detector and analyzer, needs to be implemented in the form of a wearable System-on-Chip (SoC). Another alternative is a wearable accessory to a SmartPhone, connected via Wi-Fi (other applications such as heart-beat monitor, have been recently announced by Samsung, in

particular). Each of those suggested applications can be a future development for the proposed in this thesis face and face expression recognition approach using wavelet-based transforms.

Chapter Eight: **CONCLUSIONS AND FUTURE WORKS**

This chapter provides conclusions pertaining to the facial biometrics, in particular facial expression recognition, facial points detection, and expression-invariant face recognition presented in the thesis. This is followed by recommendations for future works.

8.1 Conclusion

This thesis showcases a joint facial expression and face recognition in the framework of biometric-based applications. Three main issues that are relevant to the design of joint FER and FR systems were addressed. The first issue is how to select right facial features which are suitable for both FER and FR. The second is what type of features are efficient to perform both tasks at the same time and also can be applied for infrared images as well. The last one is how to use the facial biometrics for the interview support system. This thesis proposed a solution based on a unified wavelet-based transform approach (specifically, Gauss-Laguerre) and Modified vesselness Frangi filter for facial feature detection. This approach addresses the following phases of facial biometric data processing and recognition:

Facial point detection: A new filter based on Frangi's vesselness filter modified for facial feature extraction has been proposed in this thesis. The extracted facial features can be used later for facial recognition or expression recognition. The eyebrows, eyes, nostrils, and mouth were considered as facial features in this paper. The eyes were segmented in the coarse estimation step, and the estimated position of the mouth was also obtained. The fine detection was performed with the same filter, but using a different scale. The mouth, nose and eyebrows were precisely localized by using geometric information of facial features. The system has been tested on four different databases with different expressions and poses. The average detection rate of

97.92% for eye detection (100% for JAFFE, 98.08 for CK+, 97.37 for FEI, and 96.25 for ORL databases), and 97.78% for all other detected features have been achieved.

Feature extraction: We have investigated the proposed feature extraction for FER using GL wavelet. This filter is capable of providing efficient features that represent unique characteristics of facial images for expression recognition. Local and global feature extraction scheme have been performed on different databases in this thesis to come up with the best solution in terms of accuracy. We also studied different geometric features in addition to the textural features to enhance the feature extraction technique. The experimental results led us to conclusion that the highest performance, in terms of recognition accuracy, was achieved by using Frangi filter for facial points detection, along with the global textural features in addition to 15 geometric distances. To show the efficiency of the proposed approach, three different scenarios (leave-one-out, cross validation, and expresser-based) have been applied on three different databases (JAFFE, CK, and MMI). The average recognition rate of 94.29% was achieved for JAFFE, 89.55% for CK, and 84.60% for MMI database. It should be noted that the lower recognition rate on MMI and CK compared to JAFFE does not mean the weak performance; this is rather due to the fact that those databases are more complex in terms of image quality, occlusion, and the number of image per subject showing the exact expression.

Face recognition: This study has investigated the proposed method for face recognition using GL wavelet. The FR requires the same approaches as FER at all steps (preprocessing, feature extraction and classification). The slight difference between FR and FER lies in feature extraction. However, it is the GL filter that was used, albeit with different parameters, to provide textural features for both tasks. Compared to a few available methods that can handle both FR and FER at the same time, our system provides a low cost, one-step feature extraction technique

for both classifications. Extensive comparison with other approaches on two different databases showed the efficiency of the proposed system in various classification scenarios. The average recognition rate of 96.76% was achieved for CK, and 96.61% for JAFFE.

Infrared facial biometrics: We proposed a novel approach for facial expression recognition in infrared images. The aim of this study was to use the same approach that was previously applied on visible images for facial expression recognition, in order to demonstrate the benefits of illumination-invariance of thermal images, in the task of FER, with minimum modifications. The outcome of this method is decreasing the computational cost and increasing the speed of the system equipped with multi-band cameras for behavior analysis in real-time. The advantage of the proposed GL filter to extract features from infrared images of face is its rich frequency extraction capability for texture analysis, as well as being a rotation-invariant.

In addition, fusion of visible and infrared images for expression recognition has been studied. Most systems, that utilize fusion at the feature level, use partially and sometimes completely different algorithms to extract features from infrared and visible images. This increases the computational complexity, since it should perform two different feature extraction algorithms. In the proposed approach, the feature extraction for both infrared and visible image is performed, using a single GL filter. This is due to applying a uniform feature extraction technique for both spectra. Application of fusion results in outperforming the previously proposed approach, based on the same GL based feature extraction: the performance, in terms of FER rate, is better for infrared images by 2.7%, and by 2.3% for visible images.

Decision making: In this thesis potential applications of the facial biometrics have been mentioned. These applications include biomedical, human robotic interaction, decision making and dialogue support. A numerical example of decision making support system using Bayesian

network was presented. Implementation of facial expression recognition for biomedical application, such as wearable face detector and analyzer, needs to be implemented in the form of a wearable System-on-Chip (SoC). Another alternative is a wearable accessory to a SmartPhone, connected via Wi-Fi.

8.2 Future works

Although the proposed research proved high success rate in the joint facial expression recognition and face recognition for the still images, a number of issues remain unsolved and should be addressed in future work.

1. The system has been mainly adopted and tested for still images and not videos. Since the thesis aim was to propose a joint FER/FR for the interview-support systems, we stated off using frame images which is the basis for video analysis. Thus, one important future work could be investigating the real-time video processing to joint FER/FR systems.
2. The hardware implementation of the system on wearable devices (like eyeglass) was of interest from the beginning of this thesis but due to time limitation, it is still an ongoing task. We bought an eyeglasses equipped with camera for this research. There are few limitations associated with the camera which are: lack of storage for video-processing, not having Bluetooth connectivity to communicate with the processor (like PC or SmartPhone for processing), not being equipped with System-On-Chip (SoC) to perform all the processing without using any external portable processor. Working on hardware implementation could be an interesting future work.
3. The software has been implemented mainly in MatLab and partially in OpenCV and Java. For the real-time analysis, one needs to implement the software in language such as C++. Moreover, the protocol to transfer the captured image (video) from the wearable device to

the processor and vice versa should be taken into account. For example, SmartPhone programing is a good option.

4. In this thesis, we used a simple KNN classifier. This classifier has been used by several other papers and performed well. Although we have tested other classifiers like neural network, SVM, Bayesian network, but these kind of classifiers have not been tested comprehensively. An extension of this work could be using different classifiers or the fusion of classifiers for achieving a better recognition rate.
5. One interesting work could be using Kinect® for facial expression and face recognition from video. The Kinect camera is cheap and equipped with cameras which can provide also depth information. There are lots of available tools and libraries for face analysis using Kinect that makes the study easier.

REFERENCES

- [1] C.S. Lee and A. Elgammal, "Facial expression analysis using nonlinear decomposable generative model," in *IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, 2005.
- [2] H. Wang and N. Ahuja, "Facial expression decomposition," in *IEEE International Conference on Computer Vision*, Nice, France, 2003.
- [3] I. Mpiperis, S. Malassiotis, and M. Strintzis, "Bilinear models for 3D face and facial expression recognition," *IEEE Transactions on Information Forensics and Security*, vol. 3, no. 3, pp. 498-511, 2008.
- [4] A. Colmenarez, B. Frey, and T. S. Huang, "A probabilistic framework for embedded face and facial expression recognition," in *IEEE Conference on Computer Vision and Pattern Recognition*, Ft. Collins, USA, 1999.
- [5] X. Li, G. Mori, and H. Zhang, "Expression-invariant face recognition with expression classification," in *Canadian Conference on Computer*, Canada, 2006.
- [6] S. Taheri, V. Patel and R. Chellappa, "Component-based recognition of faces and facial expressions," *IEEE Transactions on Affective Computing*, vol. 4, no. 4, pp. 360-371, 2013.
- [7] J. Tenenbaum and W. Freeman, "Separating style and content with bilinear models," *Neural Computation*, vol. 6, no. 12, pp. 1247-1283, 2000.
- [8] A. Poursaberi, M. Spicher, H. Ahmadi, S.N. Yanushkevich, "Global Gauss-Laguerre wavelets feature selection for facial expression recognition," in *8th International Conference on Digital Technologies*, Slovak Republic, 2011.
- [9] A. Poursaberi, S.N. Yanushkevich, M.L. Gavrilova, "Modified multiscale vesselness filter for facial feature detection," in *IEEE 4th International Conference on Emerging Security Technologies*, UK, 2013.
- [10] A. Poursaberi, J. Vana, S. Mracek, R. Dvora, S.N. Yanushkevich, M. Drahansky, V. Shmerko, M.L. Gavrilova, "Facial biometrics for situational awareness systems," *IET Biometrics*, vol. 2, no. 2, pp. 35-47, 2013.
- [11] A. Poursaberi, H. Ahmadi, S.N. Yanushkevich, M.L. Gavrilova, "Gauss-Laguerre wavelet textural feature fusion with geometrical information for facial expression identification," *EURASIP Journal of Image and Video Processing*, vol. 17, no. 2012, 2012.
- [12] A. Poursaberi, S.N. Yanushkevich, M.L. Gavrilova, "An efficient facial expression recognition system in infrared images," in *IEEE 4th International Conference on Emerging Security Technologies*, UK,

2013.

- [13] J. Vana, S. Mráček, M. Dražanský, A. Poursaberi, S.N. Yanushkevich, "Applying fusion in thermal face recognition," in *IEEE International Conference of the Biometrics Special Interest Group*, Germany, 2012.
- [14] A. Poursaberi, M.L. Gavrilova, S.N. Yanushkevich, "Fusion of infrared and visible images for facial expression recognition," in *IEEE International Conference on Automatic Face and Gesture Recognition*, Ljubljana, Slovenia, 2015.
- [15] A. Poursaberi, S.N. Yanushkevich, M.L. Gavrilova, V.P. Shmerko, P.S.P. Wang, "Situational awareness through biometrics," *IEEE Computer*, vol. 46, no. 5, pp. 102-104, 2013.
- [16] K. Lai, A. Poursaberi, S.N. Yanushkevich, "One-shot facial feature extraction based on Gauss-Laguerre filter," in *CCECE Canadian Conference on electrical and Computer Engineering*, Toronto, Canada, 2014.
- [17] O. Boulanov, M.L. Gavrilova, A. Poursaberi, M. Spicher, V.P. Shmerko, S.N. Yanushkevich, "Biometric-based intelligent agent systems," in *International Conference on Intelligent Systems and Agents*, Rome, Italy, 2011.
- [18] M. Yuki, W.W. Maddux, T. Masuda, "Are the windows to the soul the same in the East and West? Cultural differences in using the eyes and mouth as cues to recognize emotions in Japan and the United States," *Experimental Social Psychology*, vol. 43, no. 2, pp. 303-311, 2007.
- [19] M. Suwa, N. Sugie, N., K. Fujimora, "A preliminary note on pattern recognition of human emotional expression," in *4th International Joint Conference on Pattern Recognition*, Kyoto, Japan, 1978.
- [20] S. Bashyal, G.K. Venayagamoorthy, "Recognition of facial expressions using Gabor wavelets and learning vector quantization," *Engineering Applications of Artificial Intelligence*, vol. 21, p. 1056–1064, 2008.
- [21] H.A. Effenbein, N. Ambady, "When familiarity breeds accuracy: cultural exposure and facial emotion recognition," *Journal of Personality and Social Psychology*, vol. 2, no. 85, pp. 276-290, 2003.
- [22] J. Cohn, K. Schmidt, R. Gross, and P. Ekman, "Individual differences in facial expression: stability over time, relation to self-reported emotion, and ability to inform person identification," in *Proceedings of the International Conference on Multimodal User Interfaces*, PA, Pittsburgh, 2002.
- [23] P. Ekman and W.V. Friesen, *Manual for the facial action coding system*. Palo Alto: Consulting Psychologists Press, 1978.
- [24] A. C. M. Mtg., "MPEG Video and SNHC. "Text of ISO/IEC FDIS 14 496-3: Audio,"," Doc. ISO/MPEG N2503, Oct. 1998.

- [25] B. Cerretani, "Modelling human perception of static expressions by discrete choice models," Master thesis, Università degli Studi di Siena, 2007.
- [26] P. Ekman and W.V. Friesen, "Constants across cultures in the face and emotions," *Journal of Personality Social Psychology*, vol. 2, no. 17, pp. 124-129, 1971.
- [27] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "Coding facial expressions with gabor wavelets," in *Third IEEE International Conference on Automatic Face and Gesture Recognition*, Nara, Japan, 1998.
- [28] M.S. Bartlett, G. Littlewort, I. Fasel, and J.R. Movellan, "Real time face detection and facial expression recognition: development and applications to human computer interaction," in *IEEE Conference on Computer Vision and Pattern Recognition*, Madison, Wisconsin, 2003.
- [29] M. Lyon, S. Akamatsu, "Coding facial expressions with gabor wavelets," in *3rd International Conference on Face and Gesture Recognition*, Washington, DC, USA, 200-205.
- [30] Z. Zhang, M. Lyons, M. Schuster, and S. Akamatsu, "Comparison between geometry-based and gabor-wavelets-based facial expression recognition using multi-layer perceptron," in *3rd International Conference on Face and Gesture Recognition*, Nara, Japan, 1998.
- [31] C.C. Lee, C.Y. Shih, "Gabor feature selection for facial expression recognition," in *International Conference on Signals and Electronic Systems*, Gliwice, Poland, 2010.
- [32] G. Cottrell, J. Metcalfe., "Face, gender and emotion recognition using holons.," in *International Conference on Advances in Neural Information Processing Systems*, San Mateo, 1991.
- [33] X. Chen, T. Huang, "Facial expression recognition: A clustering based approach," *Pattern Recognition Letters*, vol. 24, pp. 1295-1302, 2003.
- [34] M. Turk, A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, pp. 71-86, 1991.
- [35] F.Y. Shin, C.F. Chuang, and P.S.P. Wang, "Performance comparisons of facial expression recognition in JAFFE database," *IEEE Transactions on Pattern Recognition and Artificial Intelligence*, vol. 3, no. 22, pp. 445-459, 2008.
- [36] A. Rahardja, A. Sowmya, and W. Wilson, "A neural network approach to component versus holistic recognition of facial expressions in images," in *Proceedings of SPIE Symposium on Intelligent Robots and Computer Vision X: Algorithms and Techniques*, Boston MA, 1991.
- [37] X. Feng, M. Pietikäinen, and A. Hadid, "Facial expression recognition based on local binary patterns," *Journal of Pattern Recognition and Image Analysis*, vol. 4, no. 17, pp. 592-598, 2007.
- [38] A. Lanitis, C. Taylor, and T. Cootes, "Automatic interpretation and coding of face images using flexible models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 7, no. 19, pp.

743-756, 1997.

- [39] L.Y. Yacoob, H. Lam, and L. Davis, "Recognizing faces showing expressions," in *International Workshop on Automatic Face and Gesture Recognition*, Zurich, 1995.
- [40] Y. Yacoob, L. Davis, "Recognizing facial expressions by spatio-temporal analysis," in *Proceedings of the International Conference on Pattern Recognition*, Jerusalem, Israel, 1994.
- [41] T. Xiang, M.K.H. Leung, and S.Y. Cho, "Expression recognition using fuzzy spatio-temporal modeling," *Pattern Recognition*, vol. 1, no. 41, pp. 204-216, 2008.
- [42] M. Bartlett, P. Viola, T. Sejnowski, L. Larsen, J. Hager, and P. Ekman, "Classifying facial action," *Advances in Neural Information Processing Systems 8*, pp. 823-829, 1996.
- [43] I. Essa, A. Pentland, "Coding, analysis, interpretation and recognition of facial expressions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 7, no. 19, pp. 757-763, 1997.
- [44] G. Donato, M.S. Bartlett, C. Hager, P. Ekman, and J. Sejnowski, "Classifying facial actions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 10, no. 21, pp. 974-989, 1999.
- [45] W. Fellenz, J. Taylor, N. Tsapatsoulis, and S. Kollias, "Comparing template-based, feature-based and supervised classification of facial expressions from static images," in *Proceedings of Circuits, Systems, Communications and Computers*, Japan, 1999.
- [46] I.R. Fasel, M.S. Bartlett, and J.R.A. Movellan, "A comparison of gabor filter methods for automatic Detection of facial landmarks," in *IEEE 5th International Conference on Automatic Face and Gesture Recognition*, Washington, DC, 2002.
- [47] B. Fasel, and J. Luettenb, "Automatic facial expression analysis: A survey," *Pattern Recognition*, vol. 1, no. 36, pp. 259-275, 2003.
- [48] S. Zafeiriou and I. Pitas, "Discriminant graph structures for facial expression recognition," *IEEE Transactions on Multimedia*, vol. 10, no. 8, pp. 1528-1540, 2008.
- [49] H. Deng, J. Zhu, M.R. Lyu, and I. King, "Two-stage multi-class AdaBoost for facial expression recognition," in *International Joint Conference on Neural Networks*, Orlando, FL, 2007.
- [50] M. Pardas and A. Bonafonte, "Facial animation parameters extraction and expression recognition using Hidden Markov Models," *Signal Processing: Image Communication*, vol. 17, pp. 675-688, 2002.
- [51] P. Lucey, J. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended cohn-kanade dataset (CK+): a complete dataset for action unit and emotion-specified expression," in *Computer Vision and Patter Recognition*, San Francisco, 2010.

- [52] K. Meissner, E. Muth, and B.M. Herbert, "Bradygastric activity of the stomach predicts disgust sensitivity and perceived disgust intensity," *Biological Psychology*, vol. 86, pp. 9-16, 2011.
- [53] ". definition", "<http://www.medterms.com/script/main/art.asp?articlekey=33843>," Retrieved 2008-04-05. [Online].
- [54] E. Huber, *Evolution of facial musculature and facial expression*, The Johns Hopkins press; London, H. Milford, Oxford University Press, 1931.
- [55] F. -. F. A. C. System, "<http://www.cs.cmu.edu/~face/facs.htm>," [Online].
- [56] Y. Tian, T. Kanade, J. Cohn, "Recognizing action units for facial expression analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, pp. 97-115, 2001.
- [57] C. Zor, "Facial expression recognition," Master thesis, University of Surrey, Guildford, Surrey, 2008.
- [58] L. Zhang, "Towards spontaneous facial expression recognition in realworld video," PhD thesis, 2012.
- [59] J. Ostermann, "Pandzic, Igor; Forchheimer, Robert. MPEG-4 facial animation: The standard, implementation and applications," in *Chapter 2: Face Animation in MPEG-4*, Wiley, August 2002, pp. 17-55.
- [60] P.S. Aleksic and A.K. Katsaggelos, "Automatic facial expression recognition using facial animation parameters and multistream Hmms," *IEEE Transactions on Information Forensics and Security*, vol. 1, pp. 3-11, 2006.
- [61] T. Hao and T.S. Huang, "3D facial expression recognition based on automatically selected features," in *Computer Vision and Pattern Recognition Workshops*, Anchorage, Alaska, USA, 2008.
- [62] Y. Tong, W. Liao, and Q. Ji, "Facial action unit recognition by exploiting their dynamic and semantic relationships," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 10, pp. 1683-1699, 2007.
- [63] A. Frangi, W. Niessen, K. Vincken, and M. Viergever, "Multiscale vessel enhancement filtering," in *Conference on Medical Image Computing and Computer Assisted Intervention*, Cambridge, UK, 1998.
- [64] P. Wang, M. Green, Q. Ji, and J. Wayman, "Automatic eye detection and its validation," in *IEEE Workshop on Computer Vision and Pattern Recognition*, San Diego, CA, 2005.
- [65] P. Wang, and Q. Ji, "Learning discriminant features for multi-view face and eye detection," in *IEEE Conference on Computer Vision and Pattern Recognition*, San Diego, CA, USA, 2005.
- [66] R.S. Feris, J. Gemmell, K. Toyama, and V. Krueger, "Hierarchical wavelet networks for facial feature localization," in *5th International Conference on Automatic Face and Gesture Recognition*,

Washington D.C. , USA, 2002.

- [67] E. Holden, R.A. Owens, "Automatic facial point detection," in *5th Asian Conference on Computer Vision (ACCV2002)*, Melbourne, 2002.
- [68] M.J.T. Reinders, R.W.C. Koch, and J.J. Gerbrands, "Locating facial features in image sequences using neural networks," in *2nd International Conference on Automatic Face and Gesture Recognition*, Killington, Vermont, 1996.
- [69] D. Vukadinovic and M. Pantic, "Fully automatic facial feature point detection using gabor feature based boosted classifiers," in *IEEE International Conference on Systems, Man and Cybernetics*, Waikoloa, Hawaii, 2005.
- [70] M.F. Valstar, B. Martinez, X. Binefa, and M. Pantic, "Facial point detection using boosted regression and graph models," in *IEEE Conference on Computer Vision and Pattern Recognition*, San Francisco, USA, 2010.
- [71] I.R. Fasel, B. Fortenberry, and J.R. Movellan, "GBoost: A generative framework for boosting with applications to realtime eye coding," *Journal of Computer Vision and Image Understanding*, p. 182–210, 2005.
- [72] T.F. Cootes, C.J. Taylor, D.H. Cooper, and J. Graham, "Active shape models - their training and application," *Journal of Computer Vision and Image Understanding*, vol. 61, pp. 38-59, 1995.
- [73] 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, 2001.
- [74] T. Coots, "An introduction to active shape models," in *Image Processing and Analysis, Chapter 7*, Oxford University Press, 2000, pp. 223-248.
- [75] S. Milborrow and F. Nicolls, "Locating facial features with an extended active shape model," in *10th European Conference on Computer Vision: Part IV*, Marseille, France, 2008.
- [76] L. Dang and F. Kong, "Facial feature point extraction using a new improved active shape model," in *3rd International Congress on Image and Signal Processing*, Yantai, China, 2010.
- [77] Y. Zhou, Y. Li, Z. Wu, and M. Ge, "Robust facial feature points extraction in color images," *Engineering Applications of Artificial Intelligence*, vol. 24, pp. 195-200, 2011.
- [78] D. Cristinacce, T. Cootes, "Facial feature detection using adaboost with shape constrains," in *British Machine Vision Conference*, 2003.
- [79] "3D shape constraint for facial feature localization using probabilistic-like output," in *6th IEEE International Conference on Automatic Face and Gesture Recognition*, Seoul, Korea, 2004.

- [80] M. Valstar, "Timing is everything, A spatio-temporal approach to the analysis of facial actions," PhD thesis, 2008.
- [81] M.H. Yang, K.J. Kriegman, and N. Ahuja, "Detecting faces in images: a survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 34-58, 2002.
- [82] P. Viola and M. Jones, "Robust real-time object detection," *Computer Vision*, vol. 2, no. 57, pp. 137-154, 2004.
- [83] F. Chen, "Facial feature point detection," Master thesis, 2011.
- [84] C. Kotropoulos and I. Pitas, "Rule-based face detection in frontal views," in *International Conference on Acoustics, Speech and Signal Processing*, Munich, Bavaria, Germany, 1997.
- [85] G. Yang and T.S. Huang, "Human face detection in complex background," *Pattern Recognition*, vol. 1, no. 27, pp. 53-63, 1994.
- [86] C.C. Han, H.M. Liao, G.j. Yu, and L.H. Chen, "Fast face detection via morphology-based pre-processing," *Pattern Recognition*, vol. 33, pp. 1701-1712, 2000.
- [87] T.K. Leung, M.C. Burl, and P. Perona, "Finding faces in cluttered scenes using random labeled graph matching," in *5th International Conference on Computer Vision*, Massachusetts, USA, 1995.
- [88] E. Saber and A.M. Tekalp, "Frontal-view face detection and facial feature extraction using color, shape and symmetry based cost functions," *Pattern Recognition Letters*, vol. 17, pp. 669-680, 1998.
- [89] K.C. Yow and R. Cipolla, "Feature-based human face detection," *Image and Vision Computing*, vol. 9, no. 15, pp. 713-735, 1997.
- [90] I. Craw, D. Tock, and A. Bennett, "Finding face features," in *2nd European Conference on Computer Vision*, Ligure, Italy, 1992.
- [91] V. Govindaraju, "Locating human faces in photographs," *Computer Vision*, vol. 19, pp. 129-146, 1996.
- [92] J. Miao, B. Yin, K. Wang, L. Shen, and X. Chen, "A hierarchical multiscale and multiangle system for human face detection in a complex background using gravity-center template," *Pattern Recognition*, vol. 32, pp. 1237-1248, 1999.
- [93] A.L. Yuille, P.W. Hallinan, and D.S. Cohen, "Feature extraction from faces using deformable templates," *Computer Vision*, vol. 8, pp. 99-111, 1992.
- [94] C. Liu, "A bayesian discriminating features method for face detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, p. 7250740, 2003.

- [95] P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, pp. 711-720, 1997.
- [96] H.A. Rowley, S. Baluja, and T. Kanade, "Neural network-based face detection," *IEEE Transactions On Pattern Analysis and Machine intelligence*, vol. 20, pp. 23-38, 1998.
- [97] C. P. Papageorgiou, M. Oren, and T. Poggio, "A general framework for object detection," in *6th International Conference on Computer Vision*, Bombay, India, 1998.
- [98] P. Li, "Adaptive feature extraction and selection for robust facial expression recognition," Master thesis, 2010.
- [99] F.H.C. Tivive, A. Bouzerdoun, S.L.P. Phung, and K.M. Liftekharuddin, "Adaptive hierarchical architecture for visual recognition," *Applied Optics*, vol. 49, pp. B1-B8, 2010.
- [100] "<http://www.mathworks.com/discovery/object-detection.html>," [Online].
- [101] C. Lorenz, I.C. Carlsen, T.M. Buzug, C. Fassnacht, J. Weese, "Multi-scale line segmentation with automatic estimation of width, contrast and tangential direction in 2D and 3D medical images," in *Joint Conference on Computer Vision, Virtual Reality and Robotics in Medicine and Medical Robotics and Computer-Assisted Surgery*, 1997.
- [102] H. Mirzaalian, and G. Hamarneh, "Vessel scale-selection using MRF optimization," in *IEEE Conference on Computer Vision and Pattern Recognition*, San Francisco, USA, 2010.
- [103] S.M. Lajevardi, M. Lech, "Facial expression recognition using neural networks and log-gabor filters," in *Proceedings of Digital Image Computing: Techniques and Applications*, Australia, 2008.
- [104] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "Coding facial expressions with Gabor wavelets," in *3rd IEEE International Conference on Automatic Face and Gesture Recognition*, 1998.
- [105] "<http://www.fei.edu.br/~cet/facedatabase.html>," [Online].
- [106] F. Samaria and A. Harter, "Parameterization of a stochastic model for human face identification," *Applications of Computer Vision*, pp. 138-142, 1994.
- [107] M. Hassaballah, T. Kanazawa, and S. Ido, "Efficient eye detection method based on grey intensity variance and independent components analysis," *Computer Vision*, pp. 261-271, 2010.
- [108] Y. Ma, X. Ding, Z. Wang, and N. Wang, "Robust precise eye location under probabilistic framework," in *IEEE International Conference on Automatic Face and Gesture Recognition*, Seoul, Korea, 2004.
- [109] S. Kim, S. T. Chung, S. Jung, D. Oh, J. Kim, and S. Cho, "Multi-scale gabor feature based eye

- localization," *World Academy of Science, Engineering and Technology*, vol. 21, pp. 483-487, 2007.
- [110] B. de Brito Leite, E.T. Pereira, H.M. Gomes, L.R. Veloso, C.E do Nascimento Santos, and J.M. de Carvalho, "A Learning-based eye detector coupled with eye candidate filtering and PCA features," in *XX Brazilian Symposium on Computer Graphics and Image Processing*, Minas Gerais, 2007.
- [111] W. Li, Y. Wang, and Y. Wang, "Eye location via a novel integral projection function and radial symmetry transform," *Journal of Digital Content Technology and its Applications*, vol. 5, no. 8, pp. 70-80, 2011.
- [112] H.J. Kim and W.Y. Kim, "Eye detection in facial images using zernike moments with SVM," *ETRI Journal*, vol. 2, no. 30, pp. 335-337, 2008.
- [113] J. Whitehill, "Automatic real-time facial expression recognition for signed language translation," Master thesis, 2006.
- [114] K. Mase, "Recognition of facial expression from optical flow," *Institute of Electronics, Information, and Communication Engineers Transactions*, vol. 10, no. 74, pp. 3474-3483, 1991.
- [115] Y. Yacoob and L.S. Davis, "Recognizing human facial expressions from long image sequences using optical flow," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, no. 18, pp. 636-642, 1996.
- [116] M.J. Black and Y. Yacoob, "Recognizing facial expressions in image sequences using local parameterized models of image motion," *Computer Vision*, vol. 1, no. 25, pp. 23-48, 1997.
- [117] G. Littlewort, I. Fasel, M.S. Bartlett, and J.R. Movellan, "Fully automatic coding of basic expressions from video," INC MPLab Tech Report 3, San Diego, La Jolla, CA, 2002.
- [118] M. Bartlett, G. Donato, J. Movellan, J. Hager, P. Ekman, and T. Sejnowski, "Image representations for facial expression coding," In S.A. Solla, T.K. Leen, and K.-R. Muller, editors, *Advances in Neural Information Processing Systems*, vol. 12, 2000.
- [119] M.S. Bartlett, J.C. Hager, P. Ekman, and T.J. Sejnowski, "Measuring facial expressions by computer image analysis," *Psychophysiology*, vol. 2, no. 36, pp. 253-263, 1999.
- [120] H.B. Deng, L.W. Jin, L.X. Zhen, J.C. Huang, "A new facial expression recognition method based on local gabor filter bank and pca plus lda," *Information Technology*, vol. 11, no. 11, pp. 86-96, 2005.
- [121] T. Lee, "Image representation using 2D Gabor wavelets," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, no. 18, pp. 959-971, 1996.
- [122] Y.L. Tian, T. Kanade, and J.F. Cohn, "Eye-state action unit detection by gabor wavelets," in *Multimodal Interfaces*, Beijing, China, 2000.

- [123] Y.L. Tian, T. Kanade, and J.F. Cohn, "Evaluation of gabor-wavelet-based facial action unit recognition in image sequences of increasing complexity," in *5th IEEE International Conference on Automatic Face and Gesture Recognition*, Washington, DC, USA, 2002.
- [124] A. Saxena, A. Anand, A. Mukerjee, "Robust facial expression recognition using spatially localized geometric model," in *International Conference on Systemics, Cybernetics and Informatics (ICSCI)*, Hyderabad, India, 2004.
- [125] I. Cohen, N. Sebe, L. Chen, A. Garg, and T.S. Huang, "Facial expression recognition from video sequences: Temporal and static modelling," *Computer Vision and Image Understanding: Special Issue on Face Recognition*, no. 91, pp. 160-187, 2003.
- [126] G. Jacovitti, A. Neri, "Multiscale image features analysis with circular harmonic wavelets," *Wavelet Applications in Signal and Image Processing III*, vol. 2569, pp. 363-372, 1995.
- [127] L. Capdiferro, V. Casieri, A. Laurenti, and G. Jacovitti, "Multiple feature based multiscale image enhancement," in *IEEE Digital Signal Processing*, Los Alamitos, 2002.
- [128] H. Ahmadi, A. Pousaberi, A. Azizzadeh, and M. Kamarei, "An efficient iris coding based on Gauss-Laguerre wavelets," in *2nd IAPR/IEEE International Conference on Biometrics*, Seoul, South Korea, 2007.
- [129] A. Sohail and P. Bhattacharya, "Classification of facial expressions using k-nearest neighbor classifier," in *3rd International Conference on Computer Vision/Computer Graphics Collaboration Techniques*, Ghent, Belgium, 2007.
- [130] T. Kanade, J.F. Cohn, and Y. Tian, "Comprehensive database for facial expression analysis," in *4th IEEE International Conference on Automatic Face and Gesture Recognition*, Grenoble, France, 2000.
- [131] M. Spicher, "Using Matlab neural network toolbox for facial expression recognition," 2008.
- [132] M. Pantic, M.F. Valstar, R. Rademaker, L. Maat, "Web-based database for facial expression analysis," in *IEEE International Conference on Multimedia and Expo*, Amsterdam, Netherlands, 2005.
- [133] C. Shan, G. Shaogang, and P.W. McOwan, "Facial expression recognition based on local binary patterns: a comprehensive study," *Image and Vision Computing*, no. 27, pp. 803-816, 2009.
- [134] M. Lyons, J. Budynek, and S. Akamatsu, "Automatic classification of single facial images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 12, pp. 1357-1362, 1999.
- [135] X. Feng, B. Lv, Z. Li, and J. Zhang, "A novel feature extraction method for facial expression recognition," in *Joint Conference on Information Sciences*, Kaohsiung, Taiwan, 2006.

- [136] A. Sánchez, J.V. Ruiz, A.B. Moreno, A.S. Montemayor, J. Hernandez, and J.J. Pantrigo, "Differential optical flow applied to automatic facial expression recognition," *Neurocomputing*, vol. 7, no. 8, pp. 1272-1282, 2011.
- [137] E. Cerezo, I. Hupont, S. Baldassarri, and S. Ballano, "Emotional facial sensing and multimodal fusion in a continuous 2D affective space," *Ambient Intelligence and Humanized Computing*, vol. 3, no. 1, pp. 31-46, 2011.
- [138] R. Zhi, and Q. Ruan, "Facial expression recognition based on two dimensional discriminant locality preserving projections," *Neuro Computing*, no. 71, pp. 1730-1734, 2008.
- [139] W. Liejun, Q. Xizhong, and Z. Taiyi, "Facial expression recognition using improved support vector machine by modifying kernels," *Information Technology Journal*, vol. 4, no. 8, pp. 595-599, 2009.
- [140] L. Zhao, G. Zhuang, and X. Xu, "Facial expression recognition based on PCA and NMF," in *7th World Congress on Intelligent Control and Automation*, Chongqing, China, 2008.
- [141] G. Guo and C. R. Dyer, "Learning from examples in the small sample case: Face expression recognition," *IEEE Transactions on System, Man, and Cybernetics*, vol. 3, no. 35, pp. 477-488, 2005.
- [142] Y. Zhan, J. Ye, D. Niu, and P. Cao, "Facial expression recognition based on Gabor wavelet transformation and elastic templates matching," *International Journal of Image Graphics*, pp. 125-138, 2006.
- [143] G. Littlewort, M. Bartlett, I. Fasel, J. Susskind, and J. Movellan, "Dynamics of facial expression extracted automatically from video," *Image Vision Computing*, vol. 6, no. 24, pp. 615-625, 2006.
- [144] P. Yang, Q. Liu, and D.N. Metaxas, "Exploring facial expressions with compositional features," in *IEEE Conference on Computer Vision and Pattern Recognition*, San Francisco, 2010.
- [145] I. Kotsia, and I. Pitas, "Facial expression recognition in image sequences using geometric deformation features and support vector machines," *IEEE Transactions on Image Processing*, vol. 1, no. 16, pp. 172-187, 2007.
- [146] A. Jorstad, D. Jacobs, and A. Trouv, "A deformation and lighting insensitive metric for face recognition based on dense correspondences," in *IEEE Conference on Computer Vision and Pattern Recognition*, Colorado Springs, 2011.
- [147] I. Naseem, R. Togneri, and M. Bennamoun, "Linear regression for face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, pp. 2106-2111, 2010.
- [148] P. Nagesh and B. Li, "A compressive sensing approach for expression- invariant face recognition," in *EEE Conference on Computer Vision Pattern Recognition*, Miami, FL, 2009.
- [149] P.H. Tsai and T. Jan, "Expression-invariant face recognition system using subspace model analysis,"

- in *IEEE International Conference on Systems, Man and Cybernetics*, Waikoloa, Hawaii, USA, 2005.
- [150] M.A.O. Vasilescu and D. Terzopoulos, "Multilinear subspace analysis of image ensembles," in *IEEE Conference on Computer Vision and Pattern Recognition*, Madison, WI, USA, 2003.
- [151] S. Li and A. Jain, *Handbook of face recognition*, Springer, 2005.
- [152] D. O. Gorodnichy, "Video-based framework for face recognition in video," in *International Workshop on Face Processing in Video*, Victoria, British Columbia, Canada, 2005.
- [153] H. C. a. A. J. U. Park, "3D model-assisted face recognition in video," in *2nd Canadian Conference on Computer and Robot Vision*, Victoria, British Columbia, Canada, 2005.
- [154] B. Amberg, R. Knothe, and T. Vetter, "Expression-invariant 3D face recognition with a morphable model," in *8th IEEE International Conference on Automatic Face and Gesture Recognition*, Amsterdam, The Netherlands, 2008.
- [155] A.M. Martinez, "Recognizing expression variant faces from a single sample image per class," in *IEEE Conference on Computer Vision and Pattern Recognition*, Madison, Wisconsin, 2003.
- [156] N. Sang, J. Wu, and K. Yu, "Local Gabor Fisher Classifier for Face Recognition," in *4th International Conference on Image and Graphics*, Chengdu, Sichuan, China, 2007.
- [157] V. Perlibakas, "Face Recognition using principal component analysis and log-gabor filters," <http://arxiv.org/pdf/cs/0605025.pdf>, 2006.
- [158] J. Zhao, Z. Zhou, F. Cao, "Human face recognition based on ensemble of polyharmonic," *Neural Computing and Applications*, vol. 24, no. 6, pp. 1317-1326, 2014.
- [159] H. Boughrara, L. Chen, C. B. Amar, and M. Chtourou, "Face recognition under varying facial expression based on perceived facial images and local feature matching," in *International Conference on Information Technology and e-Services*, Changhua, Taiwan, 2012.
- [160] X. Xie, K.M. Lam, "Face recognition using elastic local reconstruction based on a single face image," *Pattern Recognition*, vol. 41, no. 1, pp. 406-417, 2008.
- [161] L. Wei, Y. S.B. Yin, "Face recognition using common vector based on curvelet transform," in *International MultiConference of Engineers and Computer Scientists*, Hong Kong, 2011.
- [162] C. Singh, E. Walia, and N. Mittal, "Face recognition using zernike and complex zernike moment features," *IEEE Transactions on Pattern Recognition and Image Analysis*, vol. 21, no. 1, pp. 71-81, 2011.
- [163] M.E. Wibowo, D. Tjondronegoro, L. Zhang, "On the application of the probabilistic linear discriminant analysis to face recognition across expression," in *IEEE International Conference on*

Multimedia & Expo, Melbourne, 2012.

- [164] P. Tsai, T. Jan, and T. Hintz, "Expression-invariant face recognition for small class problem," in *IEEE International Conference on Computational Intelligence for Measurement Systems and Applications*, SICILY, ITALY, 2005.
- [165] M. Aharon, M. Elad, and A. M. Bruckstein, "The K-SVD: an algorithm for designing of overcomplete dictionaries for sparse representation," *EEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4311-4322, 2006.
- [166] M. Yang, L. Zhang, X. Feng, and D. Zhang, "Fisher discrimination dictionary learning for sparse representation," in *13th International Conference on Computer Vision*, Barcelona, Spain, 2011.
- [167] FDDL code: , "<http://www4.comp.polyu.edu.hk/~cslzhang/papers.htm>," [Online].
- [168] KSVD source code:, "<http://www.cs.technion.ac.il/~elad/software/>," [Online].
- [169] JSM source code:, "<http://spams-devel.gforge.inria.fr/>," [Online].
- [170] G. Ford, "Fully automatic coding of basic expressions from video," Machine Perception Lab, Institute for Neural Computing, University of California, San Diego, 2002.
- [171] Z. Wen, and T. Huang, "Capturing subtle facial motions in 3D face tracking," *International Conference On Computer Vision*, vol. 2, pp. 1343-1350, 2003.
- [172] M. M. Khan, R. D. Ward, and M. Ingleby, "Classifying pretended and evoked facial expressions of positive and negative affective states using infrared measurement of skin temperature," *ACM Transactions on Applied Perception*, vol. 1, no. 6, pp. 91-113, 2009.
- [173] L. Trujillo, G. Olague, R. Hammoud, and B. Hernandez, "Automatic feature localization in thermal images for facial expression recognition," in *IEEE Conference on Computer Vision and Pattern Recognition*, San Diego, CA, USA, 2005.
- [174] P. Shen, S. Wang, and Z. Liu, "Facial expression recognition from infrared thermal videos," in *12th International Conference on Intelligent Autonomous Systems*, Jeju Island, Korea, 2013.
- [175] Y. Koda, Y. Yoshitomi, M. Nakano, and M. Tabuse, "A facial expression recognition for a speaker of a phoneme of vowel using thermal image processing and a speech recognition system," in *18th IEEE International Symposium on Robot and Human Interactive Communication*, Japan, 2009.
- [176] S. Wang, S. Lv, X. Wang, "Infrared facial expression recognition using wavelet transform," in *International Symposium on Computer Science and Computational Technology*, Shanghai, China, 2008.
- [177] I. O. W. S. Bench, "DOE University Research Program in Robotics under Grant DOE-DE-FG02-86NE37968; DOD/TACOM/NAC/ARC Program under Grant R01-1344-18; FAA/NSSA Grant R01-

1344-48/49; Office of Naval Research under Grant No. N000143010022".

- [178] S. Wang, Z. Liu, S. Lv, Y. Lv, G. Wu, P. Peng, F. Chen, and X. Wang, "A natural visible and infrared facial expression database for expression recognition and emotion inference," *IEEE Transactions on Multimedia*, vol. 7, no. 12, pp. 682-691, 2010.
- [179] C. c. t. f. Matlab, "http://www.vision.caltech.edu/bouguetj/calib_doc/htmls/example5.html," [Online].
- [180] J. Wang and E. Sung, "Facial feature extraction in an infrared image by proxy with a visible face image," *IEEE Transactions on Instrumentation and Measurement*, vol. 65, no. 5, pp. 2057-2066, 2007.
- [181] I. Buciu, and I. Pitas, "Application of non-negative and local non negative matrix factorization to facial expression recognition," in *International Conference on Pattern Recognition*, Cambridge, UK, 2004.
- [182] I.A. Oz and M.M. Khan, "Efficacy of biophysiological measurements at FTFPs for facial expression classification: A validation," in *International Conference on Biomedical and Health Informatics*, Hong Kong, China, 2012.
- [183] S. Wang, S. He, "Spontaneous facial expression recognition by fusing thermal infrared and visible images," *Intelligent Autonomous Systems 12, Advances in Intelligent Systems and Computing*, vol. 194, pp. 263-272, 2013.
- [184] S. He, S. Wang, W. Lan, H. Fu, and Q. Ji, "Facial expression recognition using deep Boltzmann machine from thermal infrared images," in *Humaine Association Conference on Affective Computing and Intelligent Interaction*, Washington, 2013.
- [185] Z. Wang, and S. Wang, "Spontaneous facial expression recognition by using feature-level fusion of visible and thermal infrared images," in *IEEE International Workshop on Machine Learning for Signal Processing*, Beijing, China, 2011.
- [186] Z. Liu, and S. Wang, "Emotion recognition using hidden markov models from facial temperature sequence," in *International Conference on Affective Computing and Intelligent Interaction*, Memphis, TN, USA, 2011.
- [187] "Total Information Awareness DAPRA's Research Program," *Information & Security*, vol. 10, pp. 105-109, 2003.
- [188] "TSA Guidance Package: 'Biometrics for Airport Access Control' .," September 2005. [Online].
- [189] J.R. Matey, O. Naroditsky, K. Hanna, R. Kolczynski, D.J. Lolacono, S. Mangru, M. Tinker, T.M. Zappia, W.Y. Zhao, "Iris on the move: acquisition of images for iris recognition in less constrained environments," *Proceedings of the IEEE*, vol. 94, no. 11, pp. 1936-1947, 2006.

- [190] MorphoTrak; "<http://www.morphotrak.com/MorphoTrak/MorphoTrak/>," [Online].
- [191] S. Chague, B. Droit, O. Boulanov, S.N. Yanushkevich, V.Shmerko, "Biometric-based decision support assistance in physical access control systems," in *Conference on Bio-inspired, Learning and Intelligent Systems for Security*, Edinburgh, 2008.
- [192] S.N. Yanushkevich, A. Stoica, V.P. Shmerko, "Experience of design and prototyping of a multi-biometric early warning physical access control security system (PASS) and a training system (T-PASS)," in *32nd Annual IEEE Industrial Electronics Society Conference*, Paris, 2006.
- [193] "http://en.wikipedia.org/wiki/Future_Attribute_Screening_Technology," [Online].
- [194] "DHS: privacy impact assessment for the future attribute screening technology (FAST) project," dhs.gov, retrieved, 2008.
- [195] S. Yanushkevich, A. Stoica, V. Shmerko, "Fundamentals of Biometric-based Training," in S. Yanushkevich, P. Wang, M. Gavrilova, S. Srihari (Editors), *Image Pattern Recognition: Synthesis and Analysis in Biometrics, Machine Perception and Artificial Intelligence*, World Scientific Publishing Co., New Jersey-London-Singapore, 2007, pp. 365-406.
- [196] R.F. Royal and S.R. Schutt, *The gentle art of interviewing and interrogation: a professional manual and guide*, Prentice-Hall, 1976.
- [197] R. Stollhof, "Modeling prosopagnosia: computational theory and experimental investigations of a deficit in face recognition," Phd thesis, <http://www.fmi.uni-leipzig.de/promotion/abstract.stollhoff.pdf>, 2010.
- [198] M. Wimmer, B.A. MacDonald, D. Jayamuni, A. Yadav, "Facial expression recognition for human-robot interaction – A prototype," *Robot Vision*, vol. 4931, pp. 139-152, 2008.
- [199] C. Lang, S. Wachsmuth, H. Wersing, and M. Hanheide, "Facial expressions as feedback cue in human-robot interaction—a comparison between human and automatic recognition performances," in *Computer Vision and Pattern Recognition Workshops*, San Francisco, CA, 2010.
- [200] N. Sebe, M.S. Lew, T.S. Huang, "The state-of-the-art in human-computer interaction," Prague, Czech Republic, 2004.
- [201] T.K. Tews, M. Oehl, F.W. Siebert, R. Höger, H. Faasch, "Emotional human-machine interaction: cues from facial expressions," in *Human Interface and the Management of Information. Interacting with Information*, Orlando, FL, USA, 2011.
- [202] S. Marcos, J. Gomez-Garcia-Bermejo, E. Zalama, J. Lopez, "Nonverbal communication with a multimodal agent via facial expression recognition," in *IEEE International Conference on Robotics and Automation*, Shanghai, China, 2011.

- [203] B. Tu, F. Yu, "Bimodal emotion recognition based on speech signals and facial expression," vol. 122, pp. 691-696, 2012.
- [204] F. Alonso, M. Malfaz, J.F. Gorostiza, M.A. Salichs, "A multimodal emotion detection system during human–robot interaction," *Journal of Sensors*, vol. 13, no. 11, pp. 15549-15581, 2013.
- [205] A. Ernst, T. Ruf, and C. Kueblbeck, "A modular framework to detect and analyze faces for audience measurement systems," in *2nd Workshop on Pervasive Advertising*, Lubeck, Germany, 2009.
- [206] SHORE demo; available online:, "<http://www.iis.fraunhofer.de/en/bf/bsy/produkte/shore.html>," [Online].
- [207] S. Yildirim, S. Narayanan, A. Potamianos, "Detecting emotional state of a child in a conversational computer game," *Computer Speech and Language*, no. 25, pp. 29-44, 2011.
- [208] M. K. A. online, "[http://en.wikipedia.org/wiki/Kinect#Kinect on the Xbox – One](http://en.wikipedia.org/wiki/Kinect#Kinect_on_the_Xbox_One)," [Online].
- [209] "Diagnostic and statistical manual of mental disorders, Fourth Edition," American Psychiatric Association, 1994.
- [210] A. Lerner, *Diagnostic criteria in neurology*, Humana Press, 2006.
- [211] V. Bevilacqua, D. D’Ambruoso, G. Mandolino, and M. Suma, "A new tool to support diagnosis of neurological disorders by means of facial expressions," in *IEEE International Workshop on Medical Measurements and Applications*, Ottawa-Canada, 2011.
- [212] C.C.V. de Almeida Rocca, E.V.D. Heuvel, S.C. Caetano and B. Lafer, "Facial emotion recognition in bipolar disorder: a critical review," *Revista Brasileira de Psiquiatria*, vol. 31, no. 2, pp. 171-180, 2009.
- [213] C.G. Kohler, G. Anselmo-Gallagher, W. Bilker, J. Karlawish, R.E. Gue, and C.M. Clark, "Emotion-discrimination deficits in mild Alzheimer disease," *American Journal of Geriatric Psychiatry*, vol. 13, pp. 926-933, 2005.
- [214] C.J. Fox, G. Iaria, and J.J.S. Barton, "Disconnection in prosopagnosia and face processing," *Cortex*, vol. 44, pp. 996-1009, 2008.
- [215] M. Harms, A. Martin, G. Wallace, "Facial emotion recognition in autism spectrum disorders: A review of behavioral and neuroimaging studies," *Neuropsychology Review*, vol. 20, pp. 290-322, 2010.
- [216] B.C. Duchaine, and K. Nakayama, "Developmental prosopagnosia: A window to content-specific face processing," *Current Opinion in Neurobiology*, vol. 61, no. 2, pp. 166-173, 2006.
- [217] J. Barton, "Disorders of face perception and recognition," *Neurologic Clinics*, vol. 21, pp. 521-548,

2003.

- [218] J.J.S. Barton, D.Z. Press, J.P. Keenan, and M. Oconnor, "Lesions of the fusiform face area impair perception of facial configuration in prosopagnosia," *Neurology*, vol. 58, pp. 71-78, 2002.
- [219] E. De Renzi, P. Faglioni, D. Grossi, and P. Nichelli, "Apperceptive and associative forms of prosopagnosia," *Cortex*, vol. 27, pp. 213-231, 1991.
- [220] I. Kennerknecht, T. Grueter, B. Welling, S. Wentzek, J. Horst, S. Edwards, and M. Grueter, "First report of prevalence of non-syndromic hereditary prosopagnosia (HPA)," *Medical Genetics Part A*, vol. 140, no. 15, pp. 1617-1622, 2006.
- [221] X. Wang , X. Zhao, V. Prakash, W. Shi, and O. Gnawali, "Computerized-eyewear based face recognition system for improving social lives of prosopagnosics," in *7th International Conference on Pervasive Computing Technologies for Healthcare*, Venice, Italy, 2013.
- [222] I. Bacivarov, "Advances in the modeling of facial sub-regions and facial expressions using active appearance techniques," Phd Thesis, National University of Ireland, Galway , 2009 .
- [223] Y. Ren, "Facial expression recognition system," Master thesis, 2008.
- [224] Y. Tian, T. Kanade, J.F. Cohn, "Recognizing upper face action units for facial expression analysis," in *IEEE Conference on Computer Vision and Pattern Recognition*, Hilton Head, SC, USA, 2000.
- [225] L. Sorgi, N. Cimminiello, and A. Neri, "Keypoints selection in the Gauss Laguerre transformed domain," in *Proceedings of the British Machine Vision*, Edinburgh, UK, 2006.
- [226] S.M. Lajevardi, Z.M. Hussain, "Hybrid feature extraction for facial expression recognition," *Advances in Modelling Series B: Signal Processing and Pattern Recognition*, vol. 53, no. 1, pp. 34-50, 2009.
- [227] A. Kar, "Unsupervised temporal segmentation of facial behaviour," [Online].

APPENDIX A: FACIAL EXPRESSION RECOGNITION IN VIDEO

The aim of this additional study is to show the feasibility and the "extendibility" of the proposed approach when applying it to videos. Needless to say, the still image analysis is totally different from the approach used for the video processing, although the core basis remains the same. For example, face detection for both the still image and video utilizes the same approach, but for video, the face detection is usually done for the first frame (or until the face is detected no matter which frame is that), and then a tracking algorithm is used to track the face instead of detecting it in every single frame onward. In this thesis, we focus on the still images. However, we investigated how the proposed method can be applied to video, and whether the underlying principles are the same. A small video dataset has been selected from the data in the CK database to test this hypothesis.

We used the same geometric features (Figure 6-7) and the same textural features (only for visible image) in Figure 6-8. Figure 0-1 shows a sample sequence used in the experiment. The happiness expression develops starting from neutral face to the apex (last frame). CK and CK+ databases provided the location of 68 points on every single frame which have been extracted from the AAM algorithm. In real-time application, these points can be extracted and then being tracked. These 68 features and the related face textures are tracked through an image sequence. AAM method needs an initial points to fit the defined mask (See 3.1) and sometimes it is done by manually clicking the first frame. In our approach after the face is detection in the first frame (or the first frame where the face is available), we can apply the facial points detection algorithm we have proposed in this thesis. In this case, there is no need to manually select any initial position. Out of the 68 provided facial features, we use the 21 points we utilized in this thesis (Figure 6-7).

Figure 0-2 shows the 68 facial points (left) and the 21 points (right) we used in our model. Then, the Delaunay Triangulation of AAM tracked features is constructed. Delaunay triangulation of a set of points, guarantees that the circumcircle of any triangle does not include other points in its interior. This method is quite common in generating 2D or 3D meshes for FR and FER. The relative displacement of these tracked points over time can be used as features since these deformations vary from expression to expression.



Figure 0-1. Typical sequence from CK database [51] used in the experiment: Developing happy expressions.

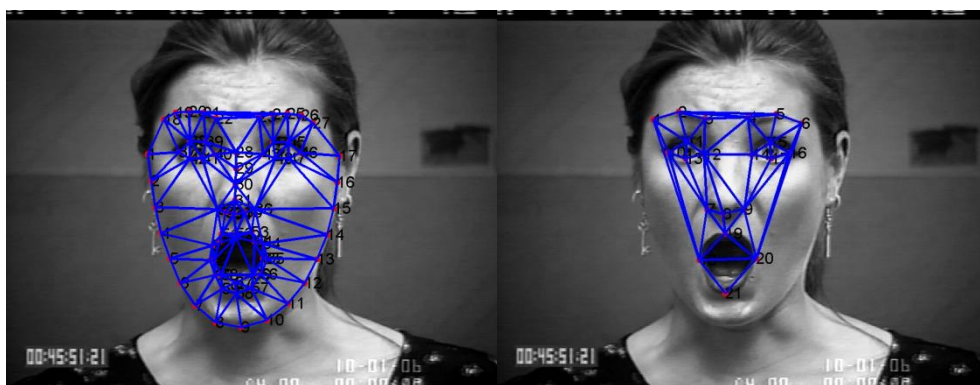


Figure 0-2. The 68 facial points from CK database (left) and the 21 points we used to construct face mesh. The Delaunay triangulation is overlaid on these points.

Figure 0-3 shows the detected facial points and corresponding Delaunay triangulation for all the images in Figure 0-1. The question in video processing is how to correctly classify the frames between neutral and apex. If the frame has the complete expression (like the second row in

Figure 0-1 where the happiness expression is existed), we can simply apply the same FER approach on images. For the frames which do not contain the “Exact” expression but the expression transient, the classifier should somehow assign a “score” or the “strength” of the expression in the frame with some probabilities to different distinct expressions. A good survey on different approaches can be found in [222].

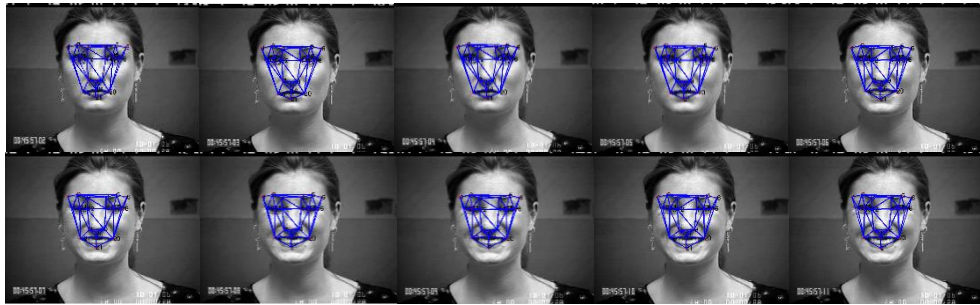


Figure 0-3. Tracking facial points for the given sequence using AAM for selected points.

For classification, we used 28 KNN classifiers. Each classifier in the form of Class1/Class2 categorizes the given feature vector from a test image into Class1 or Class2. For example in Table 0-1, The KNN#1 which is HA/DI is constructed to only distinguish between happiness and disgust expression. Each input image is classified by comparing its features with a template corresponding to each expression class. The template is constructed for each two expressions to be classified using training images. The 28 KNN classifiers we used are presented in Table 0-1.

(AN: Angry, HA: Happy, SA: Sad, SU: Surprise, DI: Disgust, FE: Fear, NE: Neutral)

For each classifier, the distance between the classifier’s trained feature vector and each image test feature vector parameters is calculated. The minimum distance between all classifiers is assigned to input image (frame).

In video face processing the identity remains constant, while pose, illumination and expression vary. For the FER and FR in video, after detecting facial points in first frame and tracking them in the next frames, the normalization we described in Figure 4-12 should be performed beforehand. The normalization (simply alignment) ensures that we are comparing corresponding texture features around each tracked nodes.

Table 0-1. The 28 KNN classifiers used for video processing.

KNN#	Class1/Class2	KNN#	Class1/Class2
1	HA/DI	15	AN/NE
2	HA/AN	16	SA/SU
3	HA/SA	17	SA/FE
4	HA/SU	18	SA/NE
5	HA/FE	19	SU/FE
6	HA/NE	20	SU/NE
7	DI/AN	21	FE/NE
8	DI/SA	22	NE/NOT NE
9	DI/SU	23	HA/NOT HA
10	DI/FE	24	DI/NOT DI
11	DI/NE	25	AN/NOT AN
12	AN/SA	26	SA/NOT SA
13	AN/SU	27	SU/NOT SU
14	AN/FE	28	FE/NOT FE

In our experiments, we used CK+ database which contains images from 123 subjects. Since some of the subjects in the database do not have all the six expressions sequences, we used only a subset of 58 subjects, for which at least four of the sequences were available. The 10-fold cross validation is used to test the performance of the system. The database was randomly partitioned to 10 segments, and 9 partitions were used for training and the remaining partition was used to

test performance. The procedure was repeated so that every equal-sized set was used once as the test set. Finally, the average of recognition rate is reported.

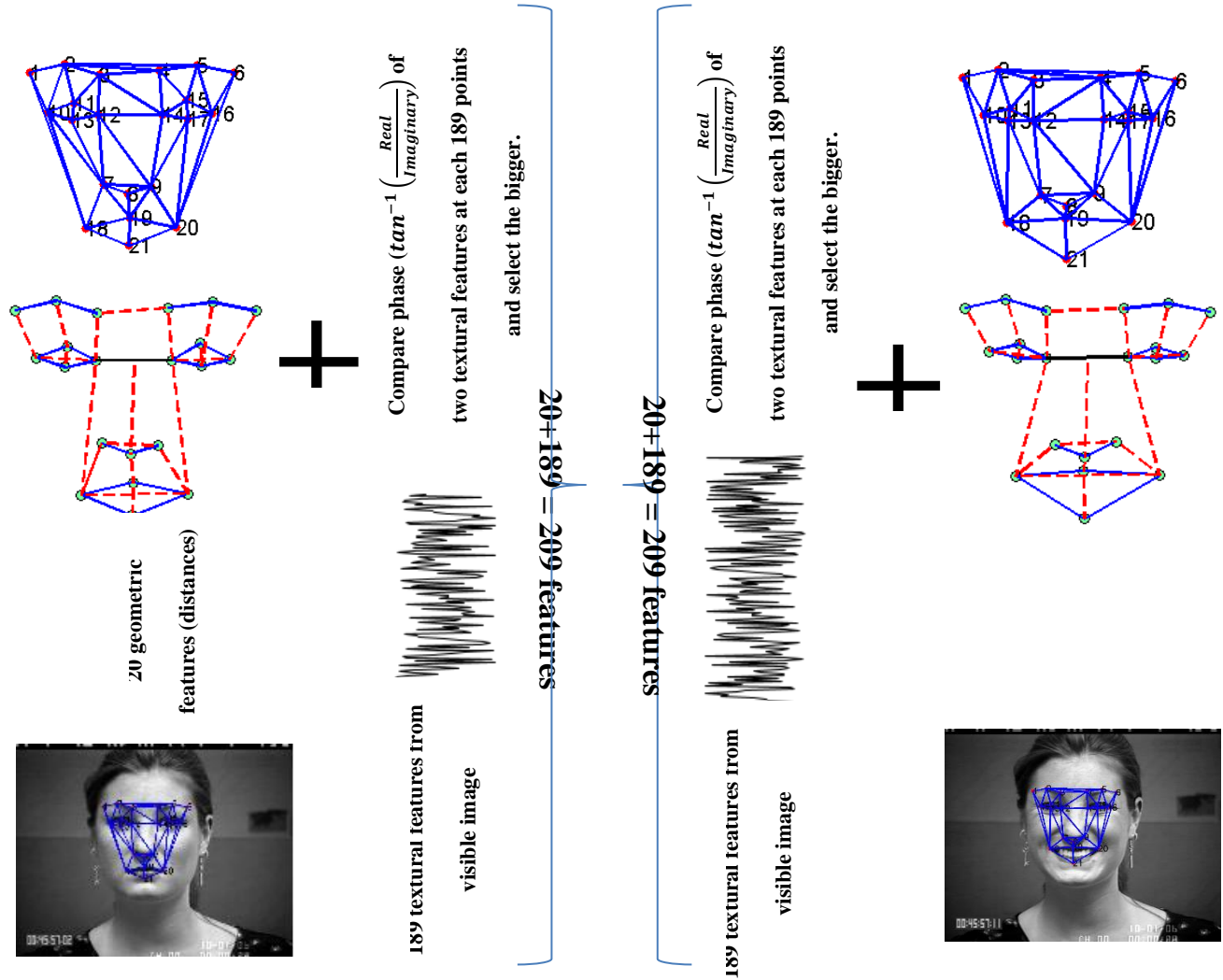


Figure 0-4. The features used in FER: From each from in a sequence, the 209 features are extracted

Table 0-2 summarized the classification rates. Each number shows the average classification rate between two expressions from the constructed KNN. For example the classification rate between DI and FE is 85.1%.The correct classification rates must be read by looking in the tables at the

line corresponding to class 1 and the column corresponding to class 2. For example, in Table 8.2, the correct discrimination between the disgusted expression and the happy one must be read at the value at line 2, column 4. The table shows that the most challenging expression to be classified is fear and the easiest one is happiness. The average recognition rate (separately) based on the numbers in the table is 87.93%.

Table 0-2. The classification rate (%) for different expressions using different KNNs.

	AN	DI	HA	FE	SA	SU	NE
AN	92.3	79.0	94.7	73.4	86.1	92.2	83.1
DI		84.3	97.5	85.1	93.9	96.8	90.5
HA			89.6	98.2	99.1	91.0	100
FE				77.6	71.4	85.2	73.3
SA					85.2	90.4	75.9
SU						91.1	97.3
NE							87.7

In summary, the proposed feature extraction technique even without using any spatio-temporal information from image sequences shows good recognition rate. There are many sophisticated methods for facial expression analysis in video which use different time-series analysis to classify the expression. In addition, more powerful classifiers for video processing like SVM can help to improve recognition rate. The aim of this short study was to show that the proposed method can be used as a good start point for video analysis as well.