

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

Αντίληψη και αναγνώριση συναισθήματος σε
εικόνες προσώπου με εφαρμογή σε συστήματα
συναισθηματικής αλληλεπίδρασης
ανθρώπου-υπολογιστή

Ιωάννα-Ουρανία Σταθοπούλου

Διδακτορική Διατριβή

Πειραιάς

Ιούνιος, 2009

UNIVERSITY OF PIRAEUS

**Emotion Perception and Recognition in Face
Images with Applications in Affective
Human-Computer Interaction Systems**

by

Ioanna-Ourania Stathopoulou

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

Doctor of Philosophy

Piraeus, Greece

June, 2009

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΡΠΑ

©Copyright 2009

by

Ioanna-Ourania Stathopoulou

All Rights Reserved

Approved thesis committee

Date and Signature

Member 1, Chairman, George A. Tsihrintzis,
Associate Professor, University of Piraeus

Member 2, Thesis Adviser, Christos Douligeris,
Professor, University of Piraeus

Member 3, Thesis Adviser, Maria Virvou,
Associate Professor, University of Piraeus

Member 4, Nikolaos Alexandris,
Professor, University of Piraeus

Member 5, Christoforos Nikou,
Assistant Professor, University of Ioannina

Member 6, Athanassia Alonistioti,
Lecturer, National and Kapodistrian University of Athens

Member 7, Aggelos Pikrakis,
Lecturer, University of Piraeus

Επιτροπή Διδακτορικού

Ημερομηνία και Υπογραφή

1. Επιβλέπων, Γεώργιος Τσιχριντζής,
Αναπληρωτής Καθηγητής Πανεπιστημίου Πειραιώς

2. Μέλος της τριμελούς επιτροπής, Χρήστος Δουληγέρης,
Καθηγητής Πανεπιστημίου Πειραιώς

3. Μέλος της συμβουλευτικής επιτροπής, Μαρία Βίρβου
Αναπληρώτρια Καθηγήτρια Πανεπιστημίου Πειραιώς

4. Νικόλαος Αλεξανδρής,
Καθηγητής Πανεπιστημίου Πειραιώς

5. Χριστόφορος Νίκου,
Επίκ. Καθηγητής Πανεπιστημίου Ιωαννίνων

6. Αθανασία Αλωνιστιώτη,
Λέκτορας Εθνικού & Καποδιστριακού Πανεπιστημίου Αθηνών

7. Άγγελος Πιχράκης,
Λέκτορας Πανεπιστημίου Πειραιώς

ABSTRACT

Emotion Perception and Recognition in Face Images with Applications in Affective Human-Computer Interaction Systems

by

Ioanna-Ourania Stathopoulou

Faces provide a wide range of information about a person's identity, race, sex, age, and emotional state. In this thesis, we study the perception of facial expressions of emotion in our aim at developing a fully automated facial expression recognition system. Our studies begin with a research in the literature about the emotion perception from the scientific - psychological and medical - point of view. Based on these studies, we came up with the following assumptions: (1) a number of brain parts play a significant role in emotion perception and expression, (2) there are six 'basic emotions', namely: 'anger', 'disgust', 'fear', 'happiness', 'sadness' and 'surprise' and, (3) there is cultural specificity in emotion perception and expression. The latter is further strengthened by our own empirical studies conducted to humans. Specifically, we developed two different questionnaires, wherein participants were shown face images and were asked to classify the emotion. In the first questionnaire we used images gathered from the web, while in the second questionnaire we used images of Greeks forming an expression. The difference in the success rates further demonstrates that there is cultural specificity in the ways people understand and express the emotion.

Moreover, from our empirical studies, we were able to identify the emotion classes that are present during a typical human-computer interaction session, which are namely: 'neutral', 'happiness', 'sadness', 'surprise', 'anger', 'disgust' and 'boredom-sleepiness'. Towards building our facial expression recognition system, we constructed our own face image database, which consisted of a two different sets of face images in front and side view: low quality images which were acquired by using web cameras and high quality face images which were acquired by using digital cameras of high resolution. Finally, we developed our own facial expression recognition system, which consists of two modules: (1) a face detection subsystem and, (2) a facial expression recognition subsystem. Our face detection subsystem is based on neural network-based classifiers. For our facial expression recognition subsystem, we considered neural network-based and other classifiers, but we concluded that Support Vector Machine-based Classifiers demonstrated better results. The feature extraction process, performance evaluations and test results are demonstrated and analyzed.

ΠΕΡΙΛΗΨΗ

Αντίληψη και Αναγνώριση Συναισθήματος σε Εικόνες Προσώπου με Εφαρμογή σε Συστήματα Συναισθηματικής Αλληλεπίδρασης Ανθρώπου-Υπολογιστή

από

Ιωάννα-Ουρανία Σταθοπούλου

Η εικόνα του προσώπου παρέχει σημαντική πληροφορία σχετικά με την ταυτότητα, τη κουλτούρα, το φύλο, την ηλικία και την ψυχολογική κατάσταση ενός ατόμου. Σε αυτή τη διατριβή, μελετάται ο τρόπος αντίληψης των εκφράσεων προσώπου, με σκοπό την ανάπτυξη ενός πλήρως αυτοματοποιημένου συστήματος αναγνώρισης εκφράσεων. Αρχικά, μελετάται η υπάρχουσα επιστημονική έρευνα, στους τομείς της φυσιολογίας και της ψυχολογίας, σχετικά με το πρόβλημα της αντίληψης συναισθήματος. Αυτή η έρευνα μας οδήγησε στα εξής συμπεράσματα: (1) υπάρχουν κάποια μέρη του εγκεφάλου που παίζουν σημαντικό ρόλο στην αντίληψη και την έκφραση των συναισθημάτων, (2) υπάρχουν κάποια βασικά συναισθήματα, τα οποία είναι: 'θυμός', 'αηδία', 'φόβος', 'χαρά', 'λύπη' και 'έκπληξη' και, (3) υπάρχουν διαφορές στον τρόπο που ο άνθρωπος αντιλαμβάνεται και εκφράζει τα συναισθήματα, ανάλογα με την κουλτούρα του. Το τελευταίο συμπέρασμα, ενισχύεται ακόμα περισσότερο από τις δικές μας εμπειρικές μελέτες. Σε αυτές τις μελέτες, δημιουργήσαμε δυο διαφορετικά ερωτηματολόγια, όπου επιδεικνύαμε μια εικόνα προσώπου στο συμμετέχοντα και του ζητούσαμε να επιλέξει το αντίστοιχο συναίσθημα. Για το πρώτο ερωτηματολόγιο, χρησιμοποιήσαμε εικόνες από το διαδίκτυο, ενώ στο δεύτερο ερωτηματολόγιο, επιλέξαμε εικόνες ελλήνων οι οποίοι σχηματίζουν

κάποια έκφραση. Η διαφορά στα ποσοστά επιτυχίας, ενισχύει το συμπέρασμα ότι υπάρχουν διαφορές, ανάλογα με την κουλτούρα, στον τρόπο που ο άνθρωπος αντιλαμβάνεται και εκφράζει κάποιο συναίσθημα. Επίσης, από τις εμπειρικές μελέτες, μπορέσαμε να ορίσουμε τις κλάσεις συναισθημάτων που μπορεί να παρουσιαστούν κατά τη διάρκεια χρήσης του υπολογιστή. Αυτές οι κλάσεις συναισθημάτων είναι οι παρακάτω: ' ουδέτερη', 'χαρά', 'λύπη', 'έκπληξη', 'θυμός', 'αηδία' και 'βαρεμάρα-νύστα'. Για την κατασκευή του συστήματος αναγνώρισης εκφράσεων, δημιουργήσαμε τις δικές μας βάσεις εικόνων, οι οποίες αποτελούνται από δύο διαφορετικές ομάδες εικόνων προσώπου και ορθή και πλάγια όψη. Η πρώτη ομάδα αποτελείται από εικόνες χαμηλής ποιότητας, η φωτογράφιση των οποίων έγινε με δικτυακές κάμερες (web cameras). Ενώ, η δεύτερη ομάδα αποτελείται από εικόνες υψηλής ποιότητας, οι οποίες αποκτήθηκαν από ψηφιακές κάμερες υψηλής ανάλυσης. Τέλος, δημιουργήσαμε το σύστημα αναγνώρισης εκφράσεων προσώπου, το οποίο αποτελείται από δύο υποσυστήματα: (1) το υποσύστημα ανίχνευσης προσώπου και, (2) το υποσύστημα αναγνώρισης εκφράσεων προσώπου. Το υποσύστημα ανίχνευσης προσώπου χρησιμοποιεί νευρωνικά δίκτυα για την ταξινόμηση των εικόνων σε 'πρόσωπο' ή 'μη-πρόσωπο'. Για το υποσύστημα αναγνώρισης εκφράσεων προσώπου, χρησιμοποιήσαμε αρχικά νευρωνικά δίκτυα, καθώς και άλλους ταξινομητές, αλλά καταλήξαμε σε ταξινομητές που βασίζονται σε Μηχανές Διανυσμάτων Υποστήριξης (Support Vector Machine-based Classifiers), οι οποίοι επέδειξαν τα καλύτερα αποτελέσματα. Επιπλέον, αναλύουμε τη διαδικασία επεξεργασίας εικόνας και εξαγωγής χαρακτηριστικών, επιδεικνύουμε αποτελέσματα και μετράμε την απόδοση των δύο υποσυστημάτων.

Acknowledgments

I would like to thank all people who have helped and inspired me during my doctoral study.

Foremost, I would like to record my gratitude to George A. Tsihrintzis for his supervision, advice, and guidance from the very early stage of this research as well as giving me extraordinary experiences through out the work. Above all and the most needed, he provided me unflinching encouragement and support in various ways. His truly scientist intuition had been extremely helpful in difficult and crucial circumstances. I consider myself extremely lucky working with him and I am indebted to him more than he knows. I gratefully acknowledge Maria Virvou for her advice, supervision, and crucial contribution, which made her a backbone of this research and so to this thesis. Her involvement has triggered and nourished my intellectual maturity that I will benefit from, for a long time to come.

I thank my fellow labmates in class 212 for the stimulating discussions, for all their help, support, interest and valuable hints, and for all the fun we have had all these years.

I would also like to thank all my friends who stood by me and showed extremely patience when I was feeling tired or just disappointed.

Support for this work was provided by the General Secretariat of Research and Technology, Greece, under the auspices of the PENED-2003 program. Their support is greatly appreciated.

My deepest gratitude goes to my family for their unflagging love and support throughout my life; this dissertation is simply impossible without them. My parents always, with difficulty and usually not so successfully, tried to understand the aims and the technological background of this work. I still remember my mother calling the 'neural networks' as 'neurotic networks', making me neurotic. Despite this fact,

their support was constant. It is to them that I dedicate this work. I would also like to dedicate this work to the newest member of our family, my nephew and his parents, my sister Helen and her husband Lambros.

Finally, I would like to thank everybody who was important to the successful realization of thesis, as well as expressing my apology that I could not mention personally one by one.

Contents

| | |
|---|----------|
| Abstract | ix |
| Acknowledgments | xiii |
| List of Illustrations | xix |
| List of Tables | xxv |
| 1 Introduction | 1 |
| 1.1 Motivation | 3 |
| 1.2 Contribution of this thesis | 4 |
| 1.3 Organization of this thesis | 5 |
| 2 Previous Related Psychological Studies on Emotion Perception | 9 |
| 2.1 Emotion vs affect vs feelings | 10 |
| 2.2 Emotions and culture | 12 |
| 2.2.1 Basic Emotions | 13 |
| 2.2.2 Culturally Specific Expressions of Emotions | 15 |
| 2.2.3 Higher Cognitive Emotions | 17 |
| 2.3 Neurobiology and Emotion Expression | 17 |
| 2.3.1 Cerebral Cortex | 18 |
| 2.3.2 Amygdala | 22 |
| 2.3.3 Superior Temporal Sulcus | 23 |
| 2.3.4 Implicit and Explicit Perception of Emotion | 25 |
| 2.3.5 Section Summary | 27 |
| 2.4 Expression of Emotion | 27 |
| 2.4.1 Written Language | 28 |

| | | |
|-------|--|----|
| 2.4.2 | Speech | 29 |
| 2.4.3 | Facial Expressions | 30 |
| 2.4.4 | Gestures and Body Language | 31 |
| 2.5 | Facial Expression of Emotion | 31 |
| 2.5.1 | Previous Attempts to Facial Emotion Quantification and Classification | 34 |
| 2.5.2 | Face and Facial Expressions: Their Role | 36 |
| 2.6 | The Importance of Understanding Emotions | 44 |
| 2.7 | Meeting Emotional Needs with the Help of Advanced Human-Computer Interaction Techniques | 47 |
| 2.7.1 | Supporting emotional skill needs | 48 |
| 2.7.2 | Supporting experiential needs | 49 |

3 Previous Related Studies and Systems on Emotion Recognition 51

| | | |
|-------|--|----|
| 3.1 | Face Databases | 51 |
| 3.1.1 | Specifying Requirements for an Ideal Facial Expression Database | 52 |
| 3.1.2 | Previous Facial Expression Databases | 54 |
| 3.1.3 | Section Summary - Results | 57 |
| 3.2 | Face Detection | 58 |
| 3.2.1 | Specifying Requirements for an Ideal Face Detection System . | 58 |
| 3.2.2 | Previous Works on Face Detection | 61 |
| 3.2.3 | Section Summary - Results | 66 |
| 3.3 | Facial Expression Classification System | 67 |
| 3.3.1 | Specifying Requirements for our Facial Expression Classification System | 67 |
| 3.3.2 | Facial Expression Classification Approaches | 72 |

| | | |
|----------|---|------------|
| 3.3.3 | Section Summary - Results | 101 |
| 4 | Face Image Databases | 105 |
| 4.1 | The Database of Low Quality Face Images (DBLQFI) | 106 |
| 4.2 | The Database of High Quality Face Images (DBHQFI) | 107 |
| 5 | Empirical Studies on Emotion Recognition | 117 |
| 5.1 | Preliminary Questionnaires | 118 |
| 5.2 | Newer (Detailed) Questionnaires | 119 |
| 5.2.1 | The detailed questionnaire structure | 120 |
| 5.2.2 | The observer and subject backgrounds | 123 |
| 5.3 | Results from Statistical Analysis | 123 |
| 5.3.1 | Statistical Analysis per Expression | 123 |
| 5.3.2 | Difficulties of Facial Expression Classification as Outlined by the Participants | 139 |
| 5.3.3 | Statistical Significance of the Results | 143 |
| 5.3.4 | Extraction of Facial Expression Classification Features | 145 |
| 5.4 | Summary - Conclusions | 152 |
| 6 | Visual-Facial Emotion Recognition System | 157 |
| 6.1 | Face Detection | 161 |
| 6.1.1 | P. Sinha's Template | 161 |
| 6.1.2 | The Face Detection Algorithm - Image Preprocessing | 162 |
| 6.1.3 | Artificial Neural Network-Based Face Detectors | 164 |
| 6.1.4 | Performance Evaluation | 168 |
| 6.1.5 | Summary and Conclusions | 180 |
| 6.2 | Introduction to our Facial Expression Recognition System | 181 |
| 6.3 | First attempts for facial expression recognition | 181 |

| | | |
|----------|---|------------|
| 6.3.1 | The Facial Expression Classification Algorithm (1st Attempts) | 183 |
| 6.3.2 | Feature Validation (First Attempts) | 188 |
| 6.3.3 | Neural Network Classifiers (First Attempts) | 190 |
| 6.3.4 | Results from neural network classifiers (First Attempts) | 192 |
| 6.4 | Facial expression recognition system | 196 |
| 6.4.1 | Feature Selection | 196 |
| 6.4.2 | Image Preprocessing and Feature Extraction | 202 |
| 6.4.3 | The extraction algorithm for the rest of facial features | 209 |
| 6.4.4 | Combination of all and computation of feature vector | 211 |
| 6.4.5 | Quantification of Feature Discrimination Power | 211 |
| 6.4.6 | Classifiers for Facial Expression Classification | 214 |
| 6.4.7 | Classification Performance Assessment | 216 |
| 6.4.8 | More Sophisticated Classifiers | 219 |
| 6.4.9 | Experimental performance evaluation | 225 |
| 6.5 | Summary - Conclusions | 226 |
| 7 | Conclusions and Future Work | 233 |
| 7.1 | Summary and Conclusions | 233 |
| 7.2 | Current and Future Work | 238 |
| 7.2.1 | Towards a multimodal emotion recognition system | 238 |
| 7.2.2 | Towards extending the visual facial expression recognition | 243 |
| | Bibliography | 245 |
| | Publications which have resulted from this research up to the instant of submission of this Thesis | 281 |

Illustrations

| | | |
|-----|--|-----|
| 2.1 | The nature of ‘emotions’, ‘affect’, ‘cognition’ and ‘feelings’ | 12 |
| 2.2 | Anatomy of the brain | 18 |
| 2.3 | Anatomy of Cerebral Cortex | 20 |
| 2.4 | Brain structures and emotion | 29 |
| 2.5 | The multidimensional affect space of James Russell | 33 |
| | | |
| 4.1 | Three Camera Configuration | 107 |
| 4.2 | Two Camera Configuration | 109 |
| | | |
| 5.1 | Error rates in recognizing the expressions in our preliminary questionnaire | 120 |
| 5.2 | The first part of the detailed questionnaire | 121 |
| 5.3 | The second part of the detailed questionnaire | 122 |
| 5.4 | Graph of the percentage to which the participants mistook the ‘angry’ emotion for other emotions | 124 |
| 5.5 | Graph of the percentage to which the ‘angry’ expression maps the equivalent emotion, based on the correct answers of the participants . | 125 |
| 5.6 | Graph of the percentage to which the ‘angry’ expression maps the equivalent emotion, based on the correct answers of the participants | 126 |
| 5.7 | Graph of the percentage to which the participants mistook the ‘bored - Sleepy’ emotion for other emotions | 127 |
| 5.8 | Graph of the percentage to which the ‘bored - Sleepy’ expression maps the equivalent emotion, based on the correct answers of the participants | 128 |

| | | |
|------|--|-----|
| 5.9 | Graph of the percentage to which the ‘bored - Sleepy’ expression maps the equivalent emotion, based on the correct answers of the participants | 129 |
| 5.10 | Graph of the percentage to which the participants mistook the ‘disgusted’ emotion for other emotions | 130 |
| 5.11 | Graph of the percentage to which the ‘disgusted’ expression maps the equivalent emotion, based on the correct answers of the participants . | 131 |
| 5.12 | Graph of the percentage to which the ‘disgusted’ expression maps the equivalent emotion, based on the correct answers of the participants | 132 |
| 5.13 | Graph of the percentage to which the participants mistook the ‘happy’ emotion for other emotions | 133 |
| 5.14 | Graph of the percentage to which the ‘happy’ expression maps the equivalent emotion, based on the correct answers of the participants . | 134 |
| 5.15 | Graph of the percentage to which the ‘happy’ expression maps the equivalent emotion, based on the correct answers of the participants | 135 |
| 5.16 | Graph of the percentage to which the participants mistook the ‘neutral’ emotion for other emotions | 136 |
| 5.17 | Graph of the percentage to which the ‘neutral’ expression maps the equivalent emotion, based on the correct answers of the participants . | 137 |
| 5.18 | Graph of the percentage to which the ‘neutral’ expression maps the equivalent emotion, based on the correct answers of the participants | 138 |
| 5.19 | Graph of the percentage to which the participants mistook the ‘sad’ emotion for other emotions | 139 |
| 5.20 | Graph of the percentage to which the ‘sad’ expression maps the equivalent emotion, based on the correct answers of the participants . | 140 |
| 5.21 | Graph of the percentage to which the ‘sad’ expression maps the equivalent emotion, based on the correct answers of the participants | 141 |

| | | |
|------|--|-----|
| 5.22 | Graph of the percentage to which the participants mistook the ‘surprised’ emotion for other emotions | 142 |
| 5.23 | Graph of the percentage to which the ‘surprised’ expression maps the equivalent emotion, based on the correct answers of the participants . | 143 |
| 5.24 | Graph of the percentage to which the ‘surprised’ expression maps the equivalent emotion, based on the correct answers of the participants | 144 |
| 5.25 | The participants’ answers regarding the level of difficulty of the facial expression recognition task | 145 |
| 5.26 | The participants’ answers regarding the most difficult emotion to recognize | 146 |
| 5.27 | The participants’ answers regarding the most difficult emotion to recognize | 147 |
| 5.28 | Error rates in recognizing the expressions in our detailed questionnaire | 148 |
| | | |
| 6.1 | The P. Sinha Template | 161 |
| 6.2 | The eigenfaces of a face image | 163 |
| 6.3 | Sample images of our Face Detection training set | 166 |
| 6.4 | Three Hidden Layer Network (Simple Demonstration) | 166 |
| 6.5 | Three Hidden Layer Network | 167 |
| 6.6 | Four Hidden Layer Network (Simple demonstration) | 168 |
| 6.7 | Four Hidden Layer Network | 169 |
| 6.8 | The face detection neural networks responses | 171 |
| 6.9 | Face Detection results for the first set of images | 174 |
| 6.10 | Face Detection results for the first set of images | 175 |
| 6.11 | Face Detection results for the second set of images | 176 |
| 6.12 | Face Detection results for the second set of images - 2 | 176 |
| 6.13 | Face Detection results for the second set of images - 3 | 177 |

| | | |
|------|---|-----|
| 6.14 | The extracted features (orange points) and the calculated distances | 184 |
| 6.15 | Typical results from our feature extraction algorithm | 186 |
| 6.16 | Facial Dimension Ratio Distribution | 190 |
| 6.17 | Mouth Dimension Ratio Distribution | 191 |
| 6.18 | Results from our 2-class system | 193 |
| 6.19 | Results from our 3-class system with the Cohn-Kanade Database | 194 |
| 6.20 | Results from our 3-class system with the Cohn-Kanade Database for training and our low quality database for testing | 195 |
| 6.21 | Results from our 3-class system with the Cohn-Kanade Database for training and our low quality, side view images for testing | 195 |
| 6.22 | The most important facial points which will help us in the extraction of the feature vector for the facial expression recognition task | 197 |
| 6.23 | The extracted features | 203 |
| 6.24 | Eye Extraction with Skin Extraction and K-means Clustering | 206 |
| 6.25 | Morphological Operations | 208 |
| 6.26 | Extracted Eye Features | 209 |
| 6.27 | Some results of the eye extraction algorithm | 209 |
| 6.28 | Probability Density of the ‘Face Size Ratio’ | 212 |
| 6.29 | Probability Density of the ‘Mouth Size Ratio’ | 213 |
| 6.30 | Probability Density of the ‘Left Eye Size Ratio’ | 214 |
| 6.31 | Probability Density of the ‘Right Eye Size Ratio’ | 215 |
| 6.32 | Probability Density of the ‘Texture of the Region of the Chin’ | 216 |
| 6.33 | Probability Density of the ‘Texture of the Region of the Forehead’ | 217 |
| 6.34 | Probability Density of the ‘Texture of the Region Between the Brows’ | 218 |
| 6.35 | The Facial Expression Neural Network Classifier | 218 |
| 7.1 | The architecture of our multimodal emotion recognition system | 241 |

7.2 Some results from combining the three modalities 242

РАНЕЕЧНО РЕПАА

Tables

| | | |
|------|---|-----|
| 2.1 | Brain areas and emotions | 28 |
| 2.2 | Summary of emotional effects in speech | 30 |
| 3.1 | Requirements for an ideal facial expression database | 54 |
| 3.2 | Review of the facial expression databases | 56 |
| 3.3 | Requirements for an ideal face detection system | 60 |
| 3.4 | Review of the face detection approaches - 1 (Before 2000) | 64 |
| 3.5 | Review of the face detection approaches - 2 (2000-2004) | 65 |
| 3.6 | Review of the face detection approaches - 3 (2005 to present) | 66 |
| 3.7 | Requirements for an ideal facial expression classification system [1] | 71 |
| 3.8 | Review of the facial expression approaches (Early Years) [1] | 74 |
| 3.9 | Categorization of the approached based on the techniques and the input media | 76 |
| 3.10 | Review of the facial expression approaches (Middle Years) [1] | 89 |
| 3.11 | Review of the facial expression approaches (Recent Years) | 102 |
| 3.12 | Performance evaluation and generalization of recent systems - 1 | 103 |
| 3.13 | Performance evaluation and generalization of recent systems - 2 | 104 |
| 4.1 | Low quality Database | 112 |
| 4.2 | Sample images of our low quality facial expression database | 113 |
| 4.3 | Low quality Database | 114 |
| 4.4 | Sample images of our facial expression database | 115 |
| 5.1 | Typical face image subsets in our questionnaire | 119 |

| | | |
|------|---|-----|
| 5.2 | Percentage to which a facial expression represents an emotion | 149 |
| 5.3 | Error rate comparison between the two parts of the questionnaire . . . | 150 |
| 5.4 | Important features for each facial expression | 151 |
| 5.5 | Mapping | 151 |
| 5.6 | Identification of the most and least important features for each expression - 1 | 153 |
| 5.7 | Identification of the most and least important features for each expression - 2 | 154 |
| 5.8 | Differences between the First and the Second Questionnaire | 155 |
| | | |
| 6.1 | The computed three clusters | 165 |
| 6.2 | Description of the two neural network classifiers | 170 |
| 6.3 | Results of the Face Detection System for the three datasets | 173 |
| 6.4 | Sample images of our facial expression database | 178 |
| 6.5 | Sample images of the fourth test set - non human faces | 179 |
| 6.6 | Demonstration of the preprocessing algorithm for low quality images | 188 |
| 6.7 | Deformations of the other six expression, compared to 'neutral' . . . | 198 |
| 6.8 | Facial action and resulting Facial Features | 199 |
| 6.9 | Facial action and resulting Facial Features - 2 | 200 |
| 6.10 | Facial action and resulting Facial Features - 3 | 201 |
| 6.11 | Results of the Facial Expression Classification System Compared to Human Classifiers | 219 |
| 6.12 | Sample images of our facial expression database | 220 |
| 6.13 | Sample images of our facial expression database | 229 |
| 6.14 | Human versus computer classifiers | 230 |
| 6.15 | Classification rates for each expression | 230 |
| 6.16 | Confusion matrix for the SVM classifier | 231 |

| | |
|--|-----|
| 6.17 Confusion matrix for the RBF classifier | 231 |
| 6.18 Confusion matrix for the KNN classifier | 232 |
| 6.19 Confusion matrix for the MLP classifier | 232 |

РАНЕЕ НЕ ПЕЧАТАЛИ

1 Introduction

Η αρχή είναι το ήμισυ του παντός. (The beginning is the most important part of the work.)

—Plato (428 BC–348 BC)

FACIAL expressions play a significant communicative role in human-to-human interaction and interpersonal relations, because they can reveal information about the affective state, cognitive activity, personality, intention and psychological state of a person and, in fact, this information may be difficult to mask. The human ability to analyze another person's facial expressions is one of the subjects of study of the scientific areas of pattern recognition and computer vision and the results of this study are applied to the design of interactive systems that support more efficient and friendlier human-computer interfaces, multimedia services, security control systems, criminology etc.

When attempting to mimic human-to-human communication, human-computer

interaction systems must determine the psychological state of a person, so that the computer can react accordingly [2]. This may be exploited in the design of advanced human-computer interfaces, which attempt to take into consideration the variations of the emotions of human users during the interaction and make the computer react accordingly. Indeed, facial expressions corresponding to the ‘neutral’, ‘happy’, ‘sad’, ‘surprised’, ‘angry’, ‘disgusted’ and ‘bored-sleepy’ psychological states arise very commonly during a typical human-computer interaction session [3, 4, 5, 6]. Thus, vision-based human-computer interactive systems with the ability to process computer user face images and extract information about the user’s identity, state and intent would prove very effective and friendly. Similar information can also be used in multimedia interactive services, security control systems or in criminology to uncover possible criminals.

Most works in automatic facial expression analysis assume that the conditions under which a facial image or an image sequence is acquired are known and controlled. Usually, the image shows the face in front view and the background is fairly uniform. In the majority of previous works, the location and the extent of the face is assumed known or easily computed. However, in real environments, this is not the case. Determining the exact location of the face in a digitized facial image is a more complex problem. First, the scale and the orientation of the face can vary from image to image. Also, even if the photos are taken from a fixed camera, there is no way to know a priori the size and the angle of the face. Thus, in order to fully automate the procedure of facial expression recognition, a task is required that consists of two steps, namely: (1) a face detection step in which the system determines whether or not there are any faces in an image and, if so, returns the location and extent of each face and, (2)

a facial expression classification step, in which the system attempts to recognize the expression formed on a detected face.

The development of such fully automated face image analysis systems, capable of detecting a face and classifying a person's facial expression with low error probabilities, is quite challenging. Some of the challenges that have to be addressed in developing such a system arise from the facts that faces are non-rigid and have a high degree of variability in size, shape, color and texture. Furthermore, variations in pose, image orientation and conditions add to the level of difficulty of the problem. Moreover, the variability in the ways people express themselves, depending on their culture, psychological state and habits, make it even more difficult to determine someone's psychological condition through his/her face image. These facts can make the analysis of the facial expressions of another person difficult and often ambiguous.

Towards building an automated facial expression classification system, we conducted a series of studies that would help us to understand how emotion perception works and set the requirements for our own facial expression classification system.

1.1 Motivation

It seems to be the case, that computer vision researchers, are still working in isolation to develop ever more sophisticated algorithms to recognize and interpret facial information without necessarily being interested in applications of their work to solve the vision problem. On the other hand, researchers in human-computer interaction are not necessarily aware of recent progress in computer vision, which may have brought the possibility of using facial information in computer interfaces closer than ever. The

question of what to do with facial information when it becomes available may actually motivate and foster ongoing research in human-computer interaction, artificial intelligence and cognitive science.

The purpose of this thesis is to identify the emotions during a typical human computer interaction process and to explore the potential of building computer interfaces which understand and respond to the richness of the information conveyed by the human face. Until recently, information has been conveyed from the computer to the user mainly via the visual channel, whereas input from the user to the computer is given through the keyboard and pointing devices. The recent emergence of multimodal interfaces might restore a better balance between our physiology and sensory/motor skills, and impact (for the better we hope), the richness of activities we find ourselves involved in. Given recent progress in user-interface primitives composed of gesture, speech, context and affect, it seems feasible to design environments which do not impose themselves as computer environments, but have a much more natural feeling associated with them.

1.2 Contribution of this thesis

In this thesis we try to justify the emotion perception from the scientific and human point of view. This is complete work in terms of analyzing the requirements, justifying and developing a facial expression recognition system to be used for the development of advanced human-computer interaction techniques. During the development of this thesis, a extensive research in the scientific areas of psychology and medicine was made. In addition with the empirical studies that we conducted on

human observers, we were able to set the requirements for our facial expression recognition system and identify the facial features that are important for the expression classification task. This work is also innovative, because, besides the ‘neutral’, ‘happiness’, ‘sadness’, ‘surprise’, ‘anger’, and ‘disgust’ emotion classes, which are common in the literature regarding the development of facial expression recognition systems, we also added the ‘boredom-sleepiness’ emotion class, which is very common during a typical human-computer interaction session. Moreover, based on our studies, there is cultural specificity in emotion perception and expression, so we developed our own face database by photo shooting Greek people forming the expressions. Finally, based on the performance evaluation of our facial expression classification system the success rates are quite significant.

1.3 Organization of this thesis

THIS thesis is organized as follows:

In Chapter 2 we study the emotion perception from the scientific - psychological and medical - point of view. This helps us to identify those brain parts that play a significant role in emotion perception and expression. Moreover, it helps us to identify the most important emotions, based on studies made by psychologists, that we need our system to recognize. Specifically, there are some ‘basic emotions’ that are considered to be similar independently of a person’s culture. Namely, these are the ‘anger’, ‘disgust’, ‘fear’, ‘happiness’, ‘sadness’ and ‘surprise’ emotion classes. At the same time, other studies show that there is a cultural specificity in emotion perception and expression.

In Chapter 3, we study previous attempts towards the development of: (1) a facial expression database, (2) a face detection system and (3) a facial expression recognition system. This, helps us to understand how to proceed in our own work in order to achieve our goal. In this chapter, we set the requirements on the performance of an ideal system. Specifically, the majority of existing methods usually succeed in several of these requirements, but there is no existing work that could match all the requirements at the same time.

Our review on previous works on facial expression databases showed that these databases were unsuitable for developing our system up to a fully operational form. This was because either the number and/or the classes of different facial expressions or the number of representative samples in existing databases were found insufficient for our purposes. This led us to the creation of two different databases, described in Chapter 4, namely: (1) A database of low quality images: this database consists of many subjects, depicting many expressions, but the image quality is very low as we used web cameras to acquire the data and, (2) A database of high quality images: this database consists of many subjects, depicting the expressions recognized by our system, and the image quality is quite high as we used digital cameras to acquire the data.

Also, in our study we tried to understand the emotion perception from the human point of view. Our findings are summarized in Chapter 5, in which we present two empirical studies that we conducted on the process of facial expression classification by humans. These studies form the basis of the facial expression classification module of our system. Specifically, we identify classification features, analyze algorithms for their extraction, and quantify their discrimination power. Moreover, we

identify the emotions that are present during a typical human-computer interaction, so facial expressions corresponding to the ‘neutral’, ‘happy’, ‘sad’, ‘surprised’, ‘angry’, ‘disgusted’ and ‘bored-sleepy’ psychological states. Finally, the results of the two empirical studies indicate that cultural exposure increases the chances of correct recognition of facial expressions.

In Chapter 6, we present our system, which consists of two modules: (1) a face detection subsystem and, (2) a facial expression recognition subsystem. Our face detection subsystem is based on neural network-based classifiers. For our facial expression recognition subsystem, we considered neural network-based and other classifiers, but we concluded that Support Vector Machine-based Classifiers demonstrated better results. Performance evaluations and test results are included in this chapter for both subsystems.

Finally, we summarize, draw conclusions and point to future work in Chapter 7, where we also discuss the possibility of using other modalities towards emotion perception, such as keyboard stroke patterns and audio-lingual information.

2

Previous Related Psychological Studies on Emotion Perception

A man should not strive to eliminate his complexes but to get into accord with them: they are legitimately what directs his conduct in the world.

—Sigmund Freud (1856–1939)

IN order to develop ways that effectively allow emotions on computers and aid our interactions with them, we need to understand first what emotions are and what exactly constitutes emotional intelligence. This section starts by providing an overview of what is currently known about emotions, including what causes them, how humans express them, and their influence on the way we feel and behave. An important part in this section is devoted to how people express their emotions using their face and how psychologists interpret and understand these emotions. Some of

the topics discussed in this section can be summarized in the following questions:

- What constitutes an emotion?
- Is there a difference between emotion, feelings and affect?
- How do people feel an emotion? Is there a proven scientific approach?
- Is there a similarity on the way people express themselves, regardless of their culture?
- Are there some ‘basic emotions’ that are similar regardless of a person’s culture?
- Can we identify these emotions?

The answers to some of these previous questions still remain under debate among the psychologists. Their beliefs on these matters will be discussed and further analyzed in this section.

2.1 Emotion vs affect vs feelings

FACIAL expressions are considered as a very important means of conveying information about a person’s affective or emotional state. Before considering facial expression of emotion, the nature of emotion itself must be examined first. There is a distinction between affect and emotion, based on the studies of the American Psychiatric Association. *Affect* is defined as ‘*a pattern of observable behaviors that is the expression of a subjectively experienced feeling state or emotion*’ [7]. The subject of emotion, in turn, has generated a variety of definitions and associated theories. Based on the studies of Iverson, Kupfermann and Kandel [8], there is a

distinction between emotions and feelings. Emotion ‘refers to the bodily state of a person’ and feeling ‘refers to the conscious experience of the bodily state’. William James proposed that ‘the bodily changes follow directly the perception of the exciting fact, and that our feeling of the same changes as they occur is the emotion’ (cited in [9]). Thus, in what has come to be known as the James-Lange theory of emotion, ‘cognition is secondary to the physiological expression of emotion’.

Accordingly, emotions are considered to be cognitive responses to information from the periphery. In a critique of the James-Lange theory, W.B. Cannon [10] proposed that bodily changes follow from cognitive processes. His assertion was based on the following observations: (a) removal of the cerebral cortex in laboratory animals does not impair emotional behavior, (b) the same visceral changes occur in different emotional states, (c) the viscera are relatively insensitive, (d) visceral changes are typically too slow to generate emotions and (e) induction of the physical changes typical of strong emotions does not result in the experience of the simulated emotion. However, these observations and conclusions have recently been drawn into question [11]. Current opinion represents a synthesis of the two theories in which emotions are viewed as the outcome of a dynamic, ongoing interaction of peripheral and central factors [8].

The nature of the four notions: ‘emotions’, ‘affect’, ‘cognition’ and ‘feelings’ is summarized in the following Figure 2.1

Our knowledge can lead to the following assumptions regarding emotions:

- emotions are fast
- triggered automatically

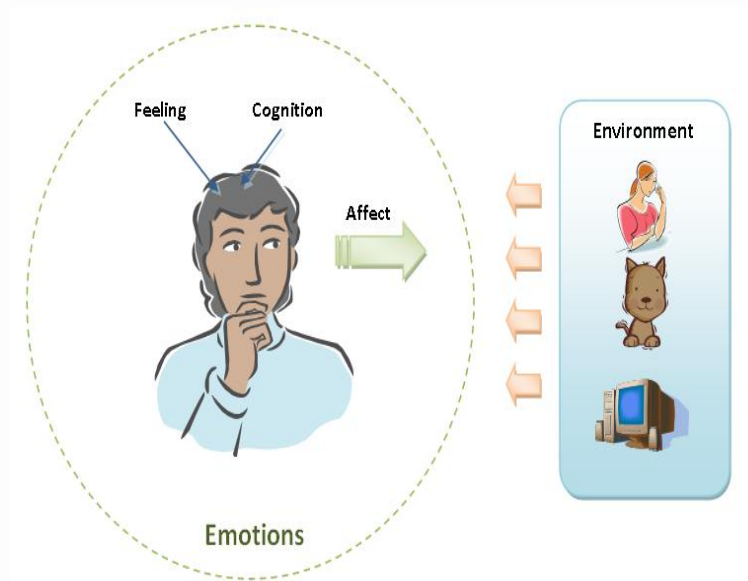


Figure 2.1 : The nature of ‘emotions’, ‘affect’, ‘cognition’ and ‘feelings’

- triggered by the environment
- moderated at least partially by cognitive processes
- transitory

2.2 Emotions and culture

EMOTION theorists have debated for centuries about what emotions are and what their primary function in human life is. This debate is far from over and there is currently no universally agreed upon definition of emotions. However, many scholars would at least agree that we experience different types of emotions in our everyday lives. Most psychologists agree that there is a difference on how people

express themselves depending on their culture. On the other hand, the majority agree that there is also a similarity in these expressions because of the theory of evolution that strengthens the connection between all these cultures. First studies on this subject were conducted by Charles Darwin [12, 13] and mostly followed by Paul Ekman [13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29] and are summarized below.

2.2.1 Basic Emotions

Although efforts were made to understand emotion and some early studies suggest the idea of universal emotions, for much of the previous century, emotion scholars generally subscribed to a cultural theory of emotion, where emotions were believed to be culturally specific learnt behaviors that could only be experienced through observing other people expressing such emotions. As mentioned above, the historical roots of basic emotion theory originate from Charles Darwin, who as a part of his evolutionary theory suggested that the emotional expressions of man were descendants from other animals [12, 13]. Darwin not only made observations on the behavior of animals but also set to study the question of whether some emotions were universal to all men. Although the idea of universal basic emotions had been mentioned already many centuries before Darwin in the writings of philosophers such as Descartes, Hobbes and Spinoza [30] and influential facial expression studies had been conducted by other 19th century scientists such as Guillaume Duchenne de Boulogne[31], Darwin appears to have conducted the first scientific evaluation studies on the recognition of emotions from faces. Darwin studied which emotions were recognized consistently from photographs of representative emotional facial expressions in England [9] and

made the first attempts to evaluate the universality of emotions by interviewing his fellow countrymen living abroad on the expression of emotions in other cultures [13]. The method of asking subjects to judge emotions from certain facial expressions has remained a part of contemporary research methodology. According to a critical review [32], different basic emotion sets ranging from two to eighteen basic emotions have been proposed by different investigators; however, most of them agree at least on the emotions of *anger*, *fear*, *happiness* and *sadness*.

However, Paul Ekman [13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29] discovered that some emotions are not necessarily learnt, as previously believed, but are in fact innate and shared across all cultures. In his study, Ekman traveled to a preliterate culture (the Fore, in New Guinea) to ensure that the people there had not been exposed to Western media and had not learned the emotional expressions of the Westerners. The subjects were told a number of stories, then asked to choose from a set of photographs of Americans expressing different emotions, the one which most closely matched the story. When tested, the Fore pointed to the same expressions that Westerners linked to the story. For further clarification, some Fore people were videotaped displaying facial expressions appropriate to each of the stories. After returning home, the experiment was completed in reverse by asking Americans to link the Fore faces to the different stories. The judgements of both the Fore people and the Americans again matched. The studies by Paul Ekman led to the identification of the '*basic emotions*' which are universally common regardless of the culture of the subject. Although researchers often disagree about how many basic emotions there are, the most commonly accepted emotions which can be classified as 'basic emotions' are: '*anger*', '*disgust*', '*fear*', '*happiness*', '*sadness*' and '*surprise*'.

Later studies by Paul Ekman [20], added '*contempt*' to the 'basic emotions' in addition to the six original ones, and 'contempt' has received some support from a study conducted in ten countries; however, no studies have been made in isolated groups. Furthermore, subjects have failed to recognize contempt from supposedly characteristic facial expressions in more recent studies [33, 34], indicating that contempt *should not be considered a basic emotion*.

2.2.2 Culturally Specific Expressions of Emotions

Besides the 'basic emotions', studies have shown that there are also cultural variations in the way in which humans express emotion. These studies have shown that the emotions can vary: (1) in terms of the expression of emotion, but also (2) in terms of the intensity of the expressed emotion.

The first assumption suggests that some emotions are culturally specific. This difference can be interpreted in many terms:

- some emotions maybe present only in some cultures: This example is demonstrated by J. Ledoux [35]. In his book, he reports about an emotion that is experienced by the Gururumba people of New Guinea that is not believed to be experienced by people of other cultures. This is known as the state of 'being a wild pig' and people who experience this state can become aggressive and often start looting, but rarely is anyone actually hurt or anything of importance stolen. This state is considered as normal among the Gururumba and as a way of relieving stress and maintaining mental health across the community.
- different ways of expression depending on the culture: For example, in many cases,

it is stated that Japanese people tend to close their mouth and widen their eyes when they feel 'surprised'. This is contrary to other cultures that tend to open their mouth widely when feeling the same emotion.

The second assumption assumes that although the emotion maybe present, the intense of the expression maybe different, depending on the culture. This, was demonstrated by another experiment conducted by Paul Ekman [14] on American and Japanese people and further discussed by Matsumoto [36]. In this experiment, both American and Japanese men were videotaped whilst watching some video clips. In these clips, there were demonstrated scenes that would provoke the corresponding emotion, so there were neutral or pleasant events (such as a canoe trip) or less pleasant events (for example, nasal surgery). There were two showings of the video clips: one where subjects watched the clips on their own and another where subjects watched the clips with an interviewer present. When subjects watched the clips in private, similar expressions were noted in both American and Japanese subjects. However, when the interviewer was present, Japanese subjects smiled more and showed less disgust than the American subjects. When the videotapes were watched back in slow motion the researchers noticed that when the interviewer was present, Japanese subjects actually started to make the same expressions of disgust as the Americans did, but they were able to mask these expressions very quickly afterwards. Therefore, it appeared that the American and Japanese participants did actually experience the same basic emotions as these were automatic responses hardwired into their brains. It was only a few hundred milliseconds later, that the Japanese subjects could apply their learnt cultural display rules and override the automatic response.

The cultural specificity of the emotions is further strengthened by studies of Izard

[37] and Elfenbein and Ambady [38].

2.2.3 Higher Cognitive Emotions

Besides the ‘basic emotions’ and the culturally-specific emotions, Paul E. Griffiths [39] believes that, in addition, there are also ‘higher cognitive emotions’. There is a similarity between those emotions and the basic emotions. This similarity lies on the fact that they are universal, but there are also variations on the way that they are expressed and experienced by different cultures. Moreover, usually, there is also no single facial expression associated with them. Higher cognitive emotions are considered to take longer than basic emotions to both develop and pass away. As an example, we can consider romantic love. This emotion usually develops gradually in people over a period of weeks and months, while ‘surprise’ (a basic emotion) is typically a very quick reaction to an event. ‘Surprise’ can also be recognized by a couple of facial expressions associated with it, while there are no universal facial expression for love. It is suggested that emotions such as *love, jealousy, pride, embarrassment and guilt* should be called ‘*higher cognitive emotions*’, because these emotions typically require more processing in the cortex of the brain. This essentially means that these emotions can be influenced more by cognitive thought processes, while basic emotions are more reactive and spontaneous in nature.

2.3 Neurobiology and Emotion Expression

ANOTHER fact that can further manifest the statement that there are universally known emotions, is the assumption that a number of brain regions are directly involved in the perception of facial expressions and, generally, emotion un-

derstanding and feeling. These include: (1) **the frontal cortex**, (2) **the superior temporal sulcus**, and (3) **the amygdala**.

The anatomy of the brain is demonstrated in Figure 2.2, where we can observe the aforementioned parts of the brain.

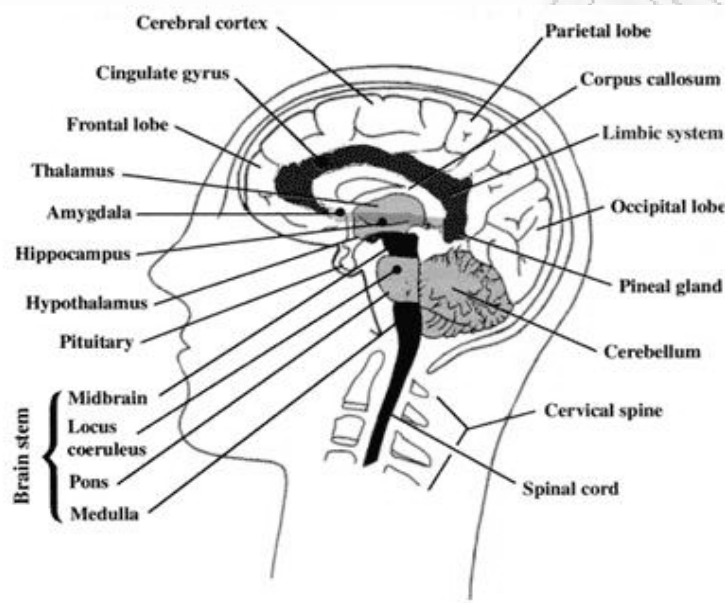


Figure 2.2 : Anatomy of the brain

2.3.1 Cerebral Cortex

The cerebral cortex is a structure within the brain that plays a key role in memory, attention, perceptual awareness, thought, language, and consciousness. In dead, preserved brains, the outermost layer of the cerebrum has a grey color, hence the name 'grey matter'. Grey matter is formed by neurons and their unmyelinated fibers, whereas the white matter below the grey matter of the cortex is formed predomi-

nantly by myelinated axons interconnecting different regions of the central nervous system. The human cerebral cortex is from 2 to 4 mm thick.

The surface of the cerebral cortex is folded in large mammals, so that more than two-thirds of the cortical surface is buried in the grooves, called 'sulci.' The phylogenetically most recent part of the cerebral cortex, the neocortex, also called isocortex, is differentiated into six horizontal layers; the more ancient part of the cerebral cortex, the hippocampus (also called archicortex), has at most three cellular layers, and is divided into subfields. Relative variations in thickness or cell type (among other parameters) allow us to distinguish between different neocortical architectonic fields. The geometry of at least some of these fields seems to be related to the anatomy of the cortical folds, and, for example, layers in the upper part of the cortical grooves (called gyri) seem to be more clearly differentiated than in its deeper parts

The cerebral cortex is divided into lobes, each of which has a specific function. For example, there are specific areas involved in vision, hearing, touch, movement, and smell. Other areas are critical for thinking and reasoning. Although many functions, such as touch, are found in both the right and left cerebral hemispheres, some functions are found in only one cerebral hemisphere. For example, in most people, language abilities are found in the left hemisphere. The lobes can be summarized as following:

- Parietal Lobe: involved in the reception and processing of sensory information from the body.
- Frontal Lobe: involved with decision-making, problem solving, and planning.
- Occipital Lobe: involved with vision.

- Temporal Lobe: involved with memory, emotion, hearing, and language.

The lobes of the cerebral cortex and their role are summarized in Figure 2.3

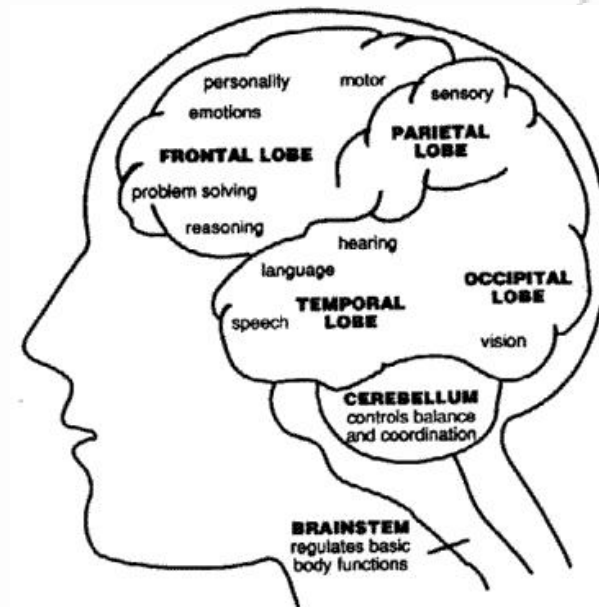


Figure 2.3 : Anatomy of Cerebral Cortex

In summary, the cerebral cortex is responsible for sensing and interpreting input from various sources and maintaining cognitive function. Sensory functions interpreted by the cerebral cortex include hearing, touch, and vision. Cognitive functions include thinking, perceiving, and understanding language. Various studies have showed that cerebral cortex plays a key role in expressing and understanding emotions by the humans. Specifically, Nakamura et al. [40], measured regional cerebral blood flow (rCBF) using positron emission tomography (PET) to determine which brain regions are involved in the assessment of facial emotion. They presented normal subjects with photographs of facial expressions. The right inferior frontal cortex showed

significant activation during the assessment of facial emotion in comparison with the other two tests. The authors suggest that the right inferior frontal cortex is involved in the processing of both visual and auditory emotional communicative information. This area is also involved in metaphor comprehension, phonological working memory, and face matching. This might indicate that the assessment of facial emotion involves the matching of perceived facial gestures with templates or prototypes in the brain.

In a study [41], Harmer, Thilo, Rothwell and Goodwin suggest that the recognition of different emotional states involves at least partly separable neural circuits. They assessed the discrimination of both anger and happiness in healthy subjects receiving transcranial magnetic stimulation (TMS) over the medial-frontal cortex or over a control site (mid-line parietal cortex). The experimental task utilized sequences of face images morphed between two prototypes, e.g. angry and neutral, in 10% increments. Subjects were presented with two forced-choice tasks which required them to differentiate between happy/neutral faces and angry/neutral faces. Recognition thresholds were established prior to the application of TMS. Recognition thresholds were defined as the level of emotional intensity at which 75% of the facial expressions were correctly identified. The recognition threshold for happiness was at a morph increment of 40% while for anger it was at an increment of 70%. Application of TMS to the medial frontal cortex was found to impair the processing of angry facial expressions but not happy facial expressions.

‘Disgust’ is another emotion that is considered to derive from the cerebral cortex. Specifically, anterior Insula and adjacent frontal operculum, play an important role in the perception of disgust. Studies from Gallese et al. [42] and Jabbi et al. [43] have demonstrated this assumption. During an experiment by Jabbi et al. [43], the

same subjects were scanned while they (a) view actors taste the content of a cup and look disgusted (b) tasted unpleasant bitter liquids to induce disgust, and (c) read and imagine scenarios involving disgust and their neutral counterparts. They found voxels in the anterior Insula and adjacent frontal operculum to be involved in all three modalities of disgust, suggesting that simulation in the context of social perception and mental imagery of disgust share a common neural substrates.

2.3.2 Amygdala

The amygdalae (Latin, also corpus amygdaloideum, singular amygdala) are almond-shaped groups of neurons located deep within the medial temporal lobes of the brain in complex vertebrates, including humans. Shown in research to perform a primary role in the processing and memory of emotional reactions, the amygdalae are considered part of the limbic system.

The regions described as amygdalae encompass several nuclei with distinct functional traits. Among these nuclei are the basolateral complex, the centromedial nucleus and the cortical nucleus. The basolateral complex can be further subdivided into the lateral, the basal and the accessory basal nuclei. Anatomically, the amygdala and more particularly, its centromedial nucleus, may be considered as a part of the basal ganglia.

In humans, the amygdalae perform primary roles in the formation and storage of memories associated with emotional events. Research indicates that, during fear conditioning, sensory stimuli reach the basolateral complexes of the amygdalae, particularly the lateral nuclei, where they form associations with memories of the stimuli. The association between stimuli and the aversive events they predict may be medi-

ated by long-term potentiation, a lingering potential for affected synapses to react more readily.

Memories of emotional experiences imprinted in reactions of synapses in the lateral nuclei elicit fear behavior through connections with the central nucleus of the amygdalae. The central nuclei are involved in the genesis of many fear responses, including freezing (immobility), tachycardia (rapid heartbeat), increased respiration, and stress-hormone release. Damage to the amygdalae impairs both the acquisition and expression of Pavlovian fear conditioning, a form of classical conditioning of emotional responses.

The amygdala is also involved in appetitive (positive) conditioning. It seems that distinct neurons respond to positive and negative stimuli, but there is no clustering of these distinct neurons into clear anatomical nuclei.

Amygdala is considered to play an important role in the perception of the emotion of 'fear', regardless of the different ways of expressing it (e.g masked fear, low spatial frequency fear, broad spatial frequency fear, conditioned response (CR) to fear. Many studies [44, 45, 46, 47] have proved this assumption. In the presence of 'fear', the subject demonstrated significantly increased amygdalar response, particularly in the left hemisphere, to images of fearful expressions. Decreased response was associated with viewing of happy expressions. Amygdalar response can occur without explicit knowledge of the stimulus [46].

2.3.3 Superior Temporal Sulcus

The superior temporal sulcus is thought to be used by humans in making simple actions, or watching other people make actions. Several research areas claim the su-

terior temporal sulcus (STS) as the host brain region for their particular behavior of interest. Some see it as one of the core structures for the theory of mind *. For others, it is the main region for audiovisual integration. It plays an important role in biological motion perception, but is also claimed to be essential for speech processing and processing of faces [49]. Dynamic changes in facial displays contribute to face recognition abilities, judgment of affect, and identity. LaBar, Crupain, Voyvodic and McCarthy [50] examined the perception of negative affect using dynamically varying facial stimuli. Prototypical expressions of fear and anger were morphed with neutral expressions resulting in a sequence of static images that, when displayed sequentially, gave the impression of dynamic changes in real-time. Identity morphs were also created across pairs of faces with neutral expressions. An fMRI study was performed to compare brain activation to static and dynamic facial displays. Amygdalar activity was enhanced by dynamic displays of facial expressions, particularly fearful expressions, relative to static displays.

Specificity for fear over anger was found in the dynamic morphs but not static images. Amygdalar response was also heightened for dynamic changes in facial identity. Rapid changes in identity are rarely encountered in nature or daily life. Such heightened activity may be due to the rapid change being perceived as a threat. The superior temporal sulcus (STS) showed enhanced activation to dynamic changes in facial expression relative to dynamic changes in identity. This supports the notion that the STS distinguishes biologically plausible motion from biologically implausible or nonbiological motion. Area MT+ was activated by all dynamic stimuli, showing no preference for emotion or identity. The anteromedial fusiform gyrus showed acti-

*The term 'theory of mind' refers to the influence of emotion and mental state [48]

vation for dynamic expression changes while the posterolateral inferotemporal cortex was activated for dynamic identity changes.

2.3.4 Implicit and Explicit Perception of Emotion

The ‘implicit-explicit’ distinction was entered in the study of memory when Graf and Schaschter [51] wrote a paper about ‘implicit’ and ‘explicit’ measurements of the memory. In the last years, these terms have been added to the study of emotion and facial expression perception [52]. As ‘implicit’ emotion is considered the unconscious emotion, whereas ‘explicit’ is usually referred to the conscious emotion. Most of what the brain does is unconscious, but attention both amplifies and prolongs activation, which allows processing at one site to affect processing at other sites, forming a network of activation that can give rise to the experience of consciousness. Studies by Gorno-Tempini et al. [53], Winston et al. [54] and, recently, by Scheuerecker et al. [55] deal with fMRI in order to examine the brain activation in the occurrence of ‘implicit’ and ‘explicit’ emotions.

Specifically, studies by Gorno-Tempini et al. [53] investigated the neural correlates of incidental and explicit processing of the emotional content of faces expressing either disgust or happiness. Subjects were examined while they were viewing neutral, disgusted, or happy faces. The incidental task required subjects to decide about face gender, the explicit task to decide about face expression. Results showed that the left inferior frontal cortex and the bilateral occipito-temporal junction responded equally to all face conditions. Several cortical and subcortical regions were modulated by task type, and by facial expression. Right neostriatum and left amygdala were activated when subjects made explicit judgements of disgust, bilateral orbitofrontal

cortex when they made judgement of happiness, and right frontal and insular cortex when they made judgements about any emotion.

Winston et al. [54], aimed to investigate whether distinct neural systems process distinct emotions. The automaticity of emotion perception was also investigated. Subjects were presented with morphed faces exhibiting various levels of disgust, fear, happiness, and sadness. Subjects were asked either to make explicit emotional judgements or to determine the gender of the displayed face. The amygdala and fusiform cortex were activated for high-intensity expression of all four emotions in both the direct and incidental tasks. The right STS was active for high-intensity expression in the explicit task. No differences were found for the different emotional expressions. Ventromedial prefrontal and somatosensory cortices were active during explicit tasks. This activation, in particular, is thought to link emotion perception with representations of somatic states. The authors also conclude that the amygdala contributes to the processing of a range of emotions in both direct and incidental processing.

Finally, Scheuerecker et al. [55] tried again to find specific regions for single types of emotions and for the cognitive demands of expression processing. Specifically, they investigated the neural correlates of incidental and explicit processing of the emotional content of faces expressing either disgust or happiness. Subjects were examined while they were viewing neutral, disgusted, or happy faces. The incidental task required subjects to decide about face gender, the explicit task to decide about face expression. In the control task, subjects were requested to detect a white square in a grayscale mosaic stimulus. Results showed that the left inferior frontal cortex and the bilateral occipito-temporal junction responded equally to all face conditions. Several cortical and subcortical regions were modulated by task type, and by facial expression. Right

neostriatum and left amygdala were activated when subjects made explicit judgements of disgust, bilateral orbitofrontal cortex when they made judgement of happiness, and right frontal and insular cortex when they made judgements about any emotion.

2.3.5 Section Summary

The results of the aforementioned studies are presented in Table 2.1. There are a number of key observations that can be made from these studies. First, the structures involved in the perception of facial expressions include subcortical structures, primary sensory areas such as the visual and somatosensory cortices, higher level visual processing areas including the fusiform cortex and the STS, premotor areas involved in motion planning, and prefrontal areas involved in more abstract processes such as analysis and discrimination. As with emotion processing in general, these structures range from phylogenetically older structures to those more recently evolved. Second, perception of facial expressions invokes conscious and unconscious processes in parallel. Third, understanding of emotion in others appears to involve an unconscious imitative process that simulates the observed emotional state. Finally, there appears to be a partial differentiation between brain areas relative to particular emotions.

2.4 Expression of Emotion

HUMANS can express emotion in a variety of ways, the primary ones being written language, facial expressions, speech, and body language (such as posture and gait).

Table 2.1 : Brain areas and emotions

| Brain Area | Activation for: | Authors: |
|--|--|--------------------|
| Right Inferior Frontal Cortex | Facial expressions | Nakamura [40] |
| Medial Frontal Cortex | Anger | Harmer [41] |
| Anterior Insular Cortex | Disgust | Gallese [56] |
| Amygdala | Masked fear | Morris [44] |
| Amygdala | Disgust | Whalen [46] |
| Amygdala | Low spatial frequency fear, Broad spatial frequency fear | Vuillemier [47] |
| Amygdala | Conditioned response (CR) to fear | Morris [44] |
| Amygdala | Gender & Expression | Gorno-Tempini [53] |
| Left Inf Frontal/ Occipito-Temporal | Disgust | Gorno-Tempini [53] |
| Rt Neostriatum & Lt Amygdala | Disgust | Gorno-Tempini [53] |
| Rt Frontal & Insular Cortices | Happiness & Disgust | Gorno-Tempini [53] |
| Amygdala / Fusiform Cortex | Gender/Intense expression | Winston [54] |
| Rt STS, Ventromedial pre-frontal, Somatosensory Cortex | Intense expression | Winston [54] |

2.4.1 Written Language

Written language is a powerful medium for expressing emotion. People often express their emotions through stories, poetry and personal letters. People can literally state

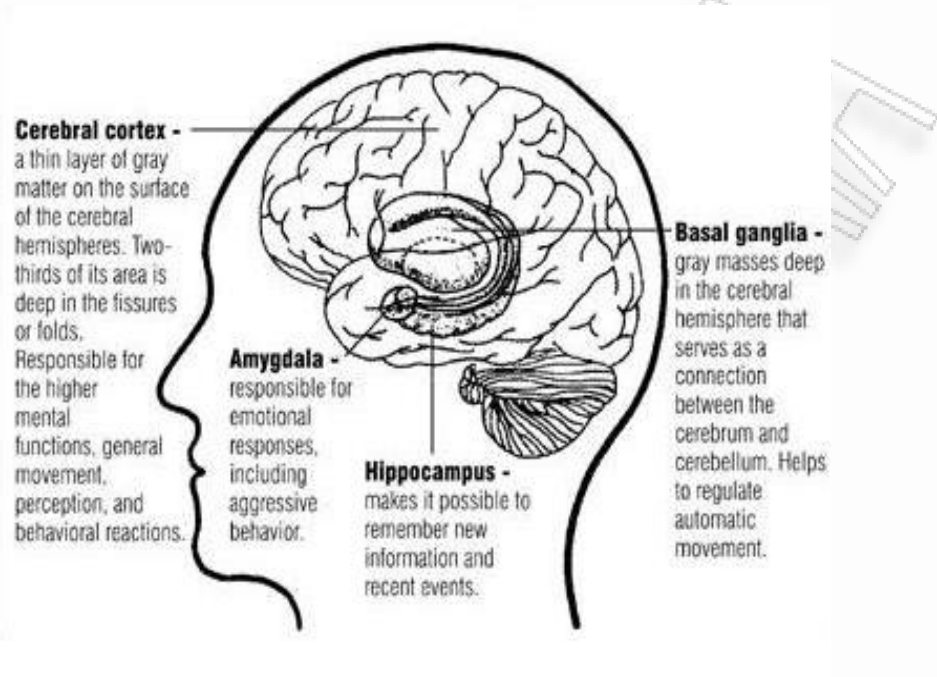


Figure 2.4 : Brain structures and emotion

how they are feeling using emotive words such as “happy”, “sad”, or “ecstatic”. The colour, size, and shape of words can also be manipulated to add emotional emphasis to content (for instance, by animating text [57]. Symbols such as emoticons • for example, ‘:-)’ or ‘:-(’ • can also be used to convey emotion, and are particularly popular within domains where emotional information is lacking, such as email, instant messaging or text messaging.

2.4.2 Speech

Another powerful method for communicating and expressing emotion is through speech. In some scenarios, it is the only channel available for communication (for

example, telephone conversations). Speech can also provide other information about a speaker such as their identity, age and gender. People can also use speech to simply communicate the emotions they are experiencing. Pitch (level, range and variability), tempo and loudness are considered the most influential parameters for expressing emotion through speech [58]. McNair et al. [59] have defined the general characteristics of a range of basic emotions (Table 2.2).

Table 2.2 : Summary of emotional effects in speech

| | Anger | Happiness | Sadness | Fear | Disgust |
|---------------|---------------------|------------------|-------------------|-------------------|------------------|
| Speech rate | slightly faster | faster or slower | slightly slower | much faster | very much slower |
| Pitch average | very much higher | much higher | slightly lower | very much higher | very much lower |
| Pitch range | much wider | much wider | slightly narrower | much wider | slightly wider |
| Intensity | higher | higher | lower | normal | lower |
| Voice quality | breathy, chest tone | breathy, blaring | resonant | irregular voicing | wide, downward |
| Articulation | tense | normal | slurring | precise | normal |

2.4.3 Facial Expressions

Facial expressions are one of the primary ways in which we can detect emotions in others and is the main research in this thesis. The importance and the ways of

expressing and understanding emotions through facial expressions is further analyzed in Section 2.5.

2.4.4 Gestures and Body Language

An overview and explanation of the meaning of different head, hand and body movements is provided in [52]. For example, a vertical (up and down) “headnod” often displays agreement or comprehension while listening. Clenched fists can signal an aroused emotional state, such as fear, anger, or excitement (for instance celebrating your team scoring at a sports event). Another example is arm-crossing, which is seen as a self-comforting and stimulating posture that is unconsciously used to ease anxiety and social stress [52].

2.5 Facial Expression of Emotion

THE question of how to best characterize perception of facial expressions has clearly become an important concern for many researchers in affective computing. Ironically somehow, this growing interest is coming at a time when the established knowledge on human facial affect is being strongly challenged in the basic research literature. In particular, recent studies have thrown suspicion on a large body of long-accepted data, even on studies previously presented by the same people.

In the past, two main studies regarding facial expression perception have appeared in the literature. The first study is the classic research by psychologist Paul Ekman and colleagues [29, 27, 26, 24, 21, 22, 15] in the early 1960s, which resulted in the identification of a small number of so-called ‘basic’ emotions: anger, disgust, fear, happiness, sadness and surprise (contempt was added only recently). In Ekman’s

theory, the basic emotions were considered to be the building blocks of more complex feeling states [15]. In newer studies, however, Eckman is sceptical about the possibility of two basic emotions occurring simultaneously [21, 22]. Following these studies, Ekman and Friesen [60] developed the, so-called, ‘facial action coding system (FACS)’, which quantifies facial movement in terms of component muscle actions. Recently automated, the FACS remains one of the most comprehensive and commonly accepted methods for measuring emotion from the visual observation of faces.

The second study by psychologist James Russell and colleagues [33, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71] challenges strongly the classic data, largely on methodological grounds. Russell argues that emotion in general (and facial expression of emotion in particular) can be best characterized in terms of a multidimensional affect space, rather than in terms of discrete emotion categories. More specifically, Russell claims that two dimensions, namely ‘pleasure’ and ‘arousal’, are sufficient to characterize facial affect space. The multidimensional affect space of James Russell is depicted in Figure 2.5

Despite the fact that divergent studies have appeared in the literature, most scientists agree that:

- Humans experience emotions in subjective ways.
- The ‘basic emotions’ deal with fundamental life tasks.
- The ‘basic emotions’ mostly occur during interpersonal relationships, but this does not exclude the possibility of their occurring in the absence of others.
- Facial expressions are important in revealing emotions and informing other people about a person’s emotional state. Indeed, studies have shown that people

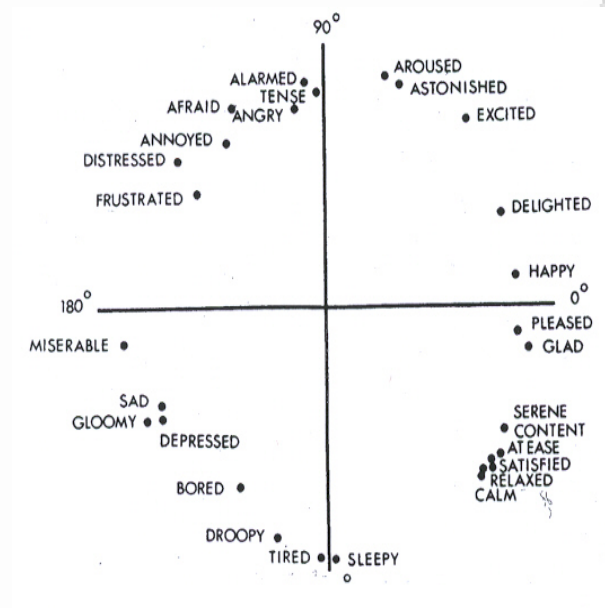


Figure 2.5 : The multidimensional affect space of James Russell [66, 67, 68, 69, 70, 71]

with congenital (Mobius Syndrome) or other (e.g. from a stroke) facial paralysis report great difficulty in maintaining and developing interpersonal relationships.

- Each time an emotion occurs, a signal will not necessarily be present. Emotions may occur without any evident signal, because humans are, to a very large extent, capable of suppressing such signals. Also, a threshold may need to be exceeded to bring about an expressive signal and this threshold may vary across individuals.
- Usually, emotions are influenced by two factors, namely social learning and evolution. Thus, similarities across different cultures arise in the way emotions are expressed because of past evolution of the human species, but differences also arise which are due to culture and social learning.

- Facial expressions are emotional signals that result into movements of facial skin and connective tissue caused by the contraction of one or more of the forty four bilaterally symmetrical facial muscles. These striated[†] muscles fall into two groups:
 - four of these muscles, innervated by the trigeminal (5th cranial) nerve, are attached to and move skeletal structures (e.g., the jaw) in mastication
 - forty of these muscles, innervated by the facial (7th cranial) nerve, are attached to bone, facial skin, or fascia and do not operate directly by moving skeletal structures but rather arrange facial features in meaningful configurations.

2.5.1 Previous Attempts to Facial Emotion Quantification and Classification

In the recent years, research work has been done that attempted at quantifying facial emotion and automating facial expression classification. Neurologists have made progress in demonstrating that emotion is one of the most important factors in a

[†]Striated muscle is a form of fibers that are combined into parallel fibers. More specifically, it can refer to:

- Skeletal muscle
- Cardiac muscle (cardiac referring to the heart).

In practice, the term ‘striated muscle’ is sometimes used to refer exclusively to skeletal muscle when distinguishing it from smooth muscle. However, different medical dictionaries report different usages of the terms. Cardiac muscle is a different type of muscle, but has almost the same structure as skeletal muscle.

person's decision making process. Recent studies have shown that automated emotion recognition can play an important role in the development of more effective and friendlier methods in multimedia interactive services and human-computer interaction systems, because how people feel may play an important role on their cognitive processes as well [72]. Picard points out that one of the major challenges in affective computers is to try to improve the accuracy of recognising people's emotions [73].

Ekman and Friesen first defined a set of universal rules to 'manage the appearance of particular emotions in particular situations' [21, 22, 19, 60, 15]. Unrestrained expressions of anger or grief are strongly discouraged in most cultures and may be replaced by an attempted smile rather than a neutral expression; detecting those emotions depends on recognizing signs other than the universally recognized archetypal expressions. D. Goren et al. made an attempt to measure emotion intensity based on a metric of deviation from a neutral face [74]. Reeves and Nass [75],[76] have shown that people's interactions with computers, TV and similar machines/media are fundamentally social and natural, just like interactions in real life. Studies have also shown that the facial expression recognition process is configural and entails an obligatory computation of gaze direction [77]. Picard, in her work in the area of affective computing, states that 'emotions play an essential role in rational decision-making, perception, learning, and a variety of other cognitive functions' [78], [73]. De Silva et al. performed an empirical study and reported results on human subjects' ability to recognize emotions [79]. They showed video clips of facial expressions to human subjects, while they had them listen to corresponding synchronised emotional speech clips in languages unfamiliar to them, namely spanish and sinhala. Then, they compared human recognition results in three tests: video only, audio only, and combined

audio and video. Finally, M. Pantic et al. [80] performed a survey of past work in solving emotion recognition problems by computer and provided a set of recommendations for developing the first part of an intelligent multimodal human-computer interaction interface.

2.5.2 Face and Facial Expressions: Their Role

Expression as Biological Adaptation

One of the most prevalent settings in which the evaluative mechanism of emotion is employed is in interpersonal relations. Each person must monitor his or her own emotional state (and expression thereof) as well as simultaneously evaluating indicators of the other persons' emotional state and intentions. Facial characteristics play a key part in the perception and expression of emotion and intent. As such, facial expression represents an essential element in social interactions. Schmidt and Cohn [81] consider human facial expressions to be biological adaptations that confer a fitness advantage in the highly social environment common to humans. Furthermore, they suggest that perception of certain signals at low levels, such as anger, is likely to confer a greater fitness advantage over those less able to detect signs of threat in their environment. Valentine [82] makes the distinction between identification and recognition in that identification requires a judgment pertaining to a specific stimulus while recognition requires only a judgment that the face has been seen before. He posits the capability to reliably distinguish friend from foe would confer an evolutionary advantage over simply knowing that a face has been seen before.

Theory of Mind

The prediction and interpretation of other people's behavior is facilitated by the assumption that others' minds are like our own. The inference of emotion and mental state is known as the theory of mind [48]. The role of facial components in this process was investigated by Baron-Cohen, Wheelwright and Jolliffe [83]. The contributions made by the eyes, the mouth, and the whole face were assessed. Adult subjects were shown images of faces depicting 10 basic emotions and 10 complex mental states. For each state, the whole face, the eyes alone, and the mouth alone were displayed. Subjects were asked to make a forced choice response. Results show considerable agreement in ascribing mental states to facial expressions. The whole face was found to be more informative for the basic emotions. However, for the complex mental states, viewing the eyes alone was as effective as seeing the whole face and both were significantly more effective than seeing the mouth alone. Adults with autism or Asperger's syndrome have shown deficits in the ability to infer mental and emotional states of others. Subjects with autism or Asperger's syndrome were tested in the same manner as the first experiment. These subjects demonstrated a significant deficit in ascribing complex mental states. This deficit was greatest for the eyes-alone condition.

Personality Attribution

Personality characteristics inferred from the face are frequently used to form an opinion about a person. Conversely, information about a person may affect the perception of that person's face. Hassin and Trope [84] refer to the first process as reading-from-

faces (RFF) and the second process as reading-into-faces (RIF). In an examination of these processes, it was found that facial information changes the interpretation of relevant information, with the reliance on facial cues increasing as other sources of information become more ambiguous. Physiognomic information is utilized, often automatically, in making assessments of others. The use of such information results in an over-confidence in personal assessments.

Facial appearance plays a role in personality attribution. Several facial features are thought to have an effect on the perception of faces: shape and symmetry, size and spacing of the eyes, nasal width, fullness of the lips, hair color, skin complexion, and the width of the cheekbones. Attractiveness is associated with facial symmetry [85], large and widely spaced eyes, light-colored hair, high forehead, small nose, full lips, and wide cheekbones [86, 84]. Facial features have also been associated with personality characteristics attributed to the target. Positive characteristics are more frequently attributed to attractive people, while negative characteristics are more frequently attributed to unattractive people [87, 88]. Babyish features in adults (i.e., large eyes, small chin, high forehead, and high eyebrows) often evoke attributions of youthfulness and immaturity. Two studies were performed by Paunonen and colleagues [86] to investigate the effects of (a) eye size and spacing, and (b) eye size and mouth fullness on observer perceptions and attributions of personality characteristics. Face images were modified to change the relative size and spacing of the eyes as well as fullness of the lips. Subjects were asked to rate the images on four physical traits (attractiveness, babyfacedness, masculinity, and physical strength) and thirteen personality variables including nurturance, extraversion, popularity, dominance, likeability, honesty, empathy, agreeableness, intelligence, ambition, conscientiousness,

culture, and neuroticism. Results indicate that eye spacing and mouth fullness have no significant impact on personality or physical characteristics attributed to a face. However, eye size demonstrated a strong effect. Analysis showed that eye size primarily impacted ratings of masculinity/femininity, babyfacedness, and attractiveness.

Facial Attractiveness

Attractiveness has been associated with the averageness and symmetry of the face [89]. According to Thornhill and Gangestad [85], natural selection should favor psychological features that evaluate bodily traits that varied with mate value and found attractive those traits correlated with high mate value. The fact that humans share views about what features are attractive implies the presence of species-typical psychological adaptations. Schmidt and Cohn [81] suggest that the evolved perceptual preference for structural symmetry may extend to a preference for symmetry in movement. Spontaneous smiles are more symmetrical than posed smiles and are considered more sincere and attractive. Rhodes et al. [90] investigated the roles of facial averageness and symmetry in signaling health information. Facial averageness and asymmetry were modified through morphing. Subjects were presented with the modified images and were asked to rate the health of the individual. Increases in both symmetry and averageness were found to result in higher perceived health ratings. Facial distinctiveness ratings and asymmetry were then compared to actual health. Facial distinctiveness was found to be correlated with poor childhood health in males. Facial asymmetry was not found to be correlated with actual health. Large deviations from facial averageness are known to be associated with certain chromosomal disorders [85]. Rhodes et al. [90] suggest that more subtle deviations may provide cues

to the health of potential mates. The preference for average faces may have evolved because it increases the likelihood of reproductive success. In normally ovulating women, facial preferences change during the menstrual cycle. Penton-Voak and Perrett [91] found that women prefer more masculine (closer to the male average) faces during the high-fertility phase. In contrast, more feminine faces were preferred during the low-fertility phase. Thornhill and Gangestad [85] suggest that natural selection designed the preferences to shift in response to variations in the relative costs and benefits of mating with a male of best condition.

Social Expectations

Social factors impact the perception of emotional expression. Women are generally considered to be more likely to show happiness while men are more likely to show anger. Hess, Adams, and Kleck [92] investigated the effect of this common stereotype on the perception of emotion in men and women. Drawings of an expressive androgynous face were coupled with either male or female hair styles. Each expression was morphed with the neutral expression resulting in a range of intensities for each emotion. Face stimuli were presented and each subject was asked to rate the level of emotion present. Results indicate that the same anger expression coupled with a male hairstyle was perceived as less angry than when coupled with a female hairstyle. The same happy expression was perceived as more intense when coupled with male hairstyle than with a female hairstyle.

Postural Factors

Changes in the posture of the head relative to the body have been shown to affect the perception of emotional expression attributed to the face. As noted by Mignault and Chaudhuri [93], it is a common perception that an upward tilt of the head relative to the body is associated with positive affect while a downward tilt is often associated with more negative affect. However, projection of the tilted head into two dimensions can reduce the upward curvature of the mouth which may be interpreted as negative affect. Full-face Noh masks used in traditional Japanese theater are capable of inducing a range of perceived expressions through changes in head orientation by skilled actors. Three experiments with British and Japanese subjects were conducted to investigate the effect of head orientation on perceived expressions [94]. The first experiment presented images of a Noh mask tilted at angles from 30 degrees to -30 degrees increments of 5 degrees. In the second experiment the same images were used, however, they were cropped such that only the internal facial features were exposed. The third experiment involved generating a 3-D scan of a Japanese female model. Two-dimensional images of the 3D model were displayed at the same angles in the previous two experiments. Lyons and his colleagues [94] found that projection of the upwardly tilted head into two dimensions can reduce the upward curvature of the mouth which may be interpreted as negative affect. Similarly, a downward tilt produces a two-dimensional image of positive affect. Thus, different postural aspects may contribute to conflicting percepts • one influenced by the shape of the internal features and the other by the orientation of external features.

Categorical Perception of Expression and Emotion

Categorical perception is a psychophysical phenomenon which may occur when a set of stimuli ranging along a physical continuum is divided into categories. Categorical perception involves a greater sensitivity to changes in a stimulus across category boundaries than when the same change occurs within a single category [95]. Categorical perception has been observed in a variety of stimuli including colors [96], musical tones [97], and speech phonology. There is significant evidence that facial expressions are perceived as belonging to distinct categories. In a study by Calder, Young, Perrett, Etcoff and Rowland [98], the categorical perception of facial expressions based on morphed photographic images was investigated. Three expression continua were employed: happiness-sadness, sadness-anger, and anger-fear. Subjects were first asked to identify the individual stimuli by placing them along particular expression continua. Subjects were then asked to perform a discrimination task in which stimuli A, B, and X were presented sequentially. Subjects were asked whether X was the same as A or B. Results indicate that each expression continuum was perceived as two distinct categories separated by a boundary. It was further found that discrimination was more accurate for across-boundary rather than within boundary pairs. Two possible confounds were addressed. First, it was proposed that the categorical perception results could be due to the experimental design in that single continua ranging between two prototypes were used. Second, the discrimination task required a short-term memory component that could account for the categorical perception observed. The experiment was repeated by using continua without fixed endpoints, yielding the same results. Finally, a same-different matching paradigm was used. No

effect of short-term memory was found.

Facial expressions were analyzed using multidimensional scaling (MDS) by Bimler and Kirkland [99]. Blends of five pure expressions were used as stimuli. Subjects were asked to describe the proximities between the expressions using similarity comparisons and sorting partitions. Results of the MDS indicate that the expressions were perceived in distinct categories. Adjacent pairs in the model were space at varying intervals resulting in clusters. The clustering was interpreted as a perceptual magnet within each category.

In a classic study, Young et al. [98] investigated whether facial expressions are perceived as continuously varying along underlying dimensions or as belonging to discrete categories. Dimensional approaches were used to predict the consequences of morphing one facial expression to another. Transitions between facial expressions vary in their effects, depending on how each expression is positioned in the emotion space. Some transitions between two expressions may involve indeterminate regions or a third emotion. In contrast, a transition from one category to another may not involve passing through a region which itself may be another category. In this case, changes in perception should be abrupt. Four experiments were conducted using facial expressions from the Ekman and Friesen series [18, 21, 22, 20, 16, 17, 23]. All possible pairwise combinations of emotions were morphed and presented randomly. Subjects identified intermediate morphs as belonging to distinct expression categories corresponding to the prototype end-points. No indeterminate regions or identification of a third emotion were observed. This supports the view that expressions are perceived categorically rather than by locating them along underlying dimensions. The authors suggest that categorical perception reflects the underlying organization

of human categorization abilities.

Shah and Lewis [100] investigated the location of the neutral expression in a face-emotion space. Objective similarity judgments were made on 25 emotions posed by the same actor. Multidimensional scaling was then applied to generate a perceptual facial-emotion space. The MDS supported the previously reported two-dimensional arrangements of emotions with pleasantness and intensity as the dimensions. The neutral face was located at the periphery suggesting that it be considered an emotion itself rather than the absence of emotion.

2.6 The Importance of Understanding Emotions

AC cording to psychologists, the fulfilment of emotional needs is essential and necessary to human well-being, as living with unmet emotional needs may cause pain, anxiety, depression, or violence eruptions [73], [101], [72], [102]. Indeed, several of the best known problems that plagued human societies in the twentieth century, such as drug and alcohol abuse, violence and criminality, derive from inability to meet such basic emotional needs. The first step in order for someone to fulfil his/her emotional needs is to be aware of them and recognize them, while the next step is to be able to meet them. Emotional needs are often categorized into two main categories: The first category consists of ‘emotional skill needs’ and refers to awareness of emotions, both one’s own and those of others, and the ability to manage them [103],[104].

The second category is referred to as ‘experiential emotional needs’ and tends to follow the Webster Dictionary definition of a need: ‘a physiological or psychological

requirement for the well-being of an organism.’ When one or more of these needs go unmet, an individual may suffer pain. In extreme cases, chronic failure to meet these needs can have very severe effects.

Below are indicative lists of the two aforementioned categories of emotional skill and experiential needs. Specifically, emotional skill needs [103],[104], [105], [72] are needs for basic skills and abilities for handling emotions, such as:

- Emotional self-awareness: a need to learn to appraise and express what one is feeling;
- Managing emotions: the need to handle and regulate feelings so that they are appropriate;
- Self-motivation: a need to learn to harness one’s emotions in the service of a goal, for example by delaying gratification;
- Affect perception: a need to accurately appraise what others are feeling as they are feeling and expressing it;
- Empathy: a need to learn to appreciate what others are feeling (closely linked in the literature to emotional self-awareness);
- Handling relationships, primarily via managing the emotions of others: This skill is a necessary component of friendship, intimacy, popularity, and leadership.

Experiential emotional needs [106], [107], [108] are mostly inherently social needs and are, therefore, usually met only with the assistance or presence of others. These include needs:

- for attention, which is strong and constant in children and fades to varying degrees in adulthood
- to feel that one's current emotional state is understood by others, particularly during strong emotional response
- to love and feel reciprocity of love;
- to express affection and feel reciprocated affection expressed;
- for reciprocity of sharing personal disclosure information;
- to feel connected to others;
- to belong to a larger group;
- for intimacy;
- to feel that one's emotional responses are acceptable to others;
- to feel accepted by others;
- to feel that emotional experience and responses are 'normal';
- for touch, to be touched;
- for security.

2.7 Meeting Emotional Needs with the Help of Advanced Human-Computer Interaction Techniques

TECHNOLOGICAL advances may help people meet their emotional needs, at least during a human-computer interaction session. Although computers cannot replace interpersonal relations, they can assist humans to fulfil their needs. Such a case may arise, for example, during e-learning, where the teacher is not present, but the encouragement of the student or the reward is needed.

Indeed, computers offer great potential for supporting human emotional needs, because of the abilities of modern computational media. More specifically, interactive media:

- are increasingly portable, smaller, and cheaper; therefore, they are increasingly able to be with their users at all times;
 - soon they will be able to sense emotion via a variety of traditional means such as facial expression, tone of voice, and gesture;
 - are able to be eternally attentive, particularly valuable for applications with young children;
 - are sometimes treated by humans as real people;
 - have the potential not only to support educational needs and enable social interaction, but can also help people to partially meet their emotional needs.
- Some of such opportunities are identified below.

2.7.1 Supporting emotional skill needs

Meeting someone's emotional needs is very important during a human-computer interaction process, especially within the framework of educational technology. This way, the computer program can enable learners to acquire academic skills and knowledge. It is conceivable that similar tools can be designed to address emotional skill needs. Software tutors could be built today for students of any age to learn about emotions; other tools could help build emotional awareness and management skills.

Emotional self-awareness is one of the basic emotional skill needs and a system able to recognize and record a person's needs is basic to modern human-computer interaction techniques. A simple way to build such a system is by prompting the user to record emotions, possibly via selection from a list of pre-specified emotions, at random moments of the day. Work on such a tool is in investigation at the MIT Media Laboratory [73], [102].

Another important task is real-time emotion sensing and recognition. This can be achieved, either by facial expression recognition, speech recognition, or gesture recognition or by combining two or three of these techniques. The realization of this technology may represent a fundamental advancement in human-computer interaction. For example, it may enable the development of an emotion-sensitive 'active listener'. Active listening is a simple, yet powerful skill used extensively by experienced therapists, and involves providing non-judgmental feedback, often about a speaker's emotional expression during conversation. While such a tool would probably rely on still-primitive speech processing capabilities, the potential benefit of such a tool is enormous.

2.7.2 Supporting experiential needs

While it is commonly assumed that experiential needs can only be met with the help of other humans, this assumption is not entirely true. People may satisfy several of these needs via other means, such as pet dogs or cats. In fact, people are able to establish relationships with a wide variety of organisms of various degrees of interactivity. During a human-computer interaction session, the system should be able to provide a bonding with the user and enable him to emotionally express himself. This, as an example, can be seen in recent products featuring computational simulations of pets, which demonstrate that interactive media can stimulate pet-like emotional bonding for both children and adults [73], [102]. Again, this conceptualization does not suggest that machines would substitute for interpersonal or even inter-organism contact, but offers a dramatic expansion in the availability and interactivity of non-human companions.

Speech and gesture recognition and facial expression classification can help machines to meet human emotional needs. It is clear that humans can and do meet many of their experiential emotional needs on a daily basis using speech or facial expressions. Computational media has much to offer today and in the future to assist in their provision, but such products require research and development.

3

Previous Related Studies and Systems on Emotion Recognition

If we knew what it was we were doing, it would not be called research, would it?

—*Albert Einstein (1879–1955)*

3.1 Face Databases

THE first step in order to develop a facial analysis system is to acquire an adequate face database. This database can be used for training and testing the system under development. So, it is very important that this database meets the needs and requirements set during development. In this section, we set the requirements of an ideal facial expression database. Based on these requirements, we review the facial expression databases that are currently publically available. This review helps us to

decide whether to use any of these databases or to develop our own facial expression database.

3.1.1 Specifying Requirements for an Ideal Facial Expression Database

For use in the development, training, and testing of facial expression classifiers, appropriate extensive facial databases are required. These databases are non-trivial to create, as they need to be sufficiently rich in both facial expression variety and representative samples of each expression. Moreover, the creators of the database need to make sure that the human models form their true facial expressions when posing.

The first thing that a facial expression classification system developer must take into account is the quality of the input media of the databases. The input media of the databases are usually either static images or image sequences (video). Also, depending on when the database was created, the input media is usually in grayscale format (as in earlier databases) or in color. Moreover, an ideal database of facial expressions should address the following issues:

- It is very important that the facial expression database should cover all the possible emotions. In many cases, the subjects in some databases are asked to form any of the Facial Actions of the Facial Action Coding System (FACS), proposed by Paul Ekman [16]
- Another important factor that it must be taken into consideration, is the human subjects' background. As there are cultural variations in the way in which humans express emotion, an ideal facial expression database should include subjects of any culture. Moreover, as there are changes in the texture of the

human skin depending on the age and the gender of the human, the facial expression database should also cover cases of different age and gender.

- The conditions during photo shooting or video acquisition should also be taken into account. In order to build a facial expression classification system that works independent of external factors, a complete facial expression database is needed for training and testing. Specifically, the facial expression database should include face images under different illumination conditions, and poses, as well as cases of faces which are partially occluded by glasses, scarfs, hats or moustache and/or beard.
- The facial expression database should be extensive enough to be used for training and testing the facial expression classification system. In order to fulfill this requirement, an adequate number of input media, static images or video, should be included in this database.
- As, in some cases, the aim is to build a user dependent facial expression classification system, there is a need of a facial expression database where the subjects are repeatedly forming the same facial expressions and repeated shots are taken.
- Finally, an ideal facial expression image database should include all the expressions formed by each subject.

The aforementioned requirements for an ideal facial expression database are summarized in the following Table 3.1.

Table 3.1 : Requirements for an ideal facial expression database

| Input media quality | |
|----------------------------------|--|
| 1 | Input media should be either images or image sequences (video) |
| 2 | Input media quality (grayscale or color) |
| Emotion Categorization | |
| 3 | Number of emotion' classes depicted by subjects |
| 4 | The subjects depict the FACS |
| 5 | Emotion' classes names |
| Human Subjects' Categorization | |
| 6 | Number of subjects |
| 7 | Subjects of any age? |
| 8 | Subjects of any gender? |
| 9 | Subjects of any culture? |
| Conditions during photo shooting | |
| 10 | The number of sessions that the photo shooting was repeated for each subject |
| 11 | Photo shooting in various illuminations |
| 12 | Photo shooting of subjects in different poses |
| 13 | Photo shooting in cases of partial occlusion of faces |
| 14 | Total number of data |
| 15 | Complete emotion sequences for all the subjects? |

3.1.2 Previous Facial Expression Databases

In the recent years, only a relatively small number of relevant face databases have been presented in the literature, including : (1) The AR Face Database [109], which

contains over 4,000 front-view color images of 126 persons' faces, forming different facial expressions under various illumination conditions and occlusion (e.g., wearing sun glasses and/or a scarf). The main disadvantage of this database is its limitation to containing only four facial expressions, namely 'neutral', 'smile', 'anger', and 'scream'. (2) The Japanese Female Facial Expression (JAFFE) Database [110], which contains 213 images of the neutral and 6 additional basic facial expressions, as formed by 10 Japanese female models. (3) The Yale Face Database [111], which contains 165 gray-scale GIF-formatted images of 15 individuals. These correspond to 11 images per subject of different facial expression or configuration, namely, center-light, with glasses, happy, left-light, without glasses, normal, right-light, sad, sleepy, surprised, and wink. (4) The Cohn-Kanade AU-Coded Facial Expression Database [112], which includes approximately 2000 image sequences from over 200 subjects and is based on the Facial Action Coding System (FACS) first proposed by Paul Ekman [16]. (5) The MMI Facial Expression Database [113], which includes more than 1500 samples of both static images and image sequences of faces in front and side view, displaying various expressions of emotion and single and multiple facial muscle activation.

The aforementioned databases are reviewed in Table 3.2, based on the requirements set on Table 3.1.

Table 3.2 : Review of the facial expression databases

| Reference | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|----------------------------|----------|---|---|---|---|-----|---|---|---|----|----|----|----|-------|----|
| AR Face Database [109] | Img | C | 3 | x | 'neu', 'joy', 'ang', 'scr' | 126 | x | x | x | 2 | ✓ | x | ✓ | 4.000 | ✓ |
| JAFFE Database [110] | Img | C | 7 | x | b.em., 'neu' | 10 | x | x | x | 1 | x | x | x | 213 | ✓ |
| Yale Face Database [111] | Img | G | 6 | x | 'neu', 'joy', 'sad', 'sle', 'sur', 'wink' | 15 | x | x | x | 1 | ✓ | x | ✓ | 165 | ✓ |
| Cohn-Kanade Database [112] | Vid, Img | C | 6 | ✓ | b.em., 'neu' | 182 | ✓ | ✓ | ✓ | 1 | x | ≈ | x | 2.105 | x |
| MMI Database [113] | Vid, Img | C | 6 | ✓ | b.em., 'neu' | 19 | ✓ | ✓ | ✓ | 1 | ≈ | ≈ | ✓ | 2894* | ≈ |

*1395 AUs and 197 emotions

Legend:

✓ : 'Yes'

x : 'No'

≈ : 'Partially available'

Img : Static Image

Vid : Image sequences (Video)

C : Color Images

G : Grayscale Images

Emotions/Expressions : 'neu': 'neutral', 'ang': 'anger', 'scr': 'scream', 'sle': 'sleep', 'sur': 'surprise'

b.em. : 6 Basic Emotions ⇒ anger, 'disgust', 'fear', 'happiness', 'sadness' and 'surprise'

3.1.3 Section Summary - Results

Our study revealed the following problems:

- Many of the aforementioned databases do *not* contain an adequate number of human subjects. Specifically, most databases have been acquired by photo shooting of fewer than 20 subjects, and only two of them, namely the AR Face Database [109] and the Cohn-Kanade Database [112], contain more than 100 subjects. The use of such databases will constrain our system performance, as the facial expression recognition task would be person-dependent and, thus, will not be able to generalize for new persons.
- On the other hand, the two databases that contain more than 100 subjects, and, thus, could lead to the development of a more user independent system, have some problems, specifically: (1) the AR Face Database [109] is constrained to only four emotions/expressions, namely ‘neutral’, ‘happiness’, ‘anger’, and ‘scream’, (2) the Cohn-Kanade Database [112] is based on Facial Action Units and not on emotion classes, which is the aim of our system. Moreover, although we could construct some emotion classes using the Cohn-Kanade Database [112], it is not complete in terms of each subject depicting all the sequence of emotions.
- Additionally, most of these databases do *not* cover the range of face expressions recognized by our system and, therefore, were found insufficient for its development.

Although we used some of these databases (the AR Face Database [109] and the Cohn-Kanade Database [112]) for the early development and testing of our system,

eventually we were forced to create our own face image database which is described in the following Chapter 4.

3.2 Face Detection

3.2.1 Specifying Requirements for an Ideal Face Detection System

FACE detection is one of the visual tasks which humans can do effortlessly. However, in computer vision terms, this task is not trivial to perform. A general statement of the problem can be defined as follows: Given a still or video image, detect and localize an unknown number (if any) of faces. The solution to the problem involves segmentation, extraction, and verification of faces and possibly facial features from an uncontrolled background. As a visual front-end processor, a face detection system should also be able to achieve the task regardless of illumination, orientation, and camera distance.

When building a face detection system, the first thing to be taken into consideration is the type of the input media that will be used to perform the face detection. The input media of the databases are usually either static images or image sequences (videos). Also, another important factor is the format of the input media. Some systems perform face detection based on the skin color, so the input media should be in color format, but, in other cases, the input media can be in grayscale format. Also, an ideal face detection system should adhere the following rules:

- It is very important that the face detection system would be able to perform well regardless of the conditions during image or video acquisition. Specifically, some of the conditions that should be taken into account are the following:

- Changes in illumination
 - De-focus and noise problems
 - Partially occluded faces
 - Rotated faces
 - Faces in side view
 - Complex background
 - Faces of different sizes and portions of image plane
 - Subjects form different facial expressions
 - Subjects of any culture and, especially, any color of the face
- As face detection is usually the first processing step of a fully automated facial analysis system, another important factor that should be taken into account is the ability to perform in real time.

The aforementioned requirements for an ideal face detection system can be summarized in the following Table 3.3.

Table 3.3 : Requirements for an ideal face detection system

| Input media quality | |
|----------------------------------|--|
| 1 | Input media should be either images or image sequences (video) |
| 2 | Input media quality (grayscale or color) |
| Conditions during photo shooting | |
| 3 | Perform well regardless of changes in illumination |
| 4 | Perform well regardless of de-focus and noise problems |
| 5 | Perform well in cases of partially occluded faces |
| 6 | Perform well in cases of different poses: rotated faces |
| 7 | Perform well in cases of different poses: faces in side view |
| 8 | Perform well in cases of complex backgrounds |
| 9 | Perform well in different face sizes |
| 10 | Perform well for different facial expressions |
| 11 | Perform well regardless of the subject's culture |
| 12 | Real time face detection |
| 13 | Success rate (%) |
| 14 | False rate (%) |

3.2.2 Previous Works on Face Detection

The goal of face detection is to determine whether or not there are any faces in an image and, if so, return the location and extent of each face. To address this problem, a number of works have appeared in the literature (e.g. [114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136]). As discussed above, the key issue and difficulty in face detection is to account for the wide range of facial pattern variations in images.

There are four main approaches to address this problem: (1) knowledge - based methods, (2) feature - or image - invariant methods, (3) template matching methods and, (4) appearance-based methods. Knowledge - based methods [115, 128], encode human knowledge of what constitutes a typical face. Usually, rules capture relationships between facial features. On the other hand, approaches based on feature - or image - invariants, aim at finding structural features that exist even when pose, viewpoint and lighting conditions vary, and then use them to detect faces. Template matching methods can be divided in two subcategories which use: (1) correlation templates and, (2) deformable templates. In approaches where correlation templates are used, we compute a difference measurement between one or more fixed target patterns and candidate image locations and the output is thresholded for matches. Deformable templates are similar in principle to correlation templates, except that the latter are more rigid. To detect faces, in this approach we try to find mathematical and geometrical patterns that depict particular regions of the face and we fit the template to different parts of the images and threshold the output for matches. Finally, approaches based on appearance can be considered the reverse of template matching

methods. In this case, the models (or templates) are learnt from a set of training images which should capture the representative variability of facial appearance. These learnt models are then used for detection.

In the past years, several systems have been developed that implement the previous approaches. The system proposed by Colmenarez et al. [114] is template-based; they try to encode face images in a particular prototype. Yang et al. [115] and Lee et al. [116] proposed systems that are knowledge-based; they encode human knowledge of what constitutes a face. Leung et al. [117], applied a local feature detector to find faces in an image. Many systems (Rowley et al [118],[119],Yang et al [120], Sung and Poggio [123] and Juell et al [121]), use artificial neural networks to find faces. Lin et al. [122], proposed a system that searches for potential face regions, based on the triangle that form the eyes and the mouth. Sung et al. [123] used a two distance metrics that measure the distance between the input image and the cluster of faces and non-faces. Lin Lin Huang et al. [124] designed three detection experts which employ different feature representation schemes of local image and then use a polynomial neural network to determine whether or not there is a face in an image. Castrillon et al. [125], developed a system to real time detect faces in video sequences, by means of cue combination. S. Phimoltares et al. [126] developed a two-stage system, which, first, detected the faces from an original image by using Canny edge detection and their proposed average face templates, and, next, use a neural visual model (NVM) to recognize all possibilities of facial feature positions. Kadoury and Levine [127] proposed a novel technique which used locally linear embedding (LLE) to determine a locally linear fit so that each data point can be represented by a linear combination of its closest neighbors and used this representation to train Support Vector Machines

to detect faces. A fairly detailed survey on the methods used for face detection is given on [137],[138]. Liu [132] developed a Bayesian discriminating feature method for face detection, in which the likelihood density was estimated by considering both the projection weights and the residual components in the eigenspace. Papageorgiou et al. [129] proposed an over-complete wavelet model to present an object class for object detection. Finally, Viola and Jones [130] presented a real-time front-view face detection system featuring a cascade of boosting classifiers based on an over-complete set of Haar-like features. Li et al. [134] modified the monotonic assumption of the Adaboost algorithm proposed by Viola and Jones in [130] to develop the so-called Floatboost algorithm for the training of face and nonface classifiers. By implementing these classifiers using a coarse-to-fine and simple-to-complex pyramidal structure, the authors successfully developed a computationally-efficient multi-view multi-face detection system. However, the proposed classifiers used in such boosted cascades operate independently of one another and therefore discard useful information between layers, resulting in convergence problems during the training process

All the studies on face detection are reviewed in Tables 3.4, 3.5 and 3.6 based on the requirements set on Table 3.3.

Table 3.4 : Review of the face detection approaches - 1* (Before 2000)

| Reference | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|-------------------|-----|-----|----|----|----|----|----|----|-------------------|----|----|---------------|---------------|----------------|
| Yang [115] | Img | G | - | - | - | - | - | ✓ | 48*60- 200*280 | - | - | 60- 120sec | 83% | n/a |
| Leung [117] | Img | A | nt | nt | ✓ | ✓ | X | ✓ | n/a | ✓ | nt | n/a | 75% | n/a |
| Dai [139] | Img | C | nt | nt | ✓ | ✓ | - | ✓ | 16*20, 20*26 | ✓ | x | n/a | 100% | 0,34% |
| Tankus [140] | Img | n/a | ✓ | nt | nt | nt | nt | nt | ≥64*96 | nt | nt | x | 92,77% | n/a |
| Colmenarez [114] | Img | G | ✓ | - | - | ✓ | x | ✓ | ≥11*11 | - | ✓ | x | 86,8%- 98% | 0,2%- 2,2 % |
| Kotropoulos [128] | Vid | n/a | ✓ | - | - | - | - | - | n/a | - | - | n/a | 86,5% | n/a |
| Rowley [118, 119] | Img | G | ✓ | - | - | ✓ | x | ✓ | ≥20*20 | - | - | n/a | 86% | n/a |
| Sung [123] | Img | G | ✓ | ✓ | nt | ✓ | x | ✓ | n/a | nt | nt | n/a | 88,1% | n/a |
| Saber [141] | Img | C | nt | nt | nt | ✓ | nt | nt | x | x | nt | ✓ | n/a | n/a |
| Jeng [142] | Img | n/a | nt | nt | nt | ✓ | nt | ✓ | ≥80*80 | nt | nt | av. | 86% | n/a |
| Wang [143] | Img | C | nt | nt | nt | x | nt | x | ≥128*128 | nt | nt | av. | 94% | n/a |
| Wei [144] | Img | C | nt | nt | nt | ✓ | nt | ✓ | n/a | nt | nt | n/a | 70%- 80% | n/a |
| Miao [142] | Img | G | nt | nt | ✓ | ✓ | nt | ✓ | ✓ | nt | nt | x | 83,8% | 3,62% |

*The review is made based on what the authors have written in the cited articles

*If not otherwise stated in the paper, the false rate, if possible, is calculated as: (false detections):(total faces tested) (%)

✓ : 'Yes'

x : 'No'

nt : Not tested

n/a or '-' : Not available

Img : Static Image

Vid : Video

av. : Not tested, but could
be available

C : Color Images

G : Grayscale Images

A : Images of any color

Table 3.5 : Review of the face detection approaches - 2* (2000-2004)

| Reference | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|--------------------|-----|---|----|----|----|----|----|----|--------|----|----|----------|-----------------|------------------|
| Han [145] | Img | G | ✓ | nt | nt | ✓ | nt | ✓ | ≥50*50 | ✓ | nt | 18sec | 94% | n/a |
| Schneiderman [133] | Img | G | ✓ | nt | nt | ✓ | ✓ | ✓ | n/a | ✓ | ✓ | 5sec | 92,3% | n/a |
| Wang [146] | Img | G | ✓ | nt | nt | ✓ | nt | ✓ | n/a | nt | nt | n/a | 84,96% | 3,47% |
| Chen [147] | Img | G | ✓ | nt | nt | ✓ | nt | ✓ | n/a | ✓ | nt | n/a | 88,2% | 1,58% |
| Viola [130, 131] | Img | G | ✓ | nt | nt | ✓ | ✓ | ✓ | ≥24*24 | ✓ | ✓ | 0,067sec | 81,1%- 93,7% | 1,97% - 32,9% |
| Yao [148] | Img | C | ✓ | nt | nt | nt | nt | ✓ | n/a | ✓ | nt | n/a | 95,4% | n/a |
| Wang [149] | Img | C | x | nt | nt | ✓ | nt | ✓ | ≥30*20 | nt | ✓ | n/a | 91,1% | 6,67% |
| Ayinde [150] | Img | A | nt | ✓ | nt | nt | nt | ✓ | n/a | nt | nt | x | 85,7% | n/a |
| Zhou [151] | Img | A | nt | nt | nt | ✓ | nt | ✓ | ≥80*80 | nt | nt | 324sec | 80% | 7% |
| Hock Koh [152] | Any | C | ✓ | nt | nt | ✓ | nt | nt | ≥16*16 | nt | nt | n/a | 95,8% | n/a |
| Liu [132] | Img | G | ✓ | ✓ | ✓ | ✓ | x | ✓ | n/a | ✓ | nt | 1sec | 97,4% | 0,44% |
| Hsieh [153] | Img | C | ✓ | nt | ✓ | nt | nt | ✓ | ≥30*30 | nt | nt | 6,16sec | 80,73% | n/a |
| Huang [154] | Img | A | ✓ | nt | nt | nt | nt | ✓ | ≥18*18 | ✓ | ✓ | n/a | 91,45% | n/a |
| Wu [155] | Img | G | ✓ | nt | x | ✓ | nt | ✓ | n/a | ✓ | nt | ✓ | 95,85% | n/a |
| Wong [156, 157] | Img | C | ✓ | nt | ✓ | ✓ | nt | ✓ | ≥30*30 | nt | nt | n/a | 91,10% | 9,94% |

*The review is made based on what the authors have written in the cited articles

*If not otherwise stated in the paper, the false rate, if possible, is calculated as: (false detections):(total faces tested) (%)

✓ : 'Yes'

x : 'No'

nt : Not tested

n/a or '-': Not available

Img : Static Image

Vid : Video

Any : Image or Video

C : Color Images

G : Grayscale Images

A : Images of any color

Table 3.6 : Review of the face detection approaches - 3* (2005 to present)

| Reference | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|-------------------|-------|-------|-----|----|-----|----|----|----|-------------|----|----|---------|--------|---------|
| Bae [158] | Img | C | ✓ | nt | nt | ✓ | nt | nt | 40*40-80*80 | nt | nt | ✓ | 98% | n/a |
| Kubleck [159] | Vid | C | nt | nt | nt | nt | nt | nt | - | nt | nt | ✓ | 94,3% | 3,6% |
| Shih [160] | Img | A | nt | nt | nt | ✓ | nt | ✓ | ≥16*16 | nt | nt | 10,5sec | 98,2% | 0,7% |
| Kondo [161] | Img | A | ✓ | nt | nt | nt | nt | nt | ≥40*48 | nt | nt | n/a | 97,78% | n/a |
| Wang [162] | Vid | C | nt | nt | nt | ✓ | ✓ | ✓ | - | nt | nt | 200msec | 84,5% | n/a |
| Lin [163] | Img | C | ✓ | nt | ✓ | ✓ | ✓ | nt | n/a | ✓ | ✓ | ✓ | 98,2% | n/a |
| Phimoltares [164] | Img | A | ✓ | ✓ | ✓ | ✓ | nt | ✓ | ≥24*24 | ✓ | nt | 20sec | 97,5% | 10,25% |
| Meynet [165] | Vid | A | ✓ | nt | ✓ | ✓ | nt | ✓ | n/a | nt | nt | n/a | 90,93% | 13,75% |
| Kadoury [166] | Img | G | ✓ | nt | ✓ | ✓ | nt | ✓ | n/a | ✓ | nt | n/a | 97% | 0,0004% |
| Castrillon [167] | Video | Color | n/a | nt | n/a | ✓ | ✓ | ✓ | ✓ | ✓ | nt | ✓ | 99,9% | 8,07% |
| Juang [168] | Image | Color | ✓ | nt | ✓ | ✓ | nt | ✓ | n/a | ✓ | nt | n/a | 95,55% | 15,26% |

*The review is made based on what the authors have written in the cited articles

*If not otherwise stated in the paper, the false rate, if possible, is calculated as: (false detections):(total faces tested) (%)

✓ : 'Yes'

x : 'No'

nt : Not tested

n/a or '-': Not available

Img : Static Image

Vid : Video

Any : Image or Video

C : Color Images

G : Grayscale Images

A : Images of any color

3.2.3 Section Summary - Results

In the recent years, some interesting studies on face detection have appeared in the literature with good success rates. Despite that fact, the majority of the studies on

face detection have some major drawbacks: (1) they have not been tested in order to cover cases of de-focus and noise problems. (2) They usually have a size limit on the detected face, which must usually be bigger than 30-by-30 pixels. (3) They cannot detect many faces (more than 3 faces) in complex backgrounds. (4) They cannot all address the problem of partial occlusion of mouth or wearing sunglasses. (5) Although some of the approaches presented above, try to deal with multi-view face images, it is not easy to detect faces in side view. Although there are some researches that can solve two or three of these problems, there is still no system that can solve all of them. Moreover, some of them can only detect faces in color images, as the use skin extraction techniques to achieve this task. Finally, many of the aforementioned methods cannot perform in real time.

3.3 Facial Expression Classification System

3.3.1 Specifying Requirements for our Facial Expression Classification System

In the first step before developing a facial expression classification system, we should specify the requirements and decide on its functionality. A very interesting and important study regarding this issue was made by Pantic and Rothkrantz [1]. In this section we discuss corresponding requirements and further enhance them with our own observations.

Firstly, there are some questions/statements we need to set:

1. Facial expression understanding was firstly used by humans during interpersonal relationships. So, the first known system which is considered to have the best

performance is the human. Some questions derive from this fact:

- (a) Has the human system truly the best performance?
- (b) In which cases, are people prone to errors when classifying an expression?
- (c) Should we try to understand the human system and mimic it through our application?
- (d) Can we use the success rates of this system as a metric for our system?

2. It is necessary that our system be fully automated. In order to achieve this, all of the stages of the facial expression analysis are to be performed automatically, namely, face detection, facial expression information extraction, and facial expression classification. So, first, before a facial expression can be analyzed, the face must be detected in a scene. In the case of static images, the process of detecting the face is referred to as *localizing* the face in the scene. In the case of facial image sequences, this process is referred to as *tracking* the face in the scene. The next step is to devise mechanisms for extracting the facial expression information from the observed facial image or image sequence. At this point, a clear distinction should be made between two terms, namely, *facial features* and *face model features*. The facial features are the prominent features of the face, such as eyebrows, eyes, nose, mouth, and chin. The face model features are the features used to represent (model) the face. The face can be represented in various ways, e.g., as a whole unit (holistic representation), as a set of features (analytic representation) or as a combination of these (hybrid approach). The applied face representation and the kind of input images determine the choice of mechanisms for automatic extraction of facial expression information. The final

step is facial expression classification, based on the classes of the expressions we have defined before building the system.

3. Based on the assumptions made in Chapter 2, besides the ‘basic emotions’ which, according to the theory of evolution and the fact that are universally common regardless culture of the subject, there are also cultural variations in the way in which humans express emotion. An ideal system should take this into account
4. Also, the system should work regardless of external factors when acquiring a face image or a sequence of face images. This means, that the system should perform robustly despite changes in lightning conditions and distractions such as glasses, changes in hair style, and facial hair like moustache, beard and eyebrows that have grown together.
5. Another important factor is setting the classes of emotion that the system is ideally expected to recognize. In our case, we will use our system for advanced human computer interaction techniques, so the corresponding set of emotions should result from those emotions that are present during a typical human computer session. Until now, the majority of the facial expression classification systems, deal with one of the following sets:
 - (a) the six ‘basic emotions’
 - (b) some other set of emotions, usually a subset of the six ‘basic emotions’
 - (c) FACS: there are no distinctive classes emotion but, rather, facial action movements

Again, based on the studies by Pantic and Rothkrantz [1], if the system is to be used for behavioral science research purposes it should perform facial expression recognition as applied to automated FACS encoding. This is explained thoroughly in Bartlett et al. [169, 170] where it is stated that the system should accomplish multiple quantified expression classification in terms of 44 AUs (Action Units) defined in FACS (Facial Action Coding System). If the system is to be used for advanced multimodal human-computer interaction techniques, the system should be able to understand the shown facial expressions and map them to the respective emotion. In this case, we have the possibility that the system will work with the six ‘basic emotions’ or some other set of emotions. Since psychological researchers disagree on existence of universal categories of emotional facial displays, an ideal system should be able to adapt the classification mechanism according to the user’s subjective interpretation of expressions, e.g., as suggested in [171]. Also, it must be taken into account that in some cases not every facial expression can be classified to only one emotion class [16].

The requirements for an ideal facial expression classification system were first set by Pantic and Rothkrantz [1], based on which the following Table 3.7 has resulted.

Table 3.7 : Requirements for an ideal facial expression classification system [1]

| | |
|----|--|
| 1 | Automatic facial image acquisition |
| 2 | Subjects of any age, ethnicity and outlook |
| 3 | Deals with variation in lightning |
| 4 | Deals with partially occluded faces |
| 5 | No special markers / make-up required |
| 6 | Deals with rigid head motions |
| 7 | Automatic face detection |
| 8 | Automatic facial expression data extraction |
| 9 | Deals with inaccurate facial expression data |
| 10 | Automatic facial expression classification |
| 11 | Distinguishes all possible expressions |
| 12 | Deals with unilateral facial changes |
| 13 | Obeys anatomical rules |
| 14 | Distinguishes all 44 facial actions (FACS) |
| 15 | Quantifies facial action codes |
| 16 | # interpretation categories unlimited |
| 17 | Features adaptive learning facility |
| 18 | Assigns quantified interpretation labels |
| 19 | Assign multiple interpretation labels |
| 20 | Features real-time processing |

3.3.2 Facial Expression Classification Approaches

Previous attempts to address problems of facial expression classification in images fall within one of two main directions in the literature: (1) methods that use image sequences (video) ([172, 173, 174, 174, 175]) and (2) methods that use static images. Approaches that use image sequence often apply optical flow analysis to the image sequence and rely on pattern recognition tools to recognize optical flow patterns associated with particular facial expression. This approach requires acquisition of multiple frames of images to recognize expressions and, thus, has limitations in real-time performance and robustness. Facial expression recognition using still images can be divided in two main categories: (1) methods based on face features ([176, 177, 19, 178, 179, 170, 180]), and (2) methods that utilize image-based representations of the face ([181, 182, 183, 184]). Methods that use facial features for facial expression recognition have fairly fast performance, but the challenge in this approach is to develop a feature extraction method that works well regardless of variations in human subjects and environmental conditions. Methods that utilize image-based representation have as an input the entire facial image which is preprocessed in various ways (e.g. Gabor Filters) or is given to a classifier that recognizes the facial expression. The aforementioned methods usually work well in generalizing for other face images, not in the database, but it is fairly difficult to train such classifier. Finally, with the advances in technology, there are some new methods based on thermal imagery (e.g. [185]), but, in this cases, there is a need for a more sophisticated hardware, which makes it difficult to use in everyday human-computer interaction.

Another fundamental issue about facial expression classification is to define a set

of expression categories we are interested in. A related issue is to devise mechanisms of categorization. Facial expressions can be classified in various ways: (1) Methods that try to classify the image face in discrete facial emotions (e.g. [186, 187, 188]) and, (2) Methods that try to classify the image in terms of facial actions that cause an expression (e.g. [189, 190, 184, 191]). The majority of these methods use the Facial Action Coding System (FACS) [17]. An extended survey about all the aforementioned methods can be found in [192].

Early Years of Facial Expression Recognition

In the early years of facial expression classification, which can be identified as the period between 1990 and 1995, three main approaches can be identified and summarized as follows: Cottrell et al. [193], Rahardja et al [194] and Matsuno et al. [195] use holistic spatial analysis to classify the expression. Whereas, Mase et al. [196], Moses et al. [197], Rosenblum et al. [198] and Yacoob et al. [199] use techniques based on spatio-temporal analysis and Kearney et al. [171], Kobayashi et al. [200], Ushida et al. [201] and Vanger et al. [202] use analytic spatial analysis.

The resulting evaluation of the aforementioned systems has already been done by Pantic and Rothkrantz [1] and is shown in Table 3.8.

Table 3.8 : Review of the facial expression approaches (Early Years) [1]

| Reference | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 16 | 17 | 18 | 19 | 20 |
|---|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|
| Input media: Static images | | | | | | | | | | | | | | | | | | | |
| Cottrell [193] | x | ✓ | x | x | ✓ | - | x | - | - | ✓ | x | ✓ | - | x | 8 | x | x | x | x |
| Kearney [171] | x | ✓ | x | x | ✓ | - | x | x | - | ✓ | x | ✓ | ✓ | 36 | n | ✓ | x | ✓ | x |
| Kobayashi [200] | x | ✓ | x | x | ✓ | - | x | x | - | ✓ | x | ✓ | ✓ | x | 6 | x | ✓ | ✓ | x |
| Matsuno [195] | x | x | ✓ | x | ✓ | - | x | x | - | ✓ | x | ✓ | ✓ | x | 4 | x | x | x | x |
| Rahardja [194] | x | - | x | - | - | - | x | - | - | ✓ | x | - | - | x | 6 | x | x | x | x |
| Ushida [201] | x | ✓ | x | x | ✓ | - | x | x | - | ✓ | x | x | ✓ | x | 3 | x | x | x | x |
| Vanger [202] | x | ✓ | x | x | ✓ | - | x | x | - | ✓ | x | ✓ | ✓ | x | 7 | x | x | x | x |
| Input media: Image Sequences (Video) | | | | | | | | | | | | | | | | | | | |
| Mase [196] | x | - | x | x | ✓ | x | x | ✓ | x | ✓ | x | ✓ | ✓ | x | 4 | x | x | x | x |
| Moses [197] | ✓ | - | ✓ | x | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | x | ✓ | ✓ | 5 | 5 | x | x | x | x |
| Rosenblum [198] | x | - | - | x | ✓ | ✓ | x | ✓ | ✓ | ✓ | x | ✓ | ✓ | x | 2 | x | x | x | x |
| Yacob [199] | - | - | - | x | ✓ | ✓ | x | ✓ | ✓ | ✓ | x | ✓ | ✓ | x | 7 | x | ✓ | x | x |

✓ : 'Yes' x : 'No' '-' : Not available

Mid-years of Facial Expression Recognition (1996-2000)

Between 1996 and 2000, related works can be summarized in the following categories:

- In terms of the form of input data: (1) Static images [203, 204, 205, 206, 207,

208, 209, 190, 210, 182] and, (2) Image sequences [211, 212, 213, 214, 215, 216].

- In terms of facial expression information extraction: (1) template-based methods [203, 204, 205, 206, 207, 208, 211, 212, 213, 214, 215, 182, 217] and, (2) feature-based methods [209, 190, 210, 216].
- In terms of the method used for the classification to the respective expression: (1) template-based methods [203, 204, 205, 218, 207, 212, 213, 214, 215], (2) Neural-network based [209, 206, 208, 210, 182] and, (3) Rule-based methods [190, 211]

The aforementioned categorization of the previous studies is summarized in the following Table 3.9

Facial Expression Recognition on static images:

Edwards et al. [203] use a holistic face representation, which they refer to as the Active Appearance Model (AAM). The AAM contains a statistical, photo-realistic model of the shape and gray-level appearance of faces. To build the AAM, they used facial images that were manually labeled with 122 points localized around the facial features and are considered the key positions to outline the main features. They generated the statistical model of shape variation, by aligning all training images into a common coordinate frame and applying PCA to get a mean shape. The statistical model of the gray-level appearance, was built by warping each training image, by using a triangulation algorithm, so that its control points match the mean shape. By applying PCA to the gray-level information extracted from the warped images, they obtained a mean normalized gray-level vector. The models were generated by combining a model of shape variation with a model of the appearance variations in a

Table 3.9 : Categorization of the approached based on the techniques and the input media

| Input media: Static Images | |
|---|--------------------------|
| Used techniques | |
| <i>Template-based</i> | <i>Feature based</i> |
| Edwards [203] | Kobayashi and Hara [209] |
| Hong [204] | Pantic [190] |
| Huang [205] | Zhao [210] |
| Padgett [206] | Dailey [182] |
| Yoneyama [207] | |
| Zhang [208] | |
| Lyons [217] | |
| Input media: Image Sequences (Video) | |
| Used techniques | |
| <i>Template-based</i> | <i>Feature based</i> |
| Black [211] | Cohn [216] |
| Essa [212] | |
| Kimura [213] | |
| Otsuka [214] | |
| Wang [215] | |

shape-normalised frame. Their goal was to identify the subject regardless of the conditions during photo acquisition (such as pose and facial expression). They assume

that there is the inter-class variation (which involves the aforementioned conditions) and which is very similar for each individual. They use the Mahalanobis distance and Linear Discriminant Analysis to linearly separate the inter-class variability from the intra-class variability. Their approach seeks to find a linear transformation of the appearance parameters which maximizes inter-class variation, based on the pooled within-class and between-class covariance matrices. They used 200 of six basic emotional expressions and 'neutral' emotion categories for training and other 200 images for testing. The system was tested against the answers of 25 human observers and achieved an accuracy of 74%. It is not known how the method will behave in the case of an unknown subject.

Kobayashi and Hara [209] tried to build an animated 3-dimensional face robot for communicative interaction with human beings which would give the impression of a realistic human-like response. So, the face robot could produce human-like facial expressions and recognize human facial expressions using facial image data obtained by a CCD camera mounted inside the left eyeball. The camera gives the brightness distributions of the face and, thus, helps in locating the iris of the subject. They normalize the input image by using an affine transformation so that the distance between these irises becomes 20 pixels. They use 13 vertical lines and their length is being computed empirically based on the distance between the irises. The range of the acquired brightness distributions is normalized to [0,1] and these data are given further to a trained neural network for expression emotional classification. They used a 243-by-50-by-6 layered neural network and achieved the correct recognition rate of 85% for six typical human computer interaction expressions of 15 subjects in 55ms.

Hong et al. [204] made an online facial expression recognition system based on personalized galleries. This system is built on the framework of the PersonSpotter system, which is able to track and detect the face of a person in a live video sequence. By utilizing the recognition method of Elastic Graph Matching, the most similar person is found, whose images are stored in the gallery. Then, the personalized gallery of this person is used to recognize the expression on the probe face. A personalized gallery consists of images of the same person showing different facial expressions. Node weighting and weighted voting in addition to Elastic Graph Matching are applied to identify the expression. The personalized galleries of nine people have been utilized, where each gallery contained 28 images (four images per expression). The personalized gallery of the best matching person is used to make the judgement on the category of the observed expression. The method has been tested on images of 25 subjects. The achieved recognition rate was 89% in the case of the familiar subjects and 73% percent in the case of unknown persons.

Huang et al. [205] introduced an automatic facial expression recognition system which consists of two parts: facial feature extraction and facial expression recognition. The system applies the point distribution model and the gray-level model to find the facial features. Then the position variations of certain designated points on the facial feature are described by 10 action parameters (APs). There are two phases in the recognition process: the training phase and the recognition phase. In the training phase, given 90 training image samples of six classes representing six basic emotional expressions, the system classifies the principal components of the APs of all training expressions into six different clusters. In the recognition phase, given a facial image sequence, the system identifies the facial expressions by extracting the 10 APs,

analyzes the principal components, and finally calculates the AP profile correlation for a higher recognition rate. The proposed method has been tested on another 90 images shown by the same subjects. The achieved correct recognition ratio was 84.5%. It is not known how the method will behave in the case of unknown subjects.

Lyons et al. [218] proposed a method for automatically classifying facial images based on labeled elastic graph matching, a 2D Gabor wavelet representation, and linear discriminant analysis. They compute the value of the Gabor transform coefficients and combine these data into a single vector which they call labeled-graph vector (LG vector). The ensemble of LG vectors from a training set of images are subjected to principal components analysis (PCA) to reduce the dimensionality of the input space. LG vectors project into the lower dimensional PCA space (LG-PCA vectors). The ensemble of LG-PCA vectors from the training set are then analyzed using linear discriminant analysis (LDA) in order to separate vectors into clusters having different facial attributes. They used binary classifiers for the presence or absence of a particular facial expression. They built six binary classifiers, one for each basic emotion category, and combined them into a single facial expression classifier. An input image that is not positively classified for any category is classified as 'neutral'. They used 10-fold cross validation of 193 images of different facial expressions displayed by nine Japanese females (Zhang et al. [208]) to train and test the classifier. The generalization rate was 92% for subjects in the database, whereas it was 75% for recognition of expression of a novel subject.

Padgett and Cottrell [206] used three different face images which are fed to an artificial neural network classifier: full face projections of the dataset onto their eigenvectors (eigenfaces); a similar projection constrained to eye and mouth areas (eigen-

features); and, finally, a projection of the eye and mouth areas onto the eigenvectors obtained from 32-by-32 random image patches from the dataset. They adopted the third approach which achieved better results. The neural network consisted of: (1) a single hidden layer with 10 nodes using as activation function the nonlinear sigmoid function and, (2) the output layer of seven units, each of which corresponds to one emotion category of the six basic emotions and the 'neutral'. They trained using the backpropagation algorithm with images of 11 subjects and tested it on the images of the 12th subject. The system achieved 86% generalization on novel face images (individuals the networks were not trained on).

Pantic and Rothkrantz [190] developed an expert system called Integrated System for Facial Expression Recognition (ISFER), which performs recognition and emotional classification of human facial expression from a still full-face image. The system consisted of two major parts: (1) the ISFER Workbench, which forms a framework for hybrid facial feature detection by applying multiple feature detection techniques in parallel and, (2) its inference engine called HERCULES, which converts low level face geometry into high level facial actions, and then this into highest level weighted emotion labels. The multidetector performs locates the contours of the facial features, from which the model features are extracted. The difference between the currently detected model features and the same features detected in an expressionless face of the same person is computed. They use rules based on the knowledge acquired from FACS, to classify the calculated model deformation into the appropriate 31 AUs-classes. The performance of the system in automatic facial action coding from dual-view images has been tested on a set of 496 dual views (31 expressions of separate facial actions shown twice by eight human experts). The average recognition rate was

92% for the upper face AUs and 86% for the lower face AUs.

Zhang et al. [208] used two types of features extracted from 256-by-256 face images for recognizing facial expressions: (1) geometric positions of a set of 34 facial points on a face and, (2) a set of multi-scale and multi-orientation Gabor wavelet coefficients extracted from the face image at 34 facial points. These two types can be used either independently or jointly. In order to classify the expressions, they built a two-layer perceptron. The recognition performance with different types of features has been compared, which shows that Gabor wavelet coefficients are much more powerful than geometric positions. They finally conclude to a 680-by-7-by-7 neural network, which classifies the six basic emotions and the 'neutral' emotion. The neural network performs a nonlinear reduction of the input dimensionality and makes a statistical decision about the category of the observed expression. Each output unit gives an estimation of the probability of the examined expression belonging to the associated category. The network has been trained using a resilient propagation. The input to the network consists of the geometric position of the 34 facial points and 18 Gabor wavelet coefficients sampled at each point. The neural network's output is a 7-by-1 vector where each output unit gives an estimation of the probability of the examined expression belonging to the associated category. The network has been trained using the 10-fold cross validation algorithm with resilient propagation for 213 images of different expressions displayed by nine Japanese females. This database has been used to train and test the used network. So, the database has been partitioned into ten segments. Nine segments have been used to train the network while the remaining segment has been used to test its recognition performance. The achieved recognition rate was 90.1%. The performance of the network is not tested for recognition of expression of a novel

subject.

Yoneyama et al [207] compute the outer corners of the eyes, the height of the eyes, and the height of the mouth in an automatic way. Once these features are identified, the examined face image is divided into 8×10 regions. Accordingly, 8×10 ternary values [+1 (moving upward), 0 (no movement), -1 (moving downward)] computed from averaged value of the optical flows in each region are used as the feature parameters. They use two kinds of discrete type Hopfield neural networks in order to recognize four facial expressions, namely: 'surprise', 'anger', 'happiness' and 'sadness'. The two neural networks trained by different learning data are cascade-connected to compensate each other. As the experimental result, the averaged recognition rate for those four expressions was obtained 92.2%.

Zhao and Kearney [210] also use artificial neural network classifiers. They manually extract and compute 10 facial distances from 94 images from the Ekman and Friesen database [16] which feature the six basic emotions. The difference between a distance measured in an examined image and the same distance measured in an expressionless face of the same person was normalized. Then, each such measure was mapped into one of the eight signaled intervals of the appropriate standard deviation from the corresponding average. These intervals formed the input to the neural network. The output of the neural network represents the associated emotion (e.g., the string '001' is used to represent happiness). The neural network was trained and tested on the whole set of data (94 images) with 100% recognition rate. It is not known how the method will behave in the case of an unknown subject.

Dailey et al [182] also use artificial neural networks. The model begins by computing a biologically plausible representation of its input, which is a static image of

an actor portraying a prototypical expression of the six basic emotions plus the ‘neutral’. The input data are computed using the Gabor Lattice Representation which first computes the basic feature which is a 2-D Gabor wavelet filter and, secondly, combines two filters to get phase insensitivity, modeling complex cell responses in primary visual cortex. The dimensionality of the resulting vector is further reduced using Principal Component Analysis (PCA). The resulting feature vector consists of 50 inputs which are fed to a 6-unit softmax neural network with no hidden layers. The network was trained with a dataset of 12 actors forming the seven expressions, whereas the 13th was used for testing between the training epochs to test the network’s performance and adjust its weights. The expression formed by the 14th actor was used for the final testing. In total, the build 182 networks and applied different combinations of the datasets for training and testing. In average, the neural networks achieved an accuracy of 85.9%.

Facial Expression Recognition on image sequences:

Black and Yacoob [211] use local parametrized models of image motion for recovering and recognizing the non-rigid and articulated motion of human faces. They assume that within local regions in space and time, such models not only accurately model non-rigid facial motions but also provide a concise description of the motion in terms of a small number of parameters. The recovered image motion parameters correspond simply and intuitively to various facial expressions and are used to derive mid-level predicates describing the image motion of the facial features. High-level recognition rules describe the temporal structure of a facial expression in terms of these mid-level predicates. For each of six basic emotional expressions, they developed a model represented by a set of rules for detecting the beginning and ending of

the expression. The rules are applied to the predicates of the midlevel representation. The method has been tested on 70 image sequences containing 145 expressions shown by 40 subjects ranged in ethnicity and age. The expressions were displayed one at the time. The achieved recognition rate was 88%.

Cohn et al. [216] developed and implemented an optical-flow based approach (feature point tracking) that is sensitive to subtle changes in facial expression. They use a model of facial landmark points localized around the facial features. This model is first marked by hand with a mouse device in the first image frame of the subject, and, afterwards, uses an hierarchical optical flow method to track the optical flows of 13-by-13 windows surrounding the landmark points. The displacement of each landmark point is calculated by subtracting its normalized position in the first frame from its current normalized position. This displacement is used as a feature vector to classify the expression. In order to do this, they apply separate discriminant function analyzes within facial regions of the eyebrows, eyes, and mouth. For training and testing, they used the image sequences of 100 human subjects depicting 872 facial actions. They used two discriminant functions for three facial actions of the eyebrow region, two discriminant functions for three facial actions of the eye region. Feature point tracking demonstrated high concurrent validity with human coding using the Facial Action Coding System (FACS).

Essa and Pentland [212] developed a computer vision system for observing facial motion by using an optimal estimation optical flow method coupled with geometric, physical and motion-based dynamic models describing the facial structure. First, by modeling the elastic nature of facial skin and the anatomical nature of facial muscles they developed a dynamic muscle-based model of the face, including FACS-

like [18] control parameters. In order to use this model, they needed to locate and extract the position of the facial features. Initially they started their estimation process by manually translating, rotating and deforming their 3-D facial model to fit a face in an image. To automate this process, in more recent efforts, they used the view-based and modular Eigenspace methods of Pentland and Moghaddam [219]. They compute the eigenfaces to automatically track the face in the scene and extract the positions of the eyes, nose, and mouth. These feature positions are used to wrap the face image to match a canonical face mesh. This allows us to extract the additional ‘canonical feature points’ on the image that correspond to the fixed (non-rigid) nodes on their mesh. After the initial registering of the model to the image the coarse-to-fine flow computation methods presented by Simoncelli [220] and Wang et al. [221] are used to compute the flow. The model on the face image tracks the motion of the head and the face correctly as long as there is not an excessive amount of rigid motion of the face during an expression. By learning ‘ideal’ 2D motion views for each expression category, they generated the spatio-temporal templates for six different expressions, two facial actions (smile and raised eyebrows) and four emotional expressions (surprise, sadness, anger, and disgust). Each template has been delimited by averaging the patterns of motion generated by two subjects showing a certain expression. The Euclidean norm of the difference between the motion energy template and the observed image motion energy is used as a metric for measuring similarity/dissimilarity. When tested on 52 front-view image sequences of eight people showing six distinct expressions, a correct recognition rate of 98% has been achieved.

Kimura and Yachida [213] not only did they recognize some kind of facial expressions which is associated with human emotion but also they estimated its degree.

Their method was based on the idea that facial expression recognition can be achieved by extracting a variation from expressionless face with considering face area as a whole pattern. For the purpose of extracting subtle changes in the face such as the degree of expressions, it is necessary to eliminate the individuality appearing in the facial image. Using an elastic net model, a variation of facial expression is represented as motion vectors of the deformed Net from a facial edge image. Then, by applying K-L expansion, the change of facial expression represented as the motion vectors of nodes is mapped into low dimensional eigen space, and estimation is achieved by projecting input images on to the Emotion Space. They constructed three kinds of expression models: 'happiness', 'anger', 'surprise'. To test the system, they prepared 10 sequential images consisting of 20 frames for these three kinds of expression. These input images were from the same person as these from model images. The system showed good results in recognizing the expressions of the same person but failed to recognize the expressions of an unknown person.

Otsuka and Ohya [222] used Hidden Markov Model for facial expression recognition. They proposed a method that can be used for spotting segments that display facial expression. The motion of the face is modeled by Hidden Markov Model in such a way that each state corresponds to the conditions of facial muscles, e.g., relaxed, contracting, apex and relaxing. The probability assigned to each state is updated iteratively as the feature vector is obtained from image processing. A spotted segment is placed into a certain category when the probability of that category exceeds a threshold value. The Hidden Markov Models were trained on 120 image sequences, shown by two male subjects. The method was tested on image sequences shown by the same subjects and Otsuka and Ohya claimed that the recognition performance

was good. Therefore, it is not known how the method will behave in the case of an unknown expresser.

Wang et al. [215] used a 19-point (12 of them are used for facial expression recognition) labeled graph to track the facial features in an input image sequence. To represent the relationship between the motion of features and change of expression, they constructed expression change models by using B-spline curves. Each curve describes the relationship between the expression change and the displacement of the corresponding facial feature point in the labeled graph. Each expression model has been defined from 10 image sequences displayed by five subjects. The category of an expression is decided by determining the minimal distance between the actual trajectory of the 12 points and the trajectories defined by the models. The method has been tested on 29 image sequences of three emotional expressions shown by eight subjects (young and of Asian ethnicity). The images were acquired under constant illumination and none of the subjects had a moustache, a beard or wear glasses. The average recognition rate was 95%.

Lien, Kanade and Cohn [174] developed a computer vision system that is sensitive to subtle changes in the face. The system included three modules to extract feature information: dense-flow extraction using a wavelet motion model, facial feature tracking, and edge and line extraction. The feature information thus extracted was fed to discriminant classifiers or hidden Markov models that classify it into FACS [18] action units. The system was tested on image sequences from 100 male and female subjects of varied ethnicity (65% female, 35% male, 85% European-American, 15% African-American or Asian, ages 18 to 35 years). Subjects sat directly in front of the camera and performed a series of facial expressions that included single action units and com-

binations of action units. Each expression sequence began from a neutral face. Each frame in the sequence was digitized into a 640 by 490 pixel array with 8-bit precision for gray scale. In the brow region, three action units or action unit combinations were analyzed, 92% were correctly classified by dense flow extraction with HMM, 91% by facial-feature tracking with discriminant analysis, 85% by facial feature-tracking with HMM, and 88% by high-gradient component detection with HMM. In the eye region, analysis was limited to facial-feature tracking with discriminant analysis. Three action units were classified with 88% accuracy. There was some disagreements between two action units, but the authors claimed that there is also difficulty to discriminate for manual FACS coders as well. In the mouth region, 6 action units were analyzed by dense-flow extraction with HMM, 9 action units by facial-feature tracking with discriminant analysis, 6 action units by facial-feature tracking with HMM, and two action units by high-gradient component detection. Accuracy was above 80% for each module. The percentage correctly classified by dense-flow extraction with HMM was 92%. The percentage correctly classified by facial-feature tracking with discriminant analysis was 81% and by facial-feature tracking with HMM was 88% . The percentage correctly classified by high-gradient component detection with HMM was 81%. In conclusion, agreement with manual FACS coding was strong for the results based on dense-flow extraction and facial feature tracking, and strong to moderate for edge and line extraction.

All the aforementioned methods are evaluated based on the requirements set by Pantic and Rothkrantz [1] and this evaluation is summarized in the following Table 3.10 [1].

Table 3.10 : Review of the facial expression approaches (Middle Years) [1]

| Reference | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 16 | 17 | 18 | 19 | 20 |
|---|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|
| Input media: Static images | | | | | | | | | | | | | | | | | | | |
| Edwards [203] | ✓ | - | ✓ | - | ✓ | ✓ | x | ✓ | - | ✓ | x | ✓ | ✓ | x | 7 | x | x | x | x |
| Hara [209] | ✓ | 1 | - | - | ✓ | x | ✓ | ✓ | - | ✓ | x | ✓ | - | x | 6 | x | x | x | ✓ |
| Hong [204] | ✓ | - | x | - | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | x | ✓ | ✓ | x | 7 | x | x | x | ✓ |
| Huang [205] | ✓ | 1 | x | - | ✓ | x | ✓ | ✓ | - | ✓ | x | ✓ | ✓ | x | 6 | x | x | x | x |
| Lyons [217] | x | ✓ | x | - | ✓ | - | x | x | - | ✓ | x | ✓ | ✓ | x | 7 | x | x | x | x |
| Padget [206] | x | ✓ | x | - | ✓ | - | x | x | - | ✓ | x | ✓ | ✓ | x | 7 | x | x | x | x |
| Pantic [190] | ✓ | 3 | x | - | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | x | ✓ | ✓ | 31 | 6 | x | ✓ | ✓ | x |
| Yoneyama [207] | ✓ | 1 | - | - | ✓ | x | ✓ | ✓ | - | ✓ | x | ✓ | ✓ | x | 4 | x | x | x | - |
| Zhang [208] | x | ✓ | x | - | ✓ | - | x | x | - | ✓ | x | ✓ | ✓ | x | 7 | x | ✓ | ✓ | x |
| Zhao [210] | x | ✓ | x | - | ✓ | - | x | x | - | ✓ | x | ✓ | - | x | 6 | x | x | x | x |
| Dailey [182] | x | - | - | - | ✓ | - | x | x | - | ✓ | x | - | x | x | 7 | x | x | x | x |
| Input media: Image Sequences (Video) | | | | | | | | | | | | | | | | | | | |
| Black [211] | ✓ | - | ✓ | - | ✓ | ✓ | x | ✓ | x | ✓ | x | ✓ | ✓ | - | 6 | x | ✓ | x | x |
| Cohn [216] | ✓ | 3 | x | - | ✓ | x | x | ✓ | x | ✓ | x | ✓ | ✓ | 15 | - | x | x | x | - |
| Essa [212] | ✓ | ✓ | ✓ | - | ✓ | - | ✓ | ✓ | ✓ | ✓ | x | ✓ | ✓ | 2 | 4 | x | x | x | ✓ |
| Kimura [213] | ✓ | x | ✓ | - | ✓ | x | ✓ | ✓ | ✓ | ✓ | x | ✓ | ✓ | x | 3 | x | ✓ | x | - |
| Otsuka [214] | ✓ | - | - | - | ✓ | ✓ | - | ✓ | x | ✓ | x | x | ✓ | x | 6 | x | x | x | x |
| Wang [215] | ✓ | 1 | x | - | ✓ | x | x | ✓ | - | ✓ | x | ✓ | ✓ | x | 3 | x | ✓ | x | x |

✓ : 'Yes'

x : 'No'

'-' : Not available

Recent studies on Facial Expression Recognition (2001-present)

In the final section of this survey, we will discuss the more recent studies in facial expression recognition, from 2001 until the time of preparation of this thesis (May 2009). The techniques used in these studies fell within the same categories as in previous studies, so, in terms of facial expression information extraction, some of the studies use template-based methods, whereas other studies use feature-based methods. For classification, more sophisticated classifiers are also used, such as Hidden Markov Models (HMMs) [175, 223, 224] and Support Vector Machines (SVMs) [170, 185, 225] and only one uses Artificial Neural Network classifiers [226].

Cohen et al. [175] developed a real time facial expression classification system for human-computer intelligent interaction (HCII), which classified the expression into the six classes of the basic emotions, plus the ‘neutral’. They first track the face in real time and afterwards use the features extracted from the face tracking for facial expression recognition. They build several different classifiers for recognizing the facial expressions. In the first class of classifiers, they used the features extracted for each frame in the video sequence to produce a classification result for that frame. In the second type of classifiers, they developed a multi-level Hidden Markov (HMM) classifier, combining the temporal information to both automatically segment the video sequence to the different expressions and perform the classification of each segment to the corresponding facial expression. Their face tracking system was based on a system developed by Tao and Huang [227] called the Piecewise Bézier Volume Deformation (PBVD) tracker. This face tracker uses a model-based approach where an explicit 3D wireframe model of the face is constructed. In the first frame of the im-

age sequence, landmark facial features such as the eye corners and mouth corners are selected interactively. The generic face model is then warped to fit the selected facial features. The face model consists of 16 surface patches embedded in Bézier volumes. Once the model is constructed and fitted, head motion and local deformations of the facial features such as the eyebrows, eyelids, and mouth can be tracked. First, the 2D image motions are measured using template matching between frames at different resolutions. Image templates from the previous frame and from the very first frame are both used for more robust tracking. The measured 2D image motions are modeled as projections of the true 3D motions onto the image plane. In the first class of classifiers, they built SNoW (Sparse Network of Windows), SNoW-NB (SNoW-Naive-Bayes (NB) classifiers), NB-Gaussian (Naive-Bayes Gaussian classifiers), NB-Cauchy (Cauchy Naive Bayes classifiers), and TAN (Tree-Augmented-Naive Bayes classifiers). In another approach, they used temporal information displayed in the video to discriminate different expressions, with Hidden Markov Models (HMM). They tested all the approaches for persons in the database and new persons forming the expressions. For persons already in the database, the average facial expression recognition accuracies were 78.53%, 86.45%, 79.36%, 80.05% and 83.31%, by using SNoW, SNoW-NB, NB-Gaussian, NB-Cauchy and TAN classifiers, respectively, whereas the average accuracy was 78.49% with Single HMM and 82.46% with Multilevel HMM. For persons not in the database, the average facial expression recognition accuracies were 57.69%, 61.31%, 58.94%, 63.58% and 65.11%, by using SNoW, SNoW-NB, NB-Gaussian, NB-Cauchy and TAN classifiers, respectively, whereas the average accuracy was 55% with Single HMM and 58% with Multilevel HMM.

Bartlett et al. [170] developed a facial expression classification system to be used

with computer animated agents and robots for ‘face to face communication’. The system automatically detects front-view faces in the video stream and codes them with respect to seven dimensions in real time which represent the six basic emotions, plus the ‘neutral’. The face finder employs a cascade of feature detectors trained with boosting techniques. The expression recognizer receives image patches located by the face detector. A Gabor representation of the patch is formed and then processed by a bank of SVM classifiers. A novel combination of Adaboost and SVM’s enhances performance. The system was tested on the Cohn-Kanade dataset [112] of posed facial expressions. The real time system has been deployed in the Aibo robot and the RoboVie robot.

Busiu, Kotsia and Pitas [184] investigated the facial expression recognition task in cases where the face is partially occluded (e.g. a person wears glasses or a mouth mask). They worked with the six basic facial expressions and their main goal was to find the part of the face that contains sufficient information in order to correctly classify these six expressions. (1) the Japanese female facial expression (JAFFE)[110], where ten expressers posed three or four examples of each of the six basic facial expressions (anger, disgust, fear, happiness, sadness, surprise) plus neutral pose, for a total of 213 images of facial expressions (2) the Cohn-Kanade AU-coded facial expression database [112] that contains single or combined action units. In the images of these two databases, they superimposed a black rectangle around the eyes and mouth regions to occlude them partially. Each image from the database was convolved with a set of Gabor filters having various orientations and frequencies. The new feature vectors are classified by using a maximum correlation classifier and the cosine similarity measure approaches. They found that, overall, the facial expression recognition

method provides robustness against partial occlusion, the classification accuracy only decreasing from 89.7 % (no occlusion) to 84 % (eyes region occlusion) and 83.5 % (mouth region occlusion) for the first database and from 94.5 % (no occlusion) to 91.5 % (eyes region occlusion) and 87.2 % (mouth region occlusion) for the second database, respectively.

Hernández et al. [185] presented an illumination independent approach for facial expression recognition based on long wave infrared imagery. Until now, all the studies we have presented for facial expression recognition are designed considering the visible spectrum and consist of three major steps: (1) Region of Interest Selection, (2) Feature Extraction, and (3) Image Classification. This makes the recognition process not robust enough to be deployed in poorly illuminated environments. In this study, a Visual Learning approach based on Evolutionary Computation is proposed, which solved the first two tasks, mentioned above, simultaneously using a single evolving process. The first task consisted of the selection of a set of suitable regions where the feature extraction is performed. The second task consisted of tuning the parameters that defines the extraction of the Gray Level Co-occurrence Matrix used to compute region descriptors, as well as the selection of the best subsets of descriptors. The output of these two tasks is used for classification by a SVM classifier. A data-set of thermal images with three different expression classes was used to validate the performance. They used 33 images for ‘surprise’ emotion, 26 images for ‘happy’ emotion and 33 images for ‘angry’ emotion, in total 92 images for training/validation of the classifier. The experimental results showed 77% of accuracy of the three expressions. Although this is a quite novel approach, the system’s performance was only measured for only three classes of emotions, which are very discriminative. Also, a major

drawback is that a specific, more advanced hardware is required, which is a thermal camera.

Hammal et al. [186] proposed a method for facial expression classification, which is based on the Transferable Belief Model (TBM) framework. This fusion method is well suited for the problem of facial expression classification: this model facilitates the integration of a priori knowledge and can deal with uncertain and imprecise data which could be the case with data measures resulting from video-based segmentation algorithm. In addition, it is able to model intrinsic doubt which can occur between facial expressions in the recognition process. It allows the classification of different expressive states like ‘pure’ expression and allows the doubt between pairs of expressions. Their emotion classes were the six basic emotions plus the neutral. The proposed classifier relied on data coming from a contour segmentation technique, which extracted an expression skeleton of facial features (mouth, eyes and eyebrows) and derived simple distance coefficients from every face image of a video sequence. The characteristic distances were fed to a rule-based decision system that relies on the TBM and data fusion in order to assign a facial expression to every face image. They compared their system’s performance with the results given by human classifiers. In order to do this, they conducted an empirical study, where they presented the skeleton images corresponding to contours of permanent facial features to 60 subjects and asked them to classify by means of emotions. Humans achieved a 60% correct classification rate, whereas this rate was further improved to 80% when the original image was also showed to them. For the TBM method, they identified that there are some emotions that they can be combined together in terms of facial expression, for this reason, besides the emotion classes, they created 2 more classes which corre-

sponded to ‘Joy and/or Disgust’ and ‘Surprise and/or Fear’, and after classification, these images will further processed and re-classified. The system showed good results in recognizing some of the basic emotions plus the two classes mentioned above.

Kim et al. [187] extended their face detection system in order to perform facial expression recognition. They first detect the face using the AdaBoost face detection algorithm proposed by Viola and Jones [130], which guarantees real-time computation. If an initial face is detected, a face search window is set to reduce the searching region on the whole image and the system locates the face within the window. After face detection, face tracking is operated using the mean shift algorithm. Whenever a face region is extracted, first, a determination of whether the face is a front-view face is performed. After face detection, if the size of the face is sufficient for facial expression recognition (in this case 50 x 50 pixels), two pre-processing algorithms are applied to the input image: (1) illumination, and (2) geometry normalization, in order to reduce false recognition rates caused by illumination changes and size changes, respectively. For facial expression data extraction, the extended the form of the 3 x 1 AdaBoost rectangle feature to all possible rectangle features in a 3 x 3 matrix form, which included: 2 two-rectangle features, 6 three-rectangle features, 5 four-rectangle features, 58 six-rectangle features, and 249 nine-rectangle features for a total of 320 rectangle features. In order to reduce the computation time for selecting weak classifiers, the best five-rectangle features for each facial expression were chosen from among the 320 rectangle features. Each rectangle feature is selected by the AdaBoost algorithm, then the error rate is measured and the top five-rectangle features, that is those which have the lowest error rate among the total 320 rectangle features, are selected. The classification is made by applying the top five features for each expression

to the input image.

Wang et al. [225] were also based their study on the AdaBoost face detection algorithm proposed by Viola and Jones [130]. Like Kim et al. [187], they first detect the face using the AdaBoost face detection algorithm proposed by Viola and Jones [130]. Afterwards, they extract three key points of human face: the pair of eyes and the mouth center, using a Simple Direct Appearance Model (SDAM) method [228] based on the texture. A weak classifier pool of simple features should be configured. The weak classifier's construction was based on the Haar feature, which is a kind of simple rectangle feature proposed by Viola and Jones [130], and can be calculated very fast through the integral image. For each Haar feature, one weak classifier is configured. Their method was tested with by using the JAFFE database [110] and was compared to SVM classifiers and demonstrated an average accuracy of 92,4% in 0.11 ms in correctly classifying the expressions, in comparison to SVM classifiers which demonstrated 91,6% accuracy in 28.7 ms. In cases where subject was already in the database, the accuracy was 98,9%.

Liang et al. [188] developed a facial expression recognition system which was based on Supervised Locally Linear Embedding (SLLE). The system consists of three modules: face detection, feature extraction with SLLE and classification. In face detection module, two independent characteristics, skin color characteristic and motion characteristic are used to detect face region, and a trained Support Vector Machine (SVM) classifier is used to verify candidate regions. In feature extraction module, SLLE, a supervised learning algorithm that can compute low dimensional, neighborhood-preserving embeddings of high dimensional data is used to reduce data dimension and extract features. In classification module, minimum-distance classifier is used to

recognize different expressions. Their method was tested with by using the JAFFE database [110] and was compared to PCA-based method. The results showed up to 95% accuracy for subjects already in the database and up to 85% for new subjects in the database.

Zhou et al. [223] used Hidden Markov Models (HMM) to build a real - time facial expression recognition system which could be used in a role playing game. Their work was based on the HMM proposed by Nefian et al. for face recognition [229]. First, they detect the face using the AdaBoost face detection algorithm proposed by Viola and Jones [130]. Second, face alignment is used in facial expression environment. The embedded HMM uses observation vectors that are composed of two-dimensional Discrete Cosine Transform (2D-DCT) coefficients opposite to previous HMM approaches which use pixel intensities to form the observation vectors. The system was tested for the classification of five emotions/states, namely: ‘normal’, ‘laugh’, ‘anger’, ‘sleep’ and ‘surprise’. The results showed 92% accuracy for subjects already in the database and up to 84% for new subjects in the database.

Xiang et al. [189] proposed a fuzzy spatio-temporal approach for real time facial expression recognition in video sequences. The proposed system first employed the Fourier transform to convert a facial expression sequence of images (displaying one expression) from the spatio-temporal domain into the spatio-frequency domain. This was followed by a fuzzy C means classification [230] for expression representation. Their system classified the input images in the six basic emotion classes. Unknown input expressions are matched to the models using the Hausdorff distance to compute dissimilarity values for classification. Since the proposed algorithm used Fourier transform, it is robust, in terms of lighting changes during image acquisition. Also,

by using fuzzy C means classification, quantified interpretation labels are assigned with degrees varied from 0 to 1. Since some expressions might look alike, these degrees represent the membership of the input image to the respective class, rather than using hard boundaries between the expressions. Moreover, the system needs only the positions of two eyes that can be detected accurately in current technology. Since fewer landmarks are needed, the system is less affected by the inaccuracy of facial features detection. The system was trained and tested using the Cohn-Kanade AU-Coded Facial Expression Database [112], as follows: they used the data of all but one subject as training data, and tested on the sequences of the subject that was left out. Each input sequence was compared with model for each of the six basic expressions. A dissimilarity score was computed to measure the dissimilarity degree for each comparison. The results showed 88,8% accuracy for new subjects in the database.

Ma and Khorasani [226] built a neural network-based facial expression recognition system. In their proposed technique, the 2-dimensional discrete cosine transform (2-D DCT) is applied over the entire difference face image for extracting relevant features for recognition purpose. In order to find a proper network size, they propose to use the constructive one-hidden-layer feed forward neural networks (OHL-FNNs). The constructive OHL-FNN will obtain in a systematic way a proper network size which is required by the complexity of the problem being considered. Furthermore, the computational cost involved in network training can be considerably reduced when compared to standard back-propagation (BP) based FNNs. So, in this study, the lower frequency 2-D DCT coefficients obtained are then used to train a constructive OHL-FNN. The proposed technique is applied to a database consisting of images of 60 men, each having 5 facial expression images (neutral, smile, anger, sadness, and

surprise). Images of 40 men are used for network training, and the remaining images are used for generalization and testing. Confusion matrices calculated in both network training and testing for 4 facial expressions (smile, anger, sadness, and surprise) are used to evaluate the performance of the trained network. The results showed 93,75% accuracy for new subjects in the database.

Pardas et al. [224] were based on the modeling of the expressions by means of Hidden Markov Models for building their facial expression recognition system. The observations used to create the models are the MPEG-4 standardized facial animation parameters (FAPs). The FAPs of a video sequence are first extracted and then analyzed using semi-continuous Hidden Markov Model. They trained and tested their system for recognizing the six basic emotions using the Cohn-Kanade AU-Coded Facial Expression Database [112]. The database consisted of 90 subjects depicting the six expressions. They trained the system using all subjects except one, and, then, tested the recognition rate with the subject that had not participated in the training process. This process was repeated 90 times, as the number of subjects. Results showed 84% recognition rate. They also tested their system in ‘talking videos’, where the subject was talking while forming the expression. The recognition rate in this case was 64%.

Kobayashi et al. [231] proposed a method of expression learning by imitating the process of a baby’s learning process. A baby cannot know what an expression means but he/she can be affected by the action that people do to him or her, and then he/she remembers this facial expression. In their system, a robot starts learning facial expression by recognizing human actions. The system detects the face regions and extracts the facial features using four direction features: horizontal, vertical and

diagonally in both directions. For face detection they used two methods: (1) skin-color information and (2) template matching. Since, in previous works, they had found skin-color information-based face detection weak in some environments, they adopted template matching methods for face detection. The system was trained and tested for three emotions: ‘anger’, ‘happiness’ and ‘neutral/normal’. They collected the data 12 days in several locations and the data was taken 3 times per day, in morning, afternoon, and evening. For each learning time, they got 30 frames for each expression. The number of images of experimental data totals 3,240 images (3 expressions x 12 days x 3 times x 30 frames). They used 6 days of the data for learning (half of data). The other were used as unknown data for testing. The system achieved a recognition rate of 97,5%, for a single subject.

Abboud and Davoine [232] proposed bilinear factorization based representations for facial expression recognition and compared this method with previously investigated methods such as linear discriminant analysis and linear regression. In order to perform facial expression recognition, they used a test set of 108 unknown face image showing each of the seven basic facial expressions and ran Active Appearance Mode (AAM) optimization to extract the corresponding appearance parameters. Their aim was to extract from each appearance vector a subset of relevant parameters that represent facial expression information. Towards this goal, searching for the vectors that best discriminate among classes, the compared Linear Discriminant Analysis (LDA) techniques and Bilinear Factorization techniques. The predictive learning is performed on a training set containing 70 images which represent 10 different persons showing each of the 7 basic facial expressions. The correct recognition rate, when using LDA, for 108 unknown test persons was 67.59%. LDA was also tested for a

training set containing 26 images per class and achieved an optimal correct recognition rate of 84.34%. When trained with the same training set of 70 images, the asymmetric bilinear factorization expression classifier achieved a correct recognition rate of 83,33%, for 108 unknown test persons.

All the aforementioned methods are evaluated based on the requirements set by Pantic and Rothkrantz [1] and this evaluation is summarized in the following Table 3.11.

Based on the previous survey, many studies included the recognitions of different facial expressions or facial action units. Moreover, many studies were centered in the development of a user dependent facial expression recognition system. This task is quite easier rather than a facial expression recognition system which would be able to perform well regardless of the subject. The evaluation of the aforementioned studies, based on those criteria, are summarized in Table 3.12 and Table 3.13, where the input data are static images or image sequences, respectively.

3.3.3 Section Summary - Results

Some of the aforementioned methods have achieved good results, but there are some drawbacks. Firstly, the majority of these methods use some of the facial expression databases described in Section 3.1, mostly the JAFFE, the Cohn-Kanade and the MMI. However, as research shows, cultural exposure increases the chances of correct recognition of facial expressions indicating cultural dependence ([233], [38], [234]). This point has become more noticeable for the expressions formed by Greek people, in following Chapter 5, where we present two different questionnaires. In the first questionnaire we used images of people forming facial expressions gathered from

Table 3.11 : Review of the facial expression approaches (Recent Years)

| Reference | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|---|---|-----|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|
| Input media: Static images | | | | | | | | | | | | | | | | | | | | |
| Busiu [184] | x | 3 | x | ✓ | ✓ | x | x | ✓ | ✓ | ✓ | x | ✓ | ✓ | x | x | 7 | x | - | - | ✓ |
| Hernández [185] | ✓ | - | ✓ | x | ✓ | x | ✓ | ✓ | - | ✓ | x | x | x | x | x | 3 | x | - | - | - |
| Kim [187] | x | ✓ | ✓ | ✓ | ✓ | x | ✓ | ✓ | - | ✓ | x | - | - | x | x | 7 | x | - | - | ✓ |
| Wang [225] | ✓ | - | x | x | ✓ | x | ✓ | ✓ | - | ✓ | x | ✓ | - | x | x | 7 | x | - | - | ✓ |
| Liang [188] | x | x | x | x | ✓ | x | ✓ | ✓ | x | ✓ | x | ✓ | - | x | x | 7 | x | - | - | - |
| Ma [226] | x | x | x | x | ✓ | x | x | ✓ | x | ✓ | x | - | - | x | x | 4 | x | - | - | x |
| Kobayashi [231] | x | age | x | x | ✓ | x | ✓ | ✓ | x | ✓ | x | - | - | x | x | 3 | x | ✓ | - | ✓ |
| Abboud [232] | x | x | x | x | ✓ | x | x | ✓ | x | ✓ | x | - | - | x | x | 7 | x | - | - | x |
| Input media: Image Sequences (Video) | | | | | | | | | | | | | | | | | | | | |
| Cohen [175] | ✓ | x | x | x | ✓ | ✓ | ✓ | ✓ | - | ✓ | x | ✓ | ✓ | x | x | 7 | x | - | - | ✓ |
| Bartlett [170] | ✓ | 3 | x | x | ✓ | ✓ | ✓ | ✓ | - | ✓ | x | ✓ | - | x | x | 7 | x | - | - | ✓ |
| Hammal [186] | x | 3 | x | x | ✓ | x | x | ✓ | - | ✓ | x | ✓ | ✓ | x | x | 7 | x | - | - | - |
| Zhou [223] | ✓ | - | x | x | ✓ | x | ✓ | ✓ | - | ✓ | x | - | - | x | x | 5 | x | - | - | ✓ |
| Xiang [189] | ✓ | 3 | ✓ | x | ✓ | x | - | ✓ | ✓ | ✓ | x | - | - | x | x | 6 | x | ✓ | - | ✓ |
| Pardas [224] | x | 3 | x | x | ✓ | x | x | ✓ | x | ✓ | x | x | x | x | x | 6 | x | - | - | x |

✓ : 'Yes' x : 'No' '-' : Not available

the web, whereas in the second we used facial expression images from our own facial expression database. The difference between the error rates of the two different questionnaires is quite noticeable. In our system we used our own facial expression

Table 3.12 : Performance evaluation and generalization of recent systems - 1

| Study reference | Success Rate (User dependent) | Success Rate (User independent) | User independent tests? | Number of emotion classes |
|-----------------------------------|--|---------------------------------|-------------------------|-------------------------------|
| Input media: Static images | | | | |
| Busiu [184] | 88,96% (no occlusion), 83,46%(mouth occlusion), 85,45% (eye occlusion) | - | - | 6 b.em. + 'neu' |
| Hernández [185] | - | 77% | ✓ | 3 |
| Kim [187] | n/a | n/a | ✓ | 6 b.em. + 'neu' |
| Wang [225] | 98,9% | 92,46% | ✓ | 6 b.em. + 'neu' |
| Liang [188] | 95% | 85% | ✓ | 6 b.em. + 'neu' |
| Ma [226] | - | 93,75% | ✓ | 4: 'joy', 'sur', 'ang', 'sad' |
| Kobayashi [231] | 93,75% | X | - | 3: 'ang', 'joy', 'neu' |
| Abboud [232] | | 83,33% | ✓ | 6 b.em. + 'neu' |

Legend:

✓ : 'Yes'

- : 'No' / 'Not available'

n/a : 'Not available'

Emotions/Expressions : 'neu': 'neutral', 'ang': 'anger', 'scr': 'scream', 'sle': 'sleep', 'sur': 'surprise', 'lau': 'laughter'

b.em. : 6 Basic Emotions ⇒ anger, 'disgust', 'fear', 'happiness', 'sadness' and 'surprise'

database which consists of Greek people forming seven different expressions. Secondly, the methods which attempt to classify the expression in discrete emotions, usually assume six emotions, namely: 'neutral', 'anger', 'happiness', 'sadness', 'disgust' and

Table 3.13 : Performance evaluation and generalization of recent systems - 2

| Study reference | Success Rate (User dependent) | Success Rate (User independent) | User independent tests? | Number of emotion classes |
|---|-------------------------------|--|-------------------------|--------------------------------------|
| Input media: Image Sequences (Video) | | | | |
| Cohen [175] | 80,66% | 59,95% | ✓ | 6 b.em. |
| Bartlett [170] | n/a | n/a | n/a | 6 b.em. + 'neu' |
| Hammal [186] | - | - | - | 6 b.em. + 'neu' |
| Zhou [223] | 92% | 84% | ✓ | 5: 'neu', 'lau', 'ang', 'sle', 'sur' |
| Xiang [189] | - | 88,8% | ✓ | 6 b.em. |
| Pardas [224] | - | 84%(Cohn Database), 64%(subject is talking) | - | 6 b.em |

Legend:

✓ : 'Yes'

- : 'No' / 'Not available'

n/a : 'Not available'

Emotions/Expressions : 'neu': 'neutral', 'ang': 'anger', 'scr': 'scream', 'sle': 'sleep', 'sur': 'surprise', 'lau': 'laughter'

b.em. : 6 Basic Emotions ⇒ anger, 'disgust', 'fear', 'happiness', 'sadness' and 'surprise'

'fear'. Our system has been designed to be implemented in human computer interaction. Again, based on our studies, which are pointed out in following Chapter 5, the emotions that are most commonly observed during a typical human computer interaction session are namely: 'neutral', 'anger', 'happiness', 'sadness', 'disgust' and 'bored-sleepiness' which are the categories used by our system.

4

Face Image Databases

Every man builds his world in his own image. He has the power to choose, but no power to escape the necessity of choice.

—Ayn Rand (1905–1982)

OUR review on previous facial expression databases, as described on Section 3.1, led us to the assumption that we must create our own facial expression database [3, 5]. Our work to the creation of two different databases:

1. The database of low quality images: this database consists of many subjects, depicting many expressions, but the image quality is quite low as we used web cameras to acquire the data
 2. The database of high quality images: this database consists again of many subjects, depicting the expressions recognized by our system, and the image quality
-

is quite high as we used digital cameras to acquire the data

In this chapter, we present thoroughly these two databases and the means of developing them.

4.1 The Database of Low Quality Face Images (DBLQFI)

THE first database was created by photographing individuals aged 19-35 years old while they were forming various expressions. To acquire image data, we built a three-camera system, as in Fig. 4.1. Specifically, three identical cameras of 320-by-240 pixel resolution were placed with their optical axes on the same horizontal plane and successively forming 30-degree angles. Subjects were placed in front of the camera configuration and formed facial expressions, which were simultaneously photographed by the three cameras.

The database consisted of 300 subjects: 233 male and 67 female. All participants and expression-forming subjects were Greek, so they were used to the Greek culture and the Greek ways of expressing emotions. The study participants aged 19 to 45 years old and their majority were either under-graduate or graduate students in the University of Piraeus. A small number of the participants were employees of the University. They were asked to form the 10 following expressions/emotions: *'neutral, 'happiness', 'sadness', 'surprise' 'boredom-sleepiness', 'disappointment', 'scream', 'anger', 'disgust' and 'talking expression'.*

The resulting database consisted of **300 by 10 by 3 = 9000 color images**. The database details are summarized in the following Table 4.1, while some samples of the database are shown in Table 4.2 at the end of this Chapter.

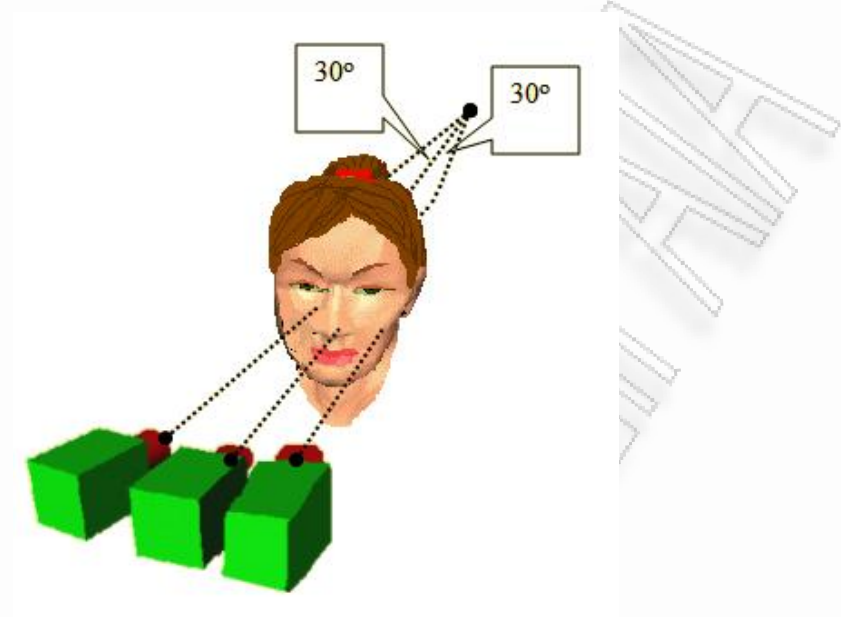


Figure 4.1 : Three Camera Configuration

4.2 The Database of High Quality Face Images (DB-HQFI)

THE first efforts in building a facial expression database resulted to the ‘Database of Low Quality Face Images (DBLQFI)’, which, although complete in terms of subjects and facial expressions, contains images of such low quality that is not useful for the purpose of building our system. Specifically, because of the low quality, many details of the expressions (such as the texture of the skin) were not clearly distinguishable in these images. Based on these assumptions, DBLQFI could only be used for testing the generalization of the system in cases of low quality images. This led us to the development of a new database of high quality images captured with

digital cameras.

The process of developing the new database consisted of three steps:

1. Observation of the user's reactions during a typical human-computer interaction session: From this step, we concluded that the facial expressions corresponding to the “**neutral**”, “**happy**”, “**sad**”, “**surprised**”, “**angry**”, “**disgusted**” and “**bored-sleepy**” psychological states arose quite commonly in human-computer interaction sessions and, thus, form the corresponding classes for our classification task.
2. Data acquisition: We created our own database of facial expressions, by photographing individuals aged 19-35 years old while they were forming various expressions. To ensure spontaneity, each subject was presented with pictures on a screen behind the camera. To acquire image data, we built a two-camera system, as in Fig. 4.2. Specifically, two identical cameras of 1600-by-1200 pixel resolution were placed with their optical axes on the same horizontal plane and successively forming 30-degree angles. Subjects were placed in front of the camera configuration and formed facial expressions, which were simultaneously photographed by the two cameras.

These pictures were expected to generate those emotional states that would map on the subject's face as the desired facial expression. For example, to have a subject assume a ‘happy’ expression, we showed him/her a picture of funny content. We photographed the resulting facial expression and then asked him/her to classify this expression. If the image shown to him/her had resulted in the desired facial expression, the corresponding photographs were saved and

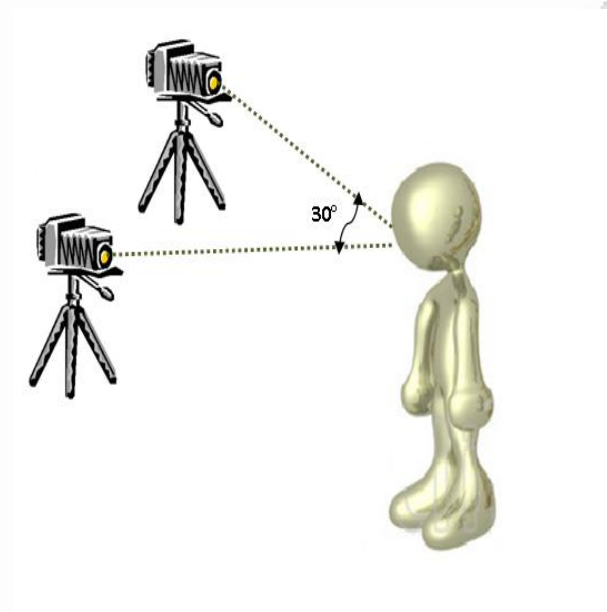


Figure 4.2 : Two Camera Configuration

labeled; otherwise, the procedure was repeated with other pictures. The final dataset consists of the images of 250 different individuals, each forming the seven expressions: 'neutral', 'happy', 'sad', 'surprised', 'angry', 'disgust', 'bored-sleepy' and 'screaming'.

The resulting dataset consisted of 2000 images of each view, which is the result of 250 persons forming 7 different expression, a total of **4000 high quality images**. The database details are summarized in the following Table 4.3.

Studying this dataset, we identified differences between the ‘neutral’ expression of a model and its deformation into other expressions. We quantified these differences into measurements of the face (such as size ratio, distance ratio, texture, or orientation), so as convert pixel data into a higher-level representation of shape, motion, color, texture and spatial configuration of the face and its components.

Furthermore, we identified that there are some differences between the ways that people express themselves in the occurrence of same emotions. For example, as we observe in Table 4.4 in the occurrence of the ‘Boredom-Sleepiness’ emotions some people tend to yawn, others to close their eyes and others to slightly tilt their head left or right, as in the case of the second subject. Specifically, we locate and extract the corner points of specific regions of the face, such as the eyes, the mouth and the brows, and compute their variations in size, orientation or texture between the neutral and some other expression. This constitutes the feature extraction process and reduces the dimensionality of the input space significantly, while retaining essential information of high discrimination power and stability. In order to validate these facial features and understand how they are used by humans to deduce someone’s emotion from his/her facial expression, we developed questionnaires where the participants were asked to determine which facial features helped them in the expression recognition/classification task.

3. Questionnaires – empirical study by observers: In order to understand aspects of the process of facial expression recognition by human observers and set target

error rates for automated systems, we conducted two relevant empirical studies based on two corresponding questionnaires, as described below. The first study was only preliminary and was based on a short ('preliminary') questionnaire. The purpose of this study was to obtain an overall idea and identify the general aspects of the facial expression recognition process in humans. The images used in this preliminary study were gathered from the Web and existing facial expression databases. The lack of a complete facial expression database, containing a sufficient number of all seven expressions of interest to us, required us to create our own database of better quality images [3, 5]. We also developed a 'detailed' questionnaire, as described below, which used images of our own database. Then, we conducted a second, more detailed empirical study. Results from both empirical studies are presented in this paper and conclusions are drawn.

A sample of this database of the two persons forming the 7 expression can be seen in Table 4.4

Table 4.1 : Low quality Database

| Emotions/Expressions | |
|-----------------------------|--|
| Number of emotions | 10 |
| Emotion Classes | 'neutral', 'happiness', 'sadness', 'surprise', 'boredom-sleepiness', 'disappointment', 'scream', 'anger', 'disgust' and 'talking expression' |
| Subjects | |
| Number of subjects | 300 |
| Subjects' culture | Greek |
| Subjects' age | 19-45 |
| Subjects' sex | Male: 77,66% and Female: 22,44% |
| Image Quality | |
| Image resolution | 320-by-240 |
| Image color | RGB |
| Data acquisition | |
| Camera | Webcamera |
| Front view | ✓ |
| Right Side view | ✓ |
| Left Side view | ✓ |
| Total images: 9000 | |













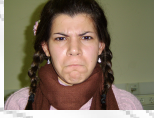















Table 4.2 : Sample images of our low quality facial expression database

| Emotions | Views | | |
|--------------------|---|--|---|
| | Left Side | Front | Right Side |
| Neutral |  |  |  |
| Happiness |  |  |  |
| Sadness |  |  |  |
| Anger |  |  |  |
| Surprise |  |  |  |
| Disgust |  |  |  |
| Boredom-Sleepiness |  |  |  |
| Disappointment |  |  |  |
| Scream |  |  |  |
| Speaking |  |  |  |

Table 4.3 : Low quality Database

| Emotions/Expressions | |
|-----------------------------|--|
| Number of emotions | 8 |
| Emotion Classes | 'neutral, 'happiness', 'sadness', 'surprise' 'boredom-sleepiness', 'scream', 'anger' and 'disgust' |
| Subjects | |
| Number of subjects | 250 |
| Subjects' culture | Greek |
| Subjects' age | 19-45 |
| Subjects' sex | Male: 77,66% and Female: 22,44% |
| Image Quality | |
| Image resolution | 1600-by-1200 |
| Image color | RGB |
| Data acquisition | |
| Camera | Webcamera |
| Front view | ✓ |
| Right Side view | ✓ |
| Left Side view | x |
| Total images: 4000 | |

Table 4.4 : Sample images of our facial expression database

| Emotions | First Subject | | Second Subject | |
|--------------------|---|--|---|---|
| | Front View | Side View | Front View | Side View |
| Neutral |  |  |  |  |
| Happiness |  |  |  |  |
| Sadness |  |  |  |  |
| Anger |  |  |  |  |
| Surprise |  |  |  |  |
| Disgust |  |  |  |  |
| Boredom-Sleepiness |  |  |  |  |

5

Empirical Studies on Emotion Recognition

We know too much and feel too little. At least we feel too little of those creative emotions from which a good life
spring.

—*Bertrand Russell (1872–1970)*

IN our attempts to understand the facial expression recognition task and set the requirements for our facial expression recognition system, we conducted two empirical studies involving human subjects and observers [3, 4, 5, 6]. The first study, as described in Section 5.1, was simpler than the second and aimed at setting an error goal for our system. We used images from facial expression databases gathered from World Wide Web [109, 112] and asked people to map the emotion based on the subject’s expression. Our second empirical study, which is described in Section 5.1, was more complicated and aimed not only at an error goal, but also, at understanding how facial expression recognition works in humans. In this study, we used our own

facial expression database [3, 4, 5, 6]. The two studies are described in detail in this Chapter, whereas in the final Section 5.4 we draw conclusions, from these studies.

5.1 Preliminary Questionnaires

TO obtain a preliminary idea of facial expression classification by humans, we developed a preliminary questionnaire in which we asked 300 study participants to classify the facial expressions that appeared in 36 images. Each participant could choose from 11 of the most common facial expressions, such as: ‘angry’, ‘happy’, ‘neutral’, ‘surprised’, or specify any other expression that he/she thought appropriate. Next, the participant had to decide which emotion he/she thought that the facial expression indicated. Our dataset consisted of 3 subsets of images, typically depicted in Table 5.1, namely:

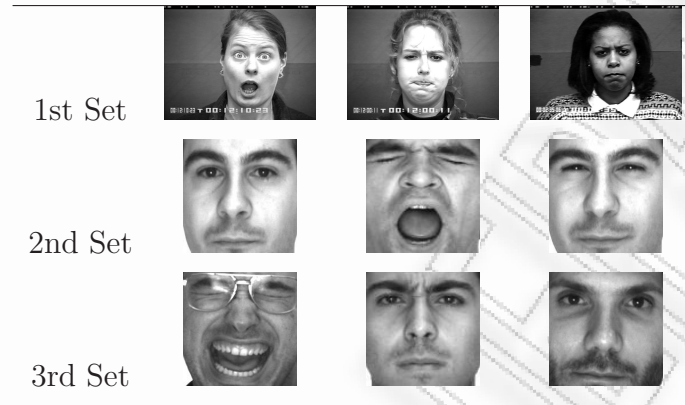
various images of individuals placed in a background and forming a facial expression,

a sequence of facial expressions of the same person without a background, and

face images of different persons without a background.

From the results of the questionnaire, we observed that the ‘surprised’ expression was the one most easily recognized, as the corresponding error rate of 22% was the lowest among all error rates. The ‘happy’ and ‘neutral’ expressions were recognized with corresponding error rates of 30% and 35%, respectively. The expression recognized with the highest difficulty was the ‘sad’ expression, as its corresponding error rate reached 88%. The ‘angry’ and ‘disappointed’ expressions had a corresponding error rate of 80% and 76%, respectively. These are summarized in Figure 5.1.

Table 5.1 : Typical face image subsets in our questionnaire



Clearly, the facial expression classification task in images is quite challenging. The reasons why humans could not achieve low classification error rates for specific expressions such as ‘angry’, ‘sad’ and ‘disappointed’, may be found in the fact that these expressions seem to differ significantly from person to person or some people may be too shy to form them clearly. This finding corroborates similar findings in previous psychological studies [14], which we discussed in Chapter 2, Section 2.2.2.

5.2 Newer (Detailed) Questionnaires

IN the newer questionnaires, we used images of subjects of the facial expression database we created ourselves [3, 4, 5, 6]. Our aim was to identify those facial features which help humans to classify a facial expression. Moreover, we wanted to know the degree to which it is possible to map a facial expression into an emotion. A third goal was to determine whether a human observer could recognize a facial expression from only portions of the face (e.g., eyes, mouth, etc.), as we expect the artificial classifier to do. Thus, we had to redesign the structure of the questionnaires.

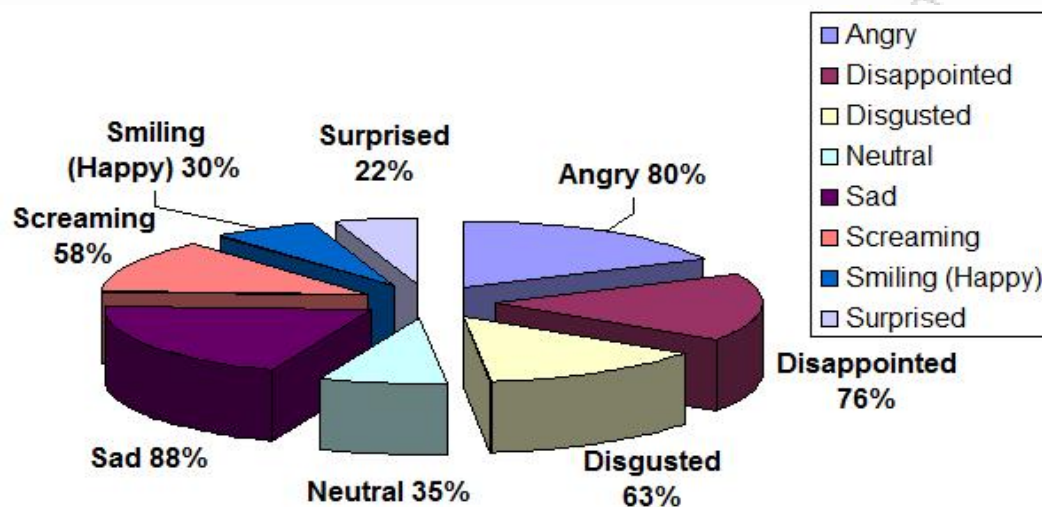


Figure 5.1 : Error rates in recognizing the expressions in our preliminary questionnaire

This led to detailed questionnaires, which are more significantly complex than the preliminary questionnaire and described next.

5.2.1 The detailed questionnaire structure

In order to understand how humans classify someone else's facial expression and set a target error rate for automated systems, we developed a questionnaire filled again by 300 study participants. These were not the same participants as those who filled the preliminary questionnaire.

Specifically, the questionnaire consisted of three parts:

1. In the first part of the questionnaire, each participant was asked to map into facial emotions the facial expressions that appeared in 14 images. Each participant could choose from the 7 common emotions 'angry', 'happy', 'neutral',

‘surprised’, ‘sad’, ‘disgusted’, ‘bored-sleepy’, or specify any other emotion that he/she thought appropriate. Next, the participant had to specify the degree (0-100%) of his/her confidence in the identified emotion. Finally, he/she had to indicate which features (such as the eyes, the nose, the mouth, the cheeks, etc.) had helped him/her make that decision. A typical print-screen of the first part of the questionnaire is depicted in Figure 5.2.

4a. What emotion does the image represent

Other...

4β. In what percent: %

4γ. Which Facial Features helped you understand the emotion?
(you can choose more than one)

Mouth Forehead texture

Eyes Texture between the brows

Shape of the face Texture of the cheeks

Other....

Figure 5.2 : The first part of the detailed questionnaire

2. In the second part of the questionnaire, each participant had to classify the emotion from portions of the face. Specifically, we showed the participant the ‘neutral’ facial image and an image of some facial expression of a subject. The latter image was cut into the corresponding facial portions, namely, the eyes, the mouth, the forehead, the cheeks, the chin and the brows. A typical print-screen of this part of the questionnaire is shown in Figure 5.3.

4a. What emotion does the image represent

Other...

4b. In what percent: %

4y. Which Facial Features helped you understand the emotion?
(you can choose more than one)

Mouth Forehead texture

Eyes Texture between the brows

Shape of the face Texture of the cheeks

Other....

Figure 5.3 : The second part of the detailed questionnaire

Again, each participant could choose from the 7 emotions or specify any other emotion that he/she thought appropriate. Next, the participant had to specify the degree (0-100%) of his/her confidence in the identified emotion. Finally, he/she had to indicate which features (such as the eyes, the nose, the mouth, the cheeks, etc.) had helped him/her make that decision.

3. In the third part of the questionnaire, we collected background information (e.g. age, interests, etc.) about the study participants. Additional information provided by the participants at this stage included:

- The level of difficulty of the questionnaire, with regards to the facial expression classification task
- Which emotion they considered the most difficult to classify
- Which emotion they considered the easiest to classify

- The degree (0-100%) to which an emotion maps into a facial expression

5.2.2 The observer and subject backgrounds

A total number of 300 participants participated in our study and filled up the detailed questionnaires. All participants and expression-forming subjects were Greek, so they were used to the Greek culture and the Greek ways of expressing emotions. The study participants aged 19 to 45 years old and their majority were either under-graduate or graduate students in the University of Piraeus, along with a small number of the participants were employees of the University.

5.3 Results from Statistical Analysis

5.3.1 Statistical Analysis per Expression

Angry

The 'angry' expression was recognized at an error rate of 23,86% in the first part of our detailed questionnaire, in which the entire facial image was depicted. The corresponding error rate was 30,30% in the second part of the questionnaire, in which only portions of the faces were depicted.

The participants mistook the 'angry' emotion mostly for the 'neutral' and the 'sad' expressions. The percentages to which the participants mistook the 'angry' emotion for some other emotion are shown in Figure 5.4 for both parts of the questionnaire. Specifically, as 'other' emotions, the participants indicated 'sceptical', 'confused', 'concerned', 'wondering', or 'suspicious', most of which are considered as 'negative' emotions. As the 'angry' emotion is considered as a negative one, one might

come to the conclusion that even though the specific emotion was not recognized, a broader class of perhaps similar emotions was recognized.

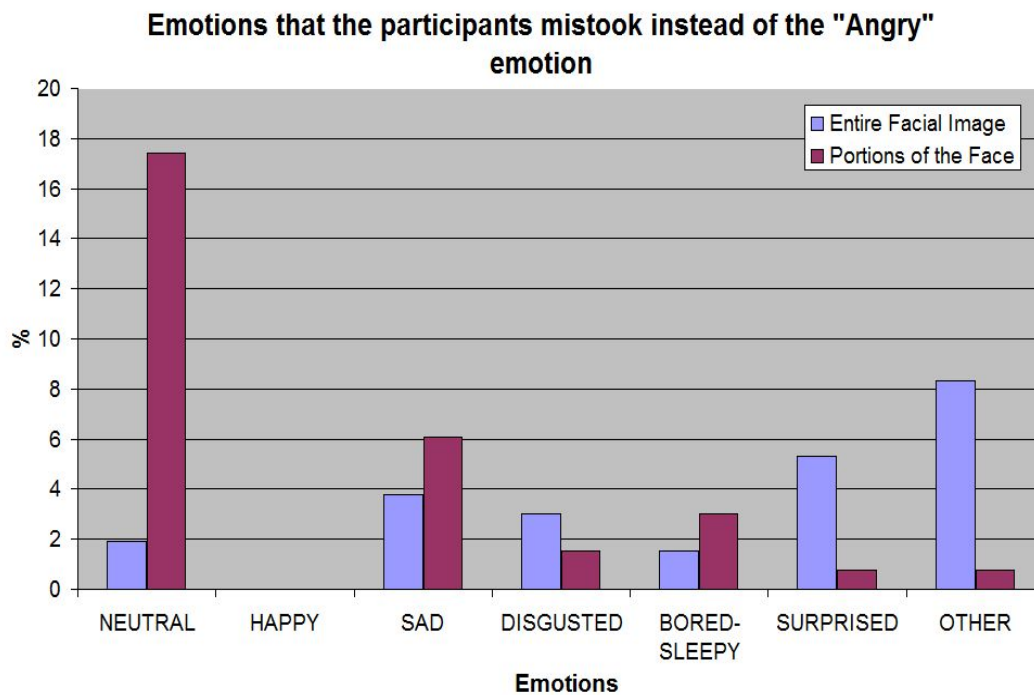


Figure 5.4 : Graph of the percentage to which the participants mistook the ‘angry’ emotion for other emotions

The facial portions that helped the participants recognize this expression were mostly the ‘eyes’, the ‘mouth’ and the ‘cheeks’. Specifically, the participants assigned percentages of significance of each facial portion as in Figure 5.5.

Finally, regarding the percentage to which the ‘angry’ expression maps the equivalent emotion, the participants’ answers are shown in Figure 5.6. Specifically, 26,24% of the participants indicated that it maps only 10% percent of the strength of the emotion, whereas 21,82% of the participants believe that the expression maps 60%

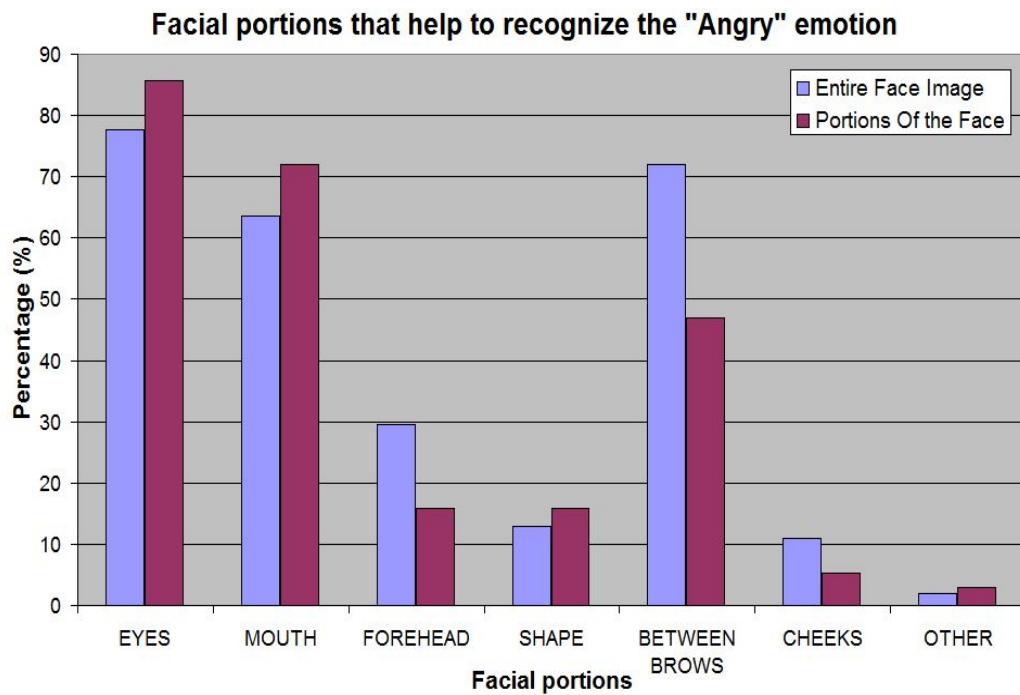


Figure 5.5 : Graph of the percentage to which the 'angry' expression maps the equivalent emotion, based on the correct answers of the participants

of the strength of the emotion. Moreover, the majority of the participants (67,4%) thought that the 'angry' expression maps more than 50% of the emotion, as opposed to 32,6% of the participants who thought that the 'angry' expression maps less than 50% of the emotion. This may again be in line with the fact that 'angry' is a negative emotion that people would tend to mask.

Bored - Sleepy

The 'bored-sleepy' expression was recognized at an error rate of 49,24% in the first part of our detailed questionnaire, in which the entire facial image was depicted. The

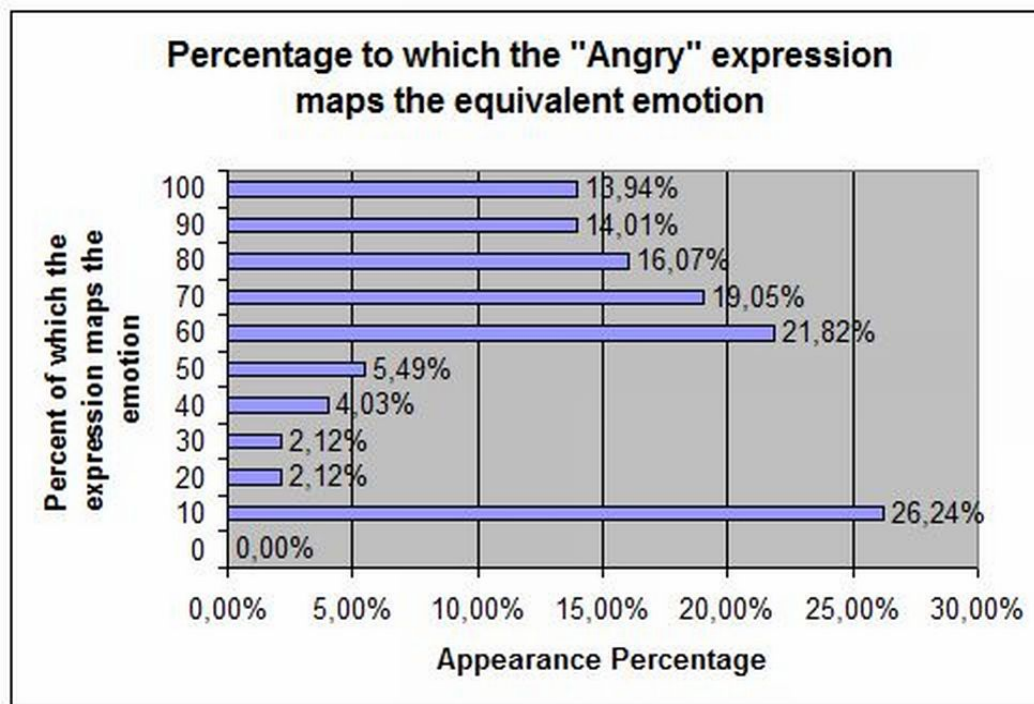


Figure 5.6 : Graph of the percentage to which the ‘angry’ expression maps the equivalent emotion, based on the correct answers of the participants

corresponding error rate was 21,96% in the second part of the questionnaire, in which only portions of the faces were depicted. The ‘angry’ expression was recognized at an error rate of 23,86% in the first part of our detailed questionnaire, in which the entire facial image was depicted. The corresponding error rate was 30,30% in the second part of the questionnaire, in which only portions of the faces were depicted.

The participants mistook the ‘bored-sleepy’ emotion mostly for the ‘sad’ expression. The percentages to which the participants mistook the ‘bored-sleepy’ emotion for some other emotion are shown in Figure 5.7 for both parts of the questionnaire. Specifically, as ‘other’ emotions, the participants indicated ‘sceptical’, ‘disappointed’,

and ‘distrustful’. Some of these emotions, especially, disappointment and the distrust may eventually lead to boredom.

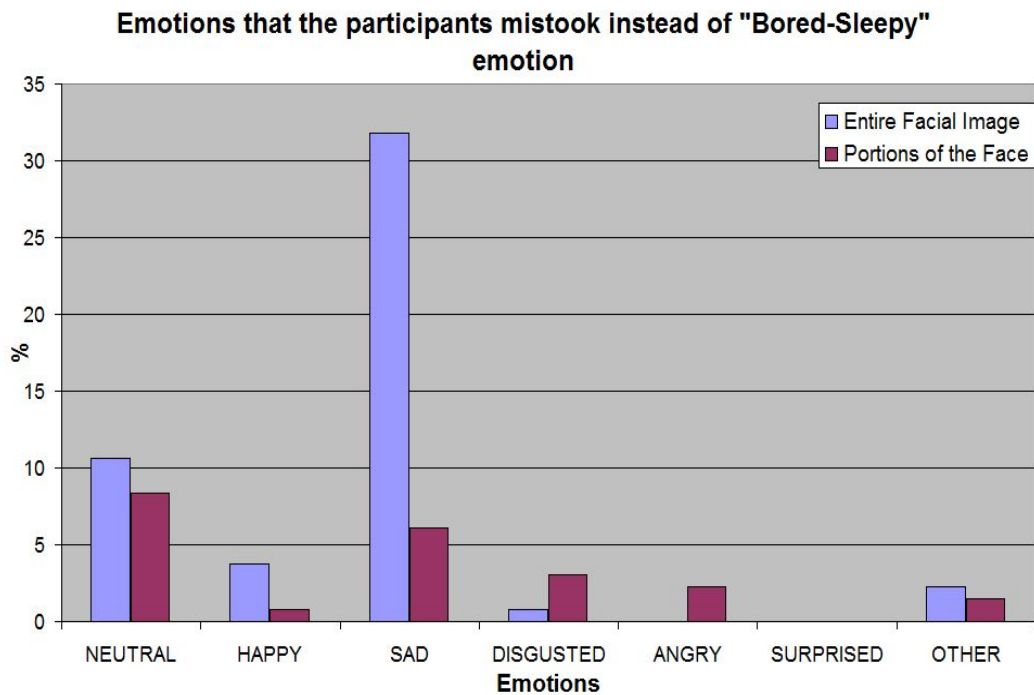


Figure 5.7 : Graph of the percentage to which the participants mistook the ‘bored - Sleepy’ emotion for other emotions

The facial portions that helped the participants recognize this expression were mostly the ‘eyes’, the ‘mouth’ and the ‘region between the brows’. Specifically, the participants assigned percentages of significance of each facial portion as in Figure 5.5.

Finally, regarding the percentages to which the ‘bored-sleepy’ expression maps the equivalent emotion, the participants’ answers are shown in Figure 5.9. Specifically, 24,95% of the participants indicated that it shows the 80% percent of the strength

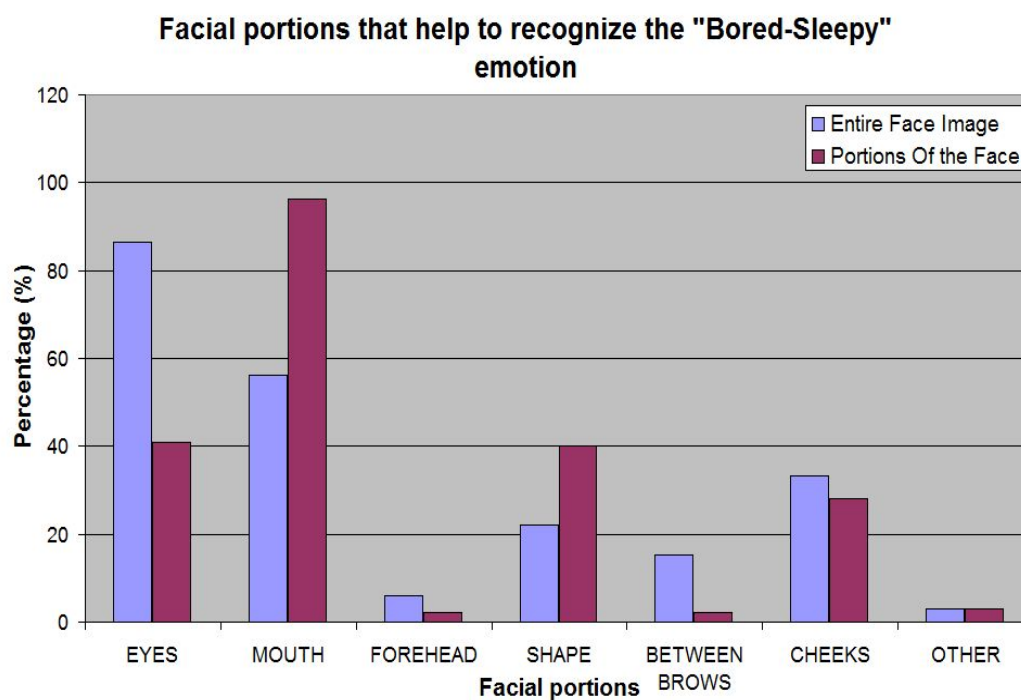


Figure 5.8 : Graph of the percentage to which the ‘bored - Sleepy’ expression maps the equivalent emotion, based on the correct answers of the participants

of the emotion. Moreover, the majority of the participants thought that the ‘angry’ expression maps more than 60% of the emotion.

Disgusted

The ‘disgusted’ expression was recognized at an error rate of 81,26% in the first part of our detailed questionnaire, in which the entire facial image was depicted. The corresponding error rate was 86,36% in the second part of the questionnaire, in which only portions of the faces were depicted. The participants mistook the ‘disgusted’ emotion mostly for the ‘happy’ expression. The percentages to which the participants

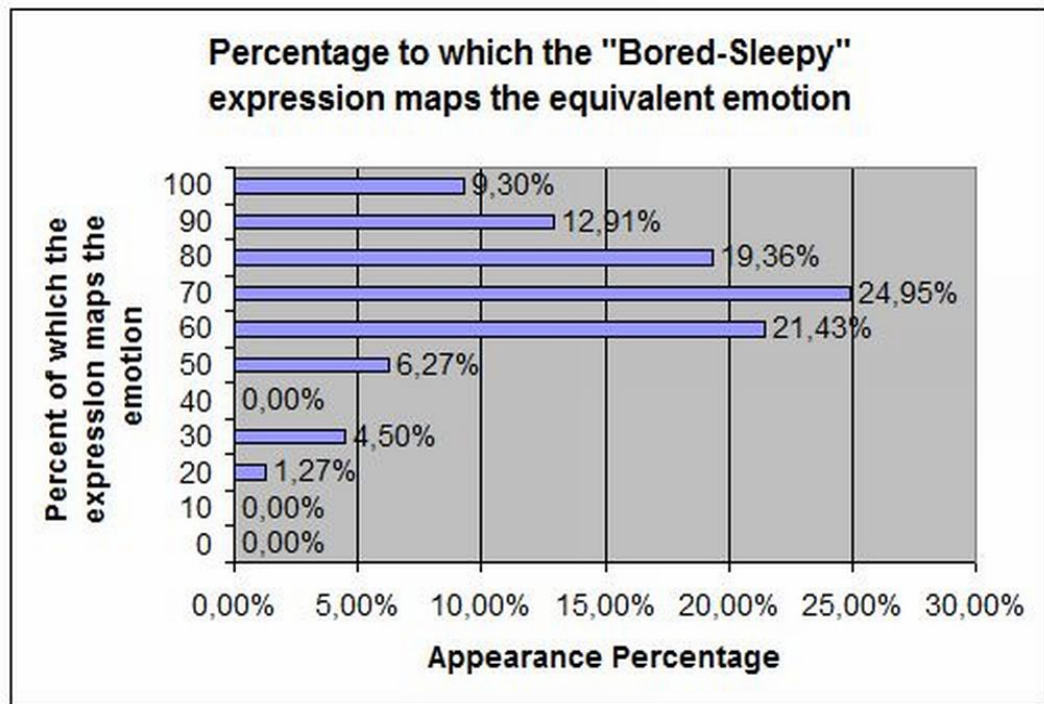


Figure 5.9 : Graph of the percentage to which the 'bored - Sleepy' expression maps the equivalent emotion, based on the correct answers of the participants

mistook the 'disgusted' emotion for some other emotion are shown in Figure 5.10 for both parts of the questionnaire. Specifically, as 'other' emotions, the participants indicated 'teasing' and 'bitter'.

The facial features that helped the participants classify the expression were mostly the 'eyes' and the 'mouth'. The percentage (%) of each facial feature is shown in Figure 5.11.

Regarding the percentage to which the 'disgusted' expression maps the equivalent emotion, the majority of the participants (76,47%) agreed that the expression maps the emotion with accuracy varying between 50% to 70%, as in Figure 5.12.

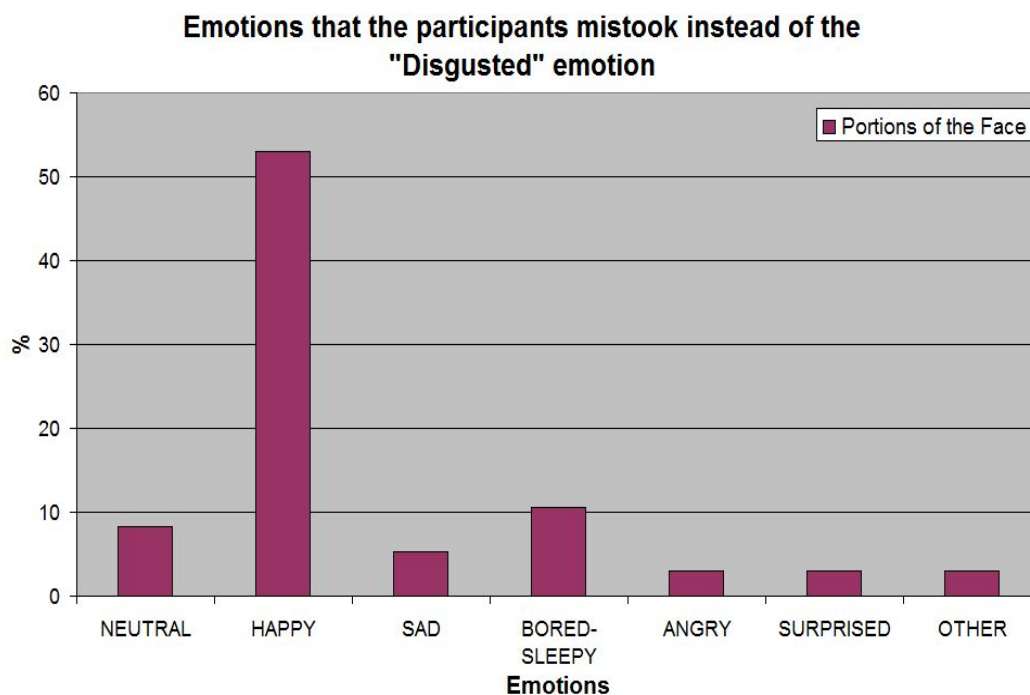


Figure 5.10 : Graph of the percentage to which the participants mistook the ‘disgusted’ emotion for other emotions

Happy

The ‘happy’ expression was recognized at an error rate of 31,06% in the first part of our detailed questionnaire, in which the entire facial image was depicted. The corresponding error rate was 3,78% in the second part of the questionnaire, in which only portions of the faces were depicted.

The participants mistook the ‘happiness’ emotion state mostly for the ‘boredom-sleepiness’ state. The percentages to which the participants mistook the ‘happiness’ for some other emotion state are shown in Figure 5.13 for both parts of the questionnaire. Specifically, as ‘other’ emotions, the participants indicated ‘ironic’, ‘satisfied’,

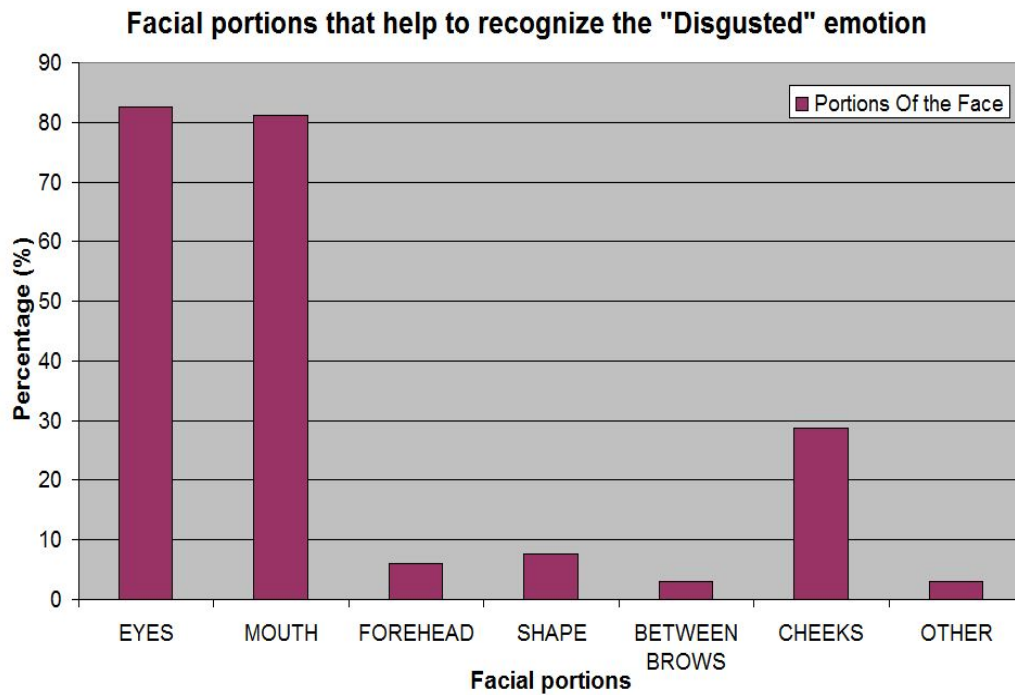


Figure 5.11 : Graph of the percentage to which the ‘disgusted’ expression maps the equivalent emotion, based on the correct answers of the participants

‘drunk’, and ‘pleased’, most of which are considered as ‘positive’ emotions. As the ‘happiness’ emotion is considered as positive, one might come to the conclusion that when it comes to a broader class of somehow similar emotions, the error rate was even lower.

The facial portions that helped the participants classify the expression were mostly the ‘eyes’, the ‘mouth’ and the ‘cheeks’. Specifically, the participants assigned percentages of significance of each facial portion as in Figure 5.14.

Finally, regarding the percentage to which the ‘happy’ expression maps the equivalent emotion, the participants’ answers are shown in Figure 15. Specifically, 18,40%

of the participants indicated that it shows 90% percent of the emotion. Moreover, the majority of the participants (77,7%) thought that the ‘happy’ expression maps more than 60% of the emotion. Regarding the percentage to which the ‘disgusted’ expression maps the equivalent emotion, the majority of the participants (76,47%) agreed that the expression maps the emotion with accuracy varying between 50% to 70%, as in Figure 5.15.

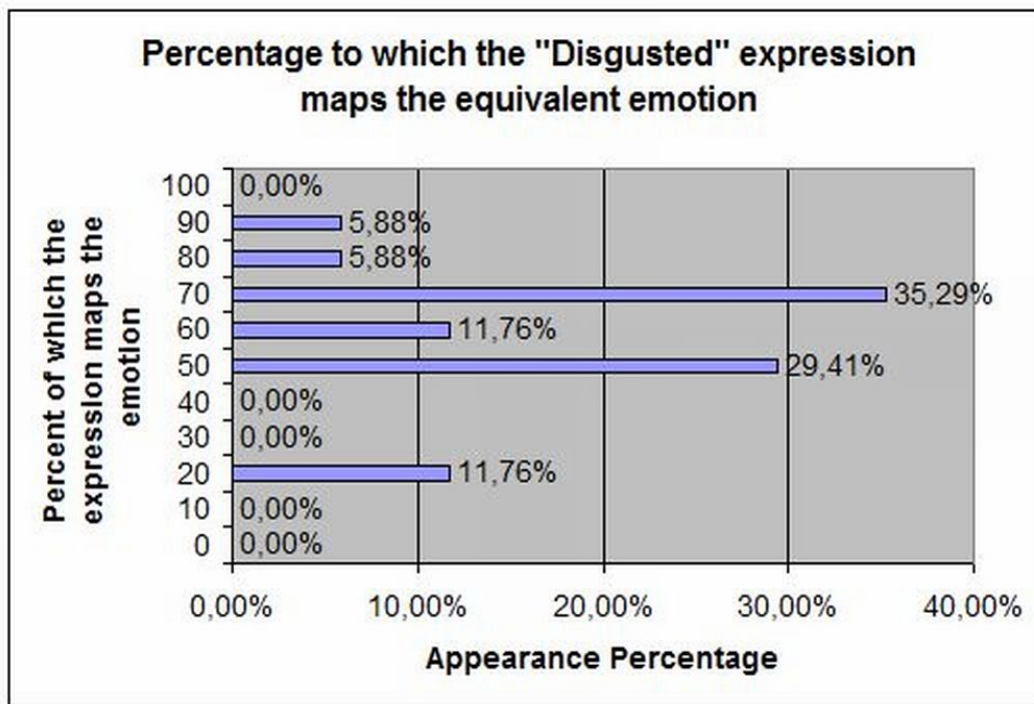


Figure 5.12 : Graph of the percentage to which the ‘disgusted’ expression maps the equivalent emotion, based on the correct answers of the participants

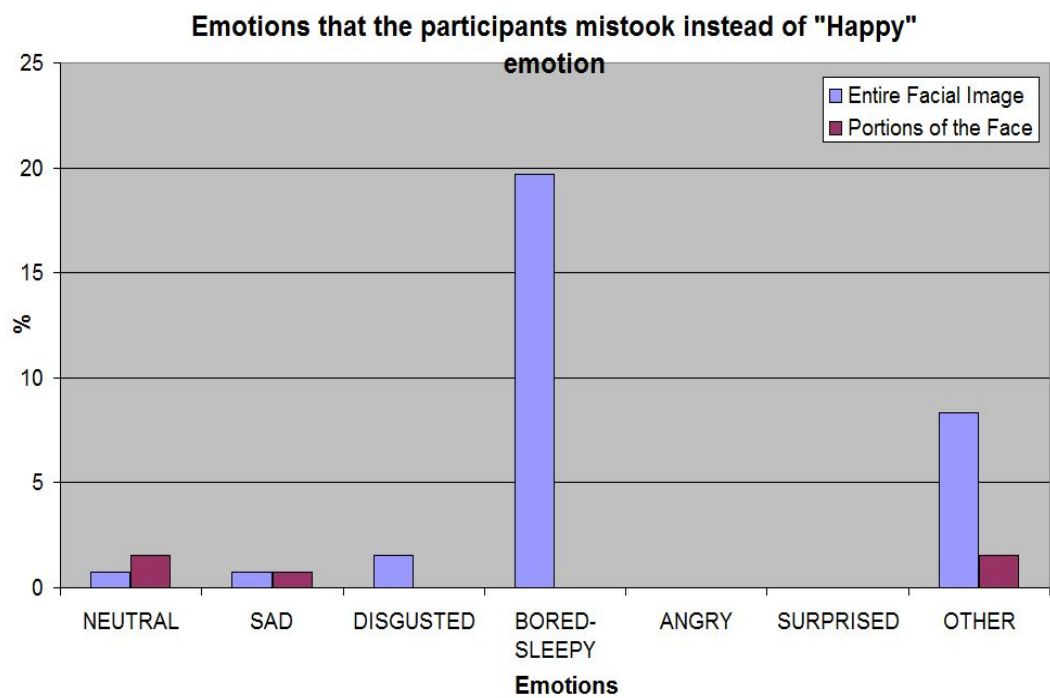


Figure 5.13 : Graph of the percentage to which the participants mistook the 'happy' emotion for other emotions

Neutral

The 'neutral' expression was recognized at an error rate of 61,74% in the first part of our detailed questionnaire, in which the entire facial image was depicted. We did not use this expression in the second part of our questionnaire, as we considered other expressions as deviations from it.

The participants mistook the 'neutral' emotion mostly for the 'happy' expression. The percentages to which the participants mistook the 'neutral' emotion for some other emotion are shown in Figure 5.16 for both parts of the questionnaire. Specifically, as 'other' emotions, the participants indicated 'sceptical', 'sleepy', 'slightly

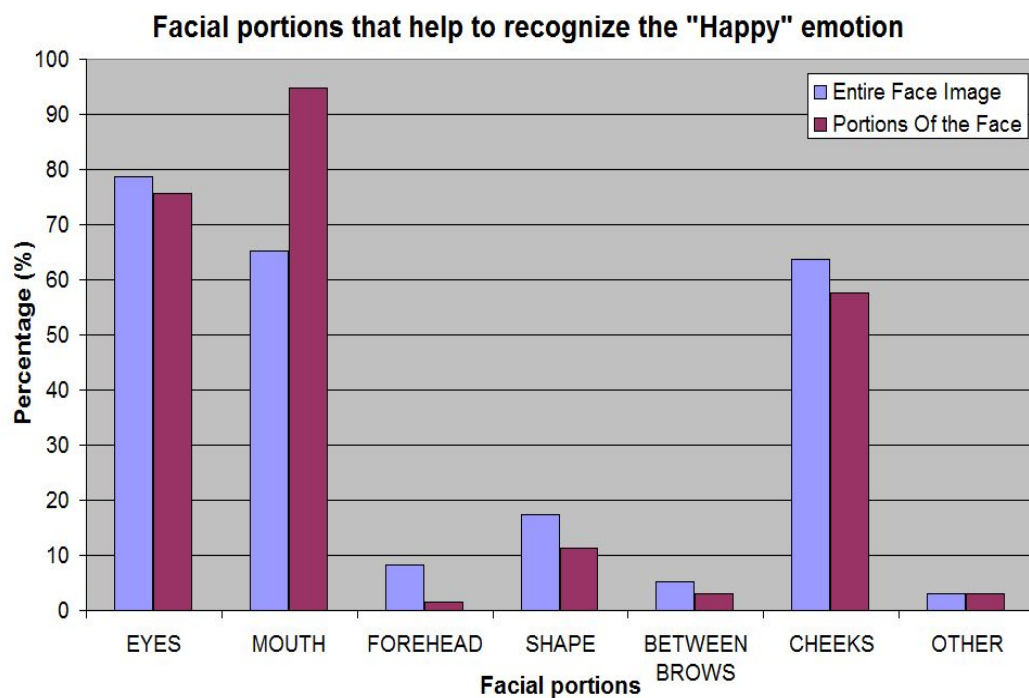


Figure 5.14 : Graph of the percentage to which the ‘happy’ expression maps the equivalent emotion, based on the correct answers of the participants

happy’ and ‘uneasy’.

The facial portions that helped the participants classify the expression were mostly the ‘eyes’ and the ‘mouth’. Specifically, the participants assigned percentages of significance of each facial portion as in Figure 5.17.

Finally, regarding the percentage to which the ‘neutral’ expression maps the equivalent emotion, the participants’ answers are shown in Figure 5.18 Specifically, 54% of the participants indicated that it shows 70-80% percent of the emotion. Moreover, the majority of the participants thought that the ‘happy’ expression maps more than 60% of the emotion.

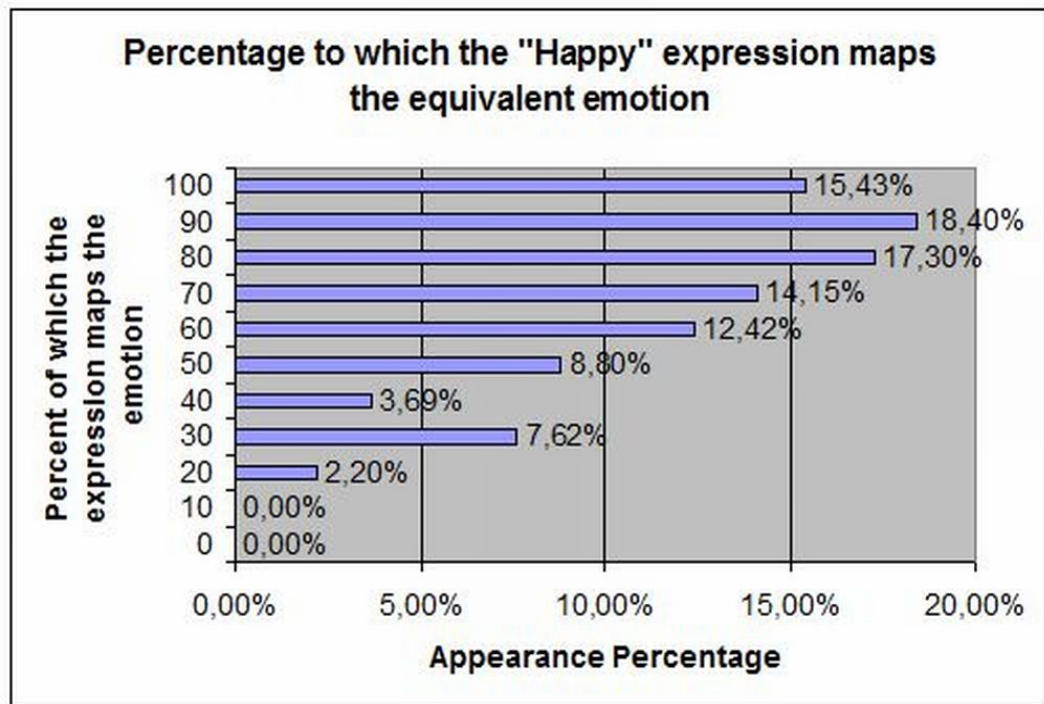


Figure 5.15 : Graph of the percentage to which the 'happy' expression maps the equivalent emotion, based on the correct answers of the participants

Sad

The 'sad' expression was recognized at an error rate of 65,9% in the first part of our detailed questionnaire, in which the entire facial image was depicted. The corresponding error rate was 17,42% in the second part of the questionnaire, in which only portions of the faces were depicted.

The participants mistook the emotion state of 'sadness' mostly for the 'neutral' and 'anger' states. This might have happened because sometimes anger causes sadness. Also, as the 'sadness' is a negative emotion that humans tend to disguise and the equivalent facial expression may not be strongly formed. The percentages to

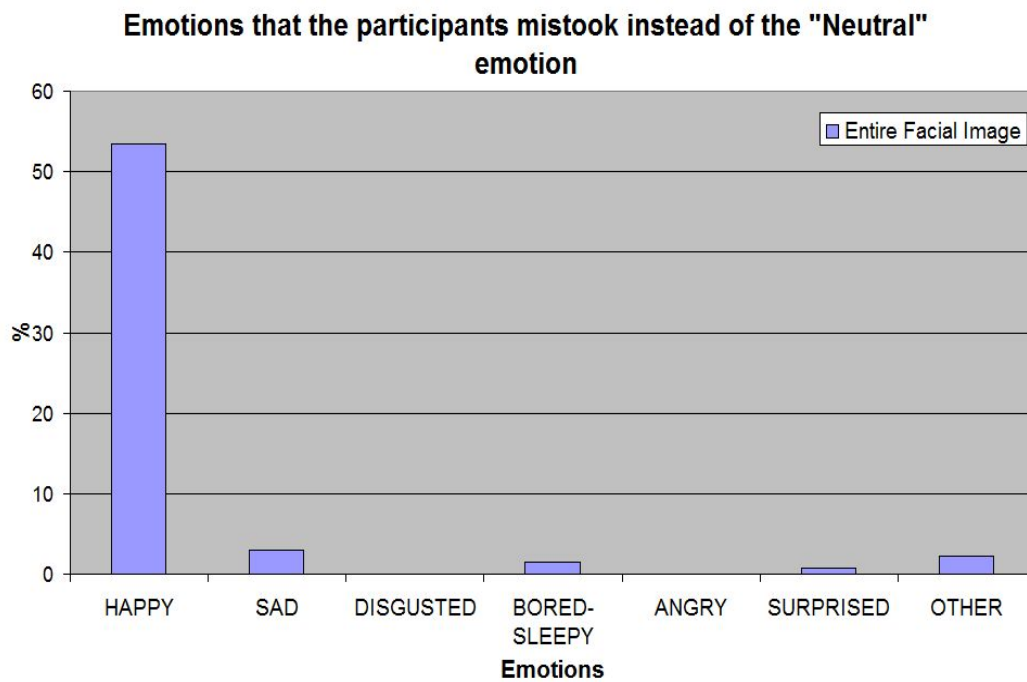


Figure 5.16 : Graph of the percentage to which the participants mistook the ‘neutral’ emotion for other emotions

which the participants mistook the ‘sad’ emotion for some other emotion are shown in Figure 5.19 for both parts of the questionnaire. Specifically, as ‘other’ emotions, the participants indicated ‘disappointed’, ‘anxious’, ‘tired’, and ‘wondering’. These emotions may happen simultaneously or, even, be caused by the emotion of ‘sadness’.

The facial portions that helped the participants classify the expression were mostly the ‘eyes’ and the ‘mouth’. Specifically, the participants assigned percentages of significance of each facial portion as in Figure 5.20.

Finally, regarding the percentage to which the ‘sad’ expression maps the equivalent emotion, the participants’ answers are shown in Figure 5.21. Specifically, the majority

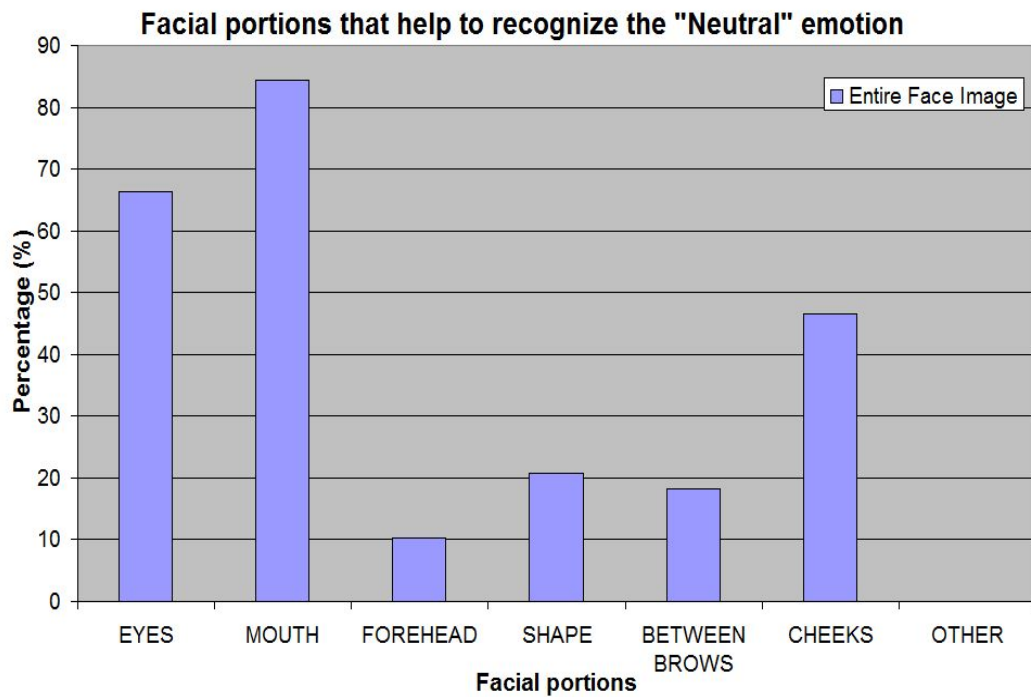


Figure 5.17 : Graph of the percentage to which the 'neutral' expression maps the equivalent emotion, based on the correct answers of the participants

of the participants thought that the 'sad' expression maps more than 60% of the emotion.

Surprised

The 'surprised' expression was recognized at an error rate of 10,22% in the first part of our detailed questionnaire, in which the entire facial image was depicted. The corresponding error rate was 4,54% in the second part of the questionnaire, in which only portions of the faces were depicted.

The participants mistook the 'surprised' emotion mostly for the 'happy' expres-

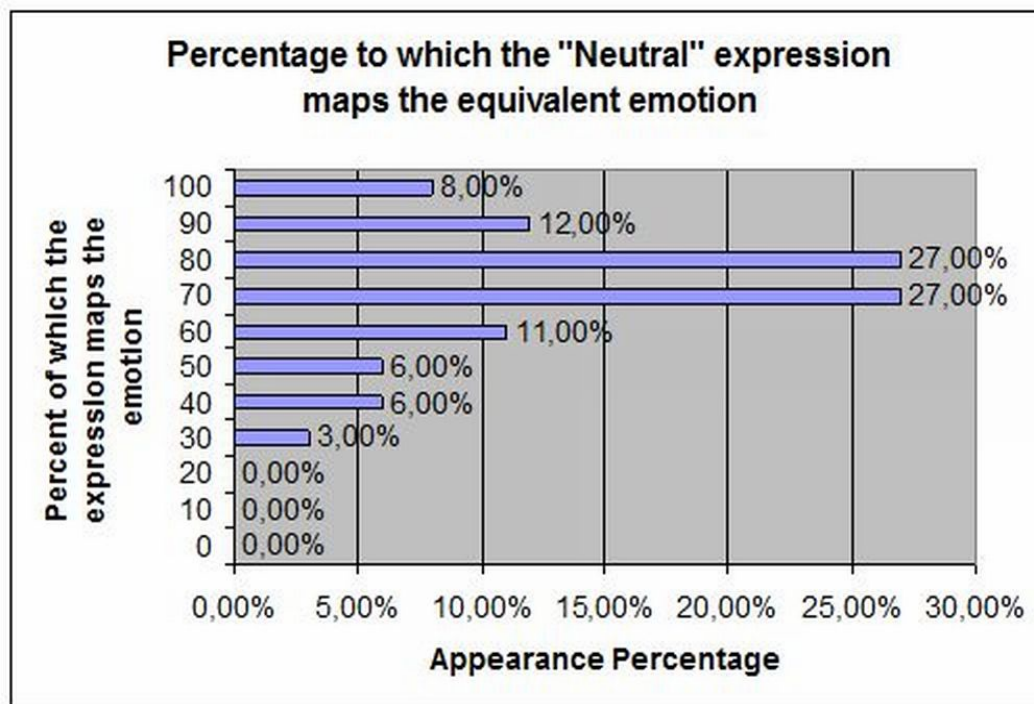


Figure 5.18 : Graph of the percentage to which the ‘neutral’ expression maps the equivalent emotion, based on the correct answers of the participants

sion. The percentages to which the participants mistook the ‘surprised’ emotion for some other emotion are shown in Figure 5.22 for both parts of the questionnaire. Specifically, as ‘other’ emotions, the participants indicated ‘afraid’, ‘excited’, and ‘relieved’. Again, some of these emotions, especially, the ‘afraid’ and the ‘excited’ can result in the formation of a facial expression of surprise.

The facial portions that helped the participants classify the expression were mostly the ‘eyes’, the ‘mouth’ and the ‘shape of the face’. Specifically, the participants assigned percentages of significance of each facial portion as in Figure 5.23.

Finally, regarding the percentage to which the ‘surprised’ expression maps the

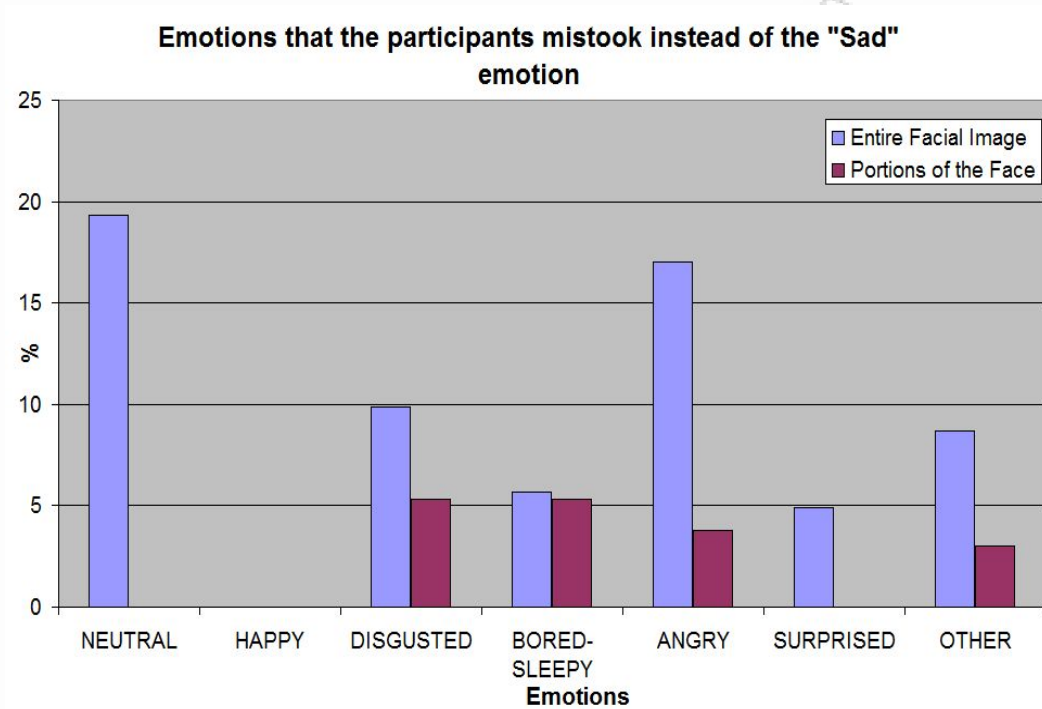


Figure 5.19 : Graph of the percentage to which the participants mistook the 'sad' emotion for other emotions

equivalent emotion, the participants' answers are shown in Figure 5.24. Specifically, 22,32% of the participants indicated that it shows 100% percent of the emotion. Moreover, the majority of the participants thought that the 'happy' expression maps more than 70% of the emotion.

5.3.2 Difficulties of Facial Expression Classification as Outlined by the Participants

In the third (final) part of our questionnaire, we asked the participants to give their opinion regarding the difficulties when classifying an emotion. When it comes to

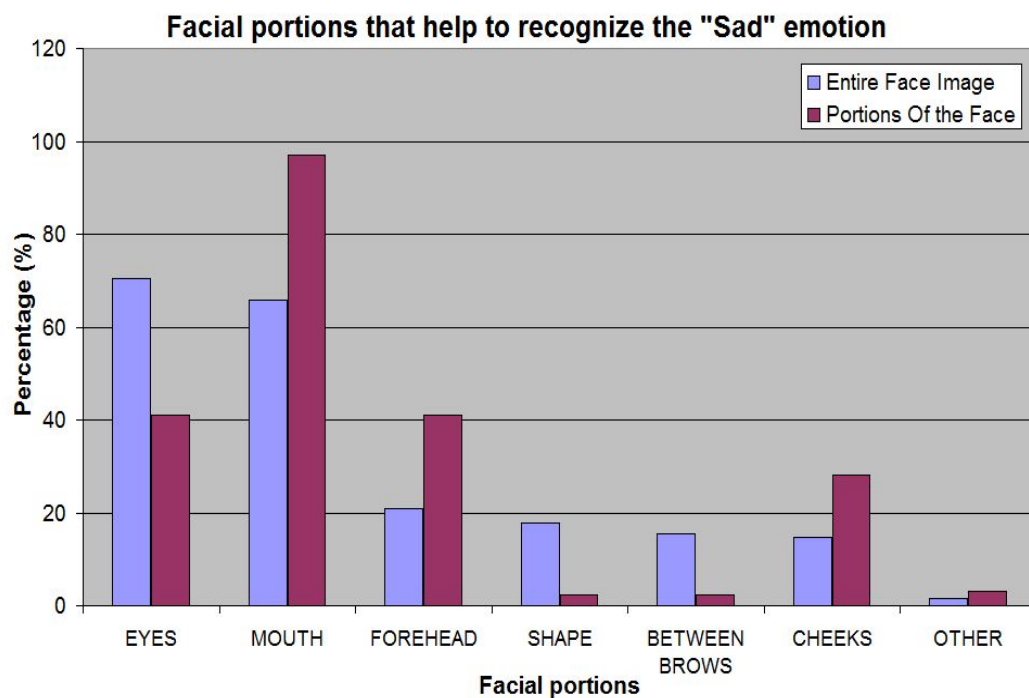


Figure 5.20 : Graph of the percentage to which the 'sad' expression maps the equivalent emotion, based on the correct answers of the participants

recognizing an emotion from someone else's facial expression, the majority of the participants consider this as a difficult task. Specific corresponding percentages are shown in Figure 5.25.

Regarding the most difficult emotion to recognize, the participants thought that this is the 'bored- sleepy' and the 'disgusted' emotions to percentages of 28% and 25%, respectively. However, the classification tasks in the first two parts of the questionnaire indicate the 'disgusted' and the 'sad' emotions as those identified with the highest error rates. Corresponding percentages for all emotions are shown in Figure 5.26.

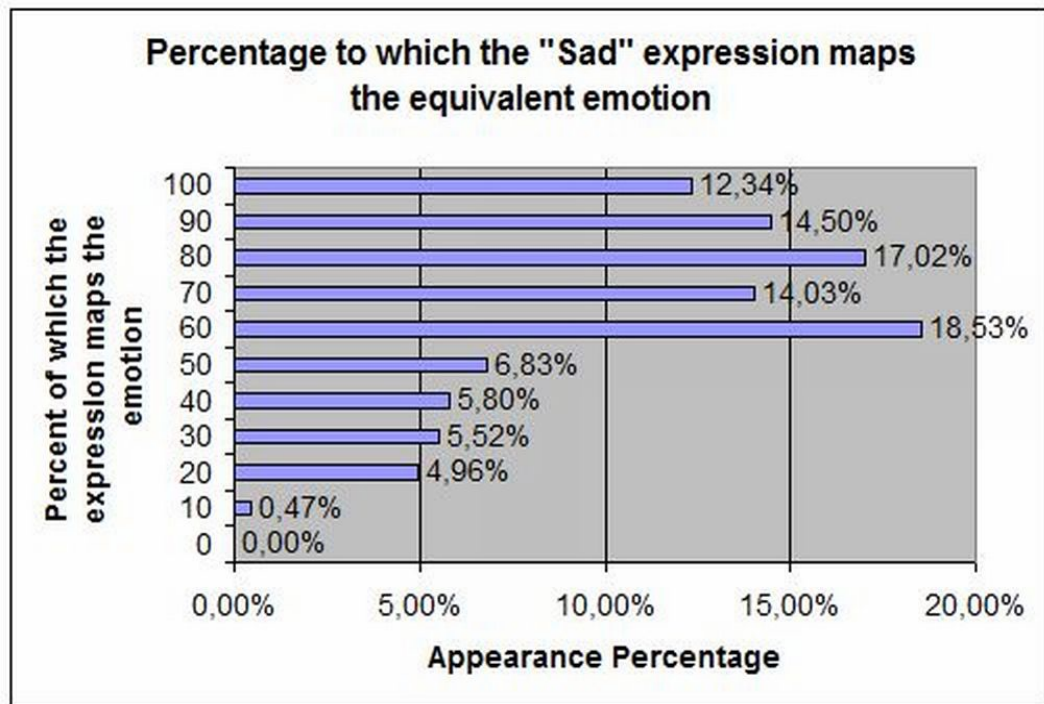


Figure 5.21 : Graph of the percentage to which the 'sad' expression maps the equivalent emotion, based on the correct answers of the participants

The majority of participants (65%) considered the emotion of 'happiness' as the easiest emotion to recognize. Generally, 'positive' emotions were considered easier to recognize than 'negative' emotions. Corresponding percentages for all emotions are shown in Figure 5.27.

The participants' opinion regarding the easiest emotion to recognize coincided with the results of the questionnaire, as the 'happiness' and 'surprise' achieved the lowest error rates, of 17% and 7%, respectively. As for the most difficult emotion to recognize, our questionnaire showed that the 'disgusted' and the 'neutral' were the most difficult emotions to recognize. Error rates corresponding to all emotions are

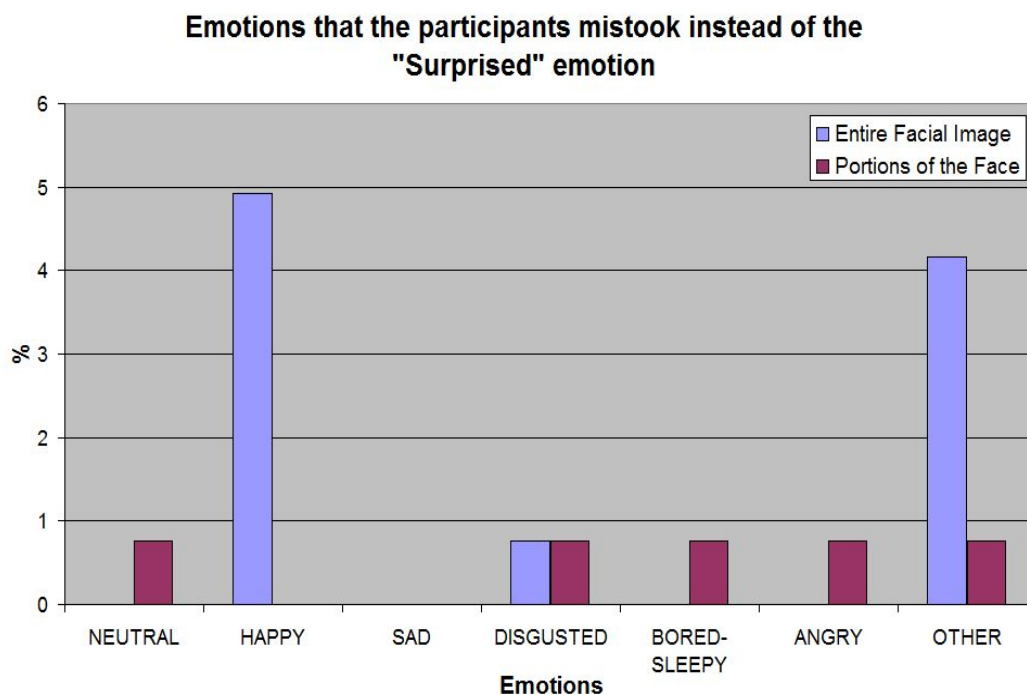


Figure 5.22 : Graph of the percentage to which the participants mistook the ‘surprised’ emotion for other emotions

shown in Figure 5.28.

The difference between the error rates of the preliminary and the detailed questionnaire is quite remarkable. This may be in line with the fact that for the development of the detailed questionnaire we used images of high resolution and quality from our own facial expression database. Moreover, the aforementioned databases were constructed by photographing Greeks and the expression on them were classified by Greeks, that is by people sharing the same culture and habits.

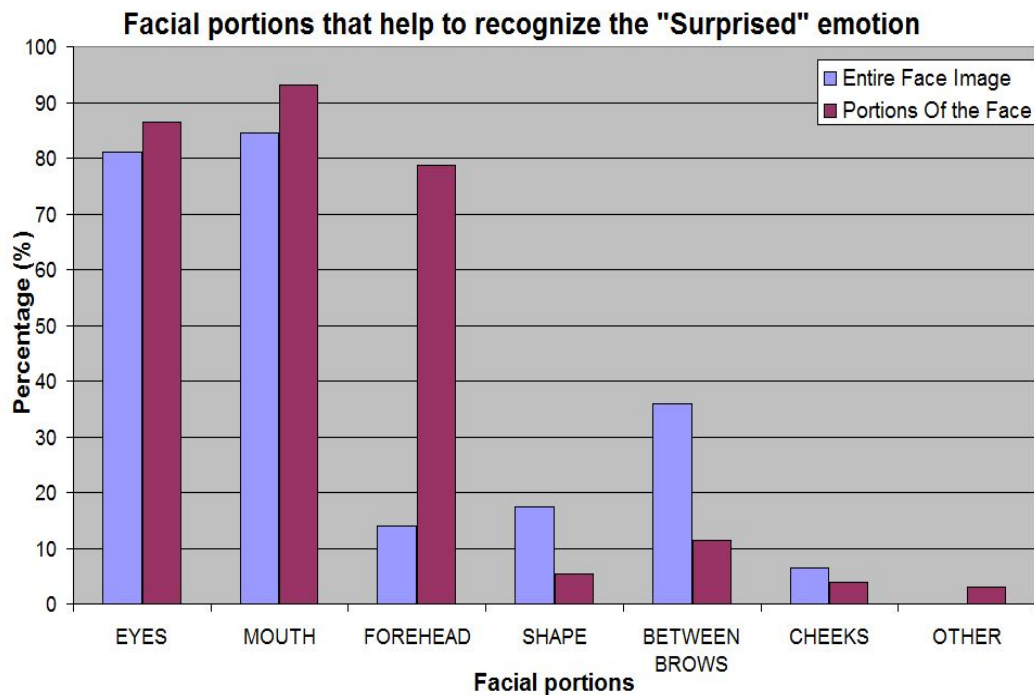


Figure 5.23 : Graph of the percentage to which the 'surprised' expression maps the equivalent emotion, based on the correct answers of the participants

5.3.3 Statistical Significance of the Results

Most of the participants agreed that a facial expression represented the equivalent emotion with a percentage of 70% or higher. The results are shown in Table 5.2.

In the first part of the detailed questionnaire, we asked the participants to map the facial emotion from the facial expression whereas in the second part the participants were asked to perform the same task but only from portions of the face. Each participant could choose from the 7 of the most common emotions that we pointed out earlier, such as: 'angry', 'happy', 'neutral', 'surprised', 'sad', 'disgusted', 'bored-sleepy', or specify any other emotion that he/she thought appropriate. Next, the

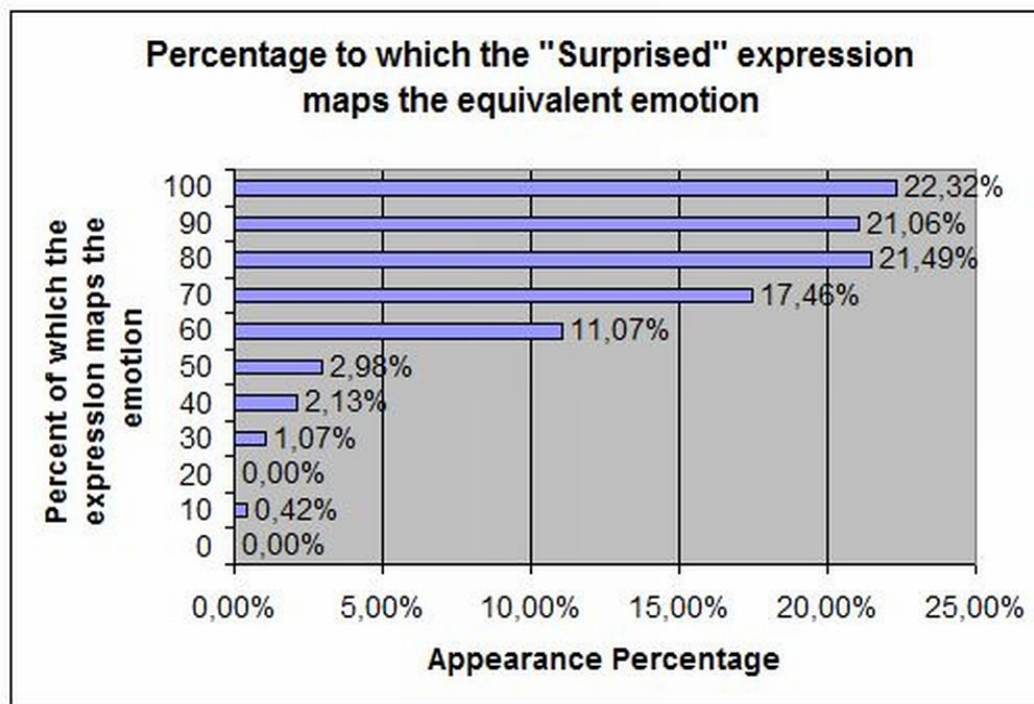


Figure 5.24 : Graph of the percentage to which the ‘surprised’ expression maps the equivalent emotion, based on the correct answers of the participants

participant had to decide the degree (0-100%) of confidence to which he/she thought that the emotion was mapped from the facial image indicated. Finally, he/she had to indicate those features (such as the eyes, the nose, the mouth, the cheeks etc.) that had helped him/her make a decision.

In the second part of our questionnaire, where we had chosen specific facial portions to display to the participants, smaller emotion classification error rates were achieved than in the first part, where the entire face image was displayed. The differences in error rates are quite significant and show that the facial portions are well chosen. The differences between the error rates are shown in Table 5.3. The statis-

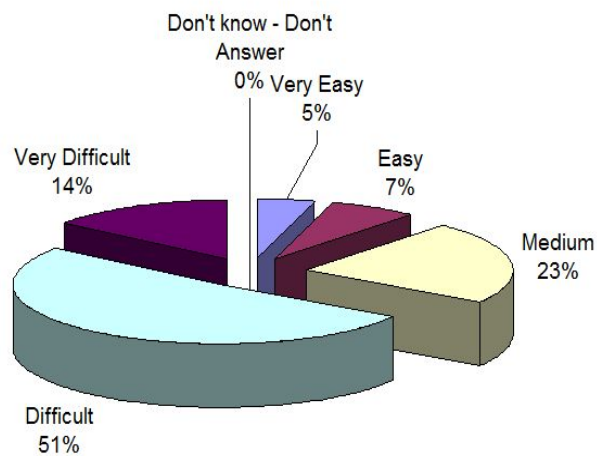
DIFFICULTY IN UNDERSTANDING EMOTIONS

Figure 5.25 : The participants' answers regarding the level of difficulty of the facial expression recognition task

tical significance of these results is shown in the last column (P-value) of Table 5.3 and is quite high.

5.3.4 Extraction of Facial Expression Classification Features

The facial portions that helped the users to understand the emotions are mostly the eyes, the mouth, and the cheeks. In some expressions, e.g. the 'angry' expression, other very important facial portions arose, such as the texture between the brows. These results are shown in Tables 5.4 and 5.5. In Table 5.4 the three most important features for the recognition of each expression are highlighted.

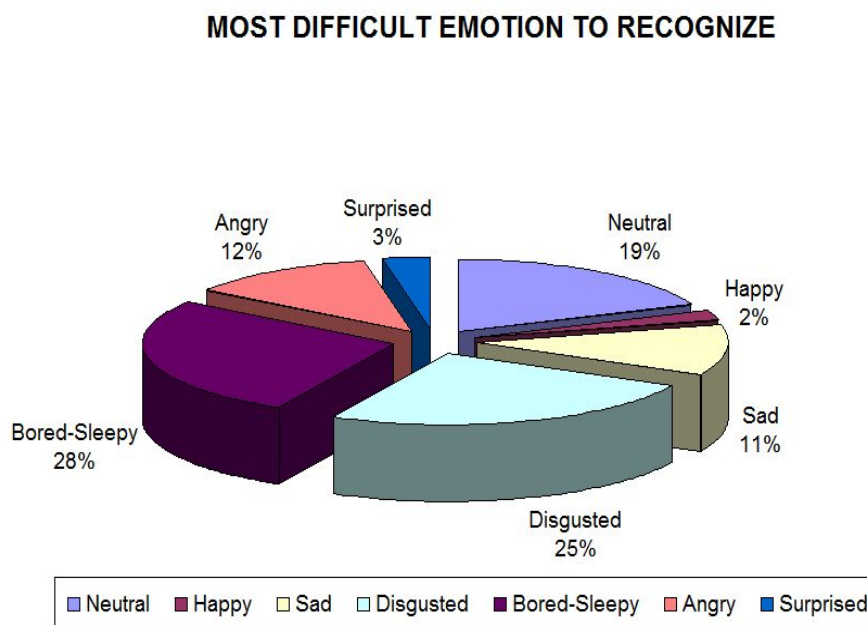


Figure 5.26 : The participants’ answers regarding the most difficult emotion to recognize

Based on the results in Table 4, we can identify the three most important features for the identification of each emotion. We examined such combinations of features for each expression to see whether the participants who misrecognized expressions based their answers on different face portions than participants who recognized expressions correctly. From our studies, we can summarize the results for each emotion, as follows:

‘angry’: (1) The combination of the three features (‘the eyes’, ‘the mouth’ and ‘the region between the brows’) was used by the participants who recognized the expression correctly at a percentage of 38,31% for the first and 30,43% for the second questionnaire part. (2) In the first questionnaire part, 38,1% of the participants who

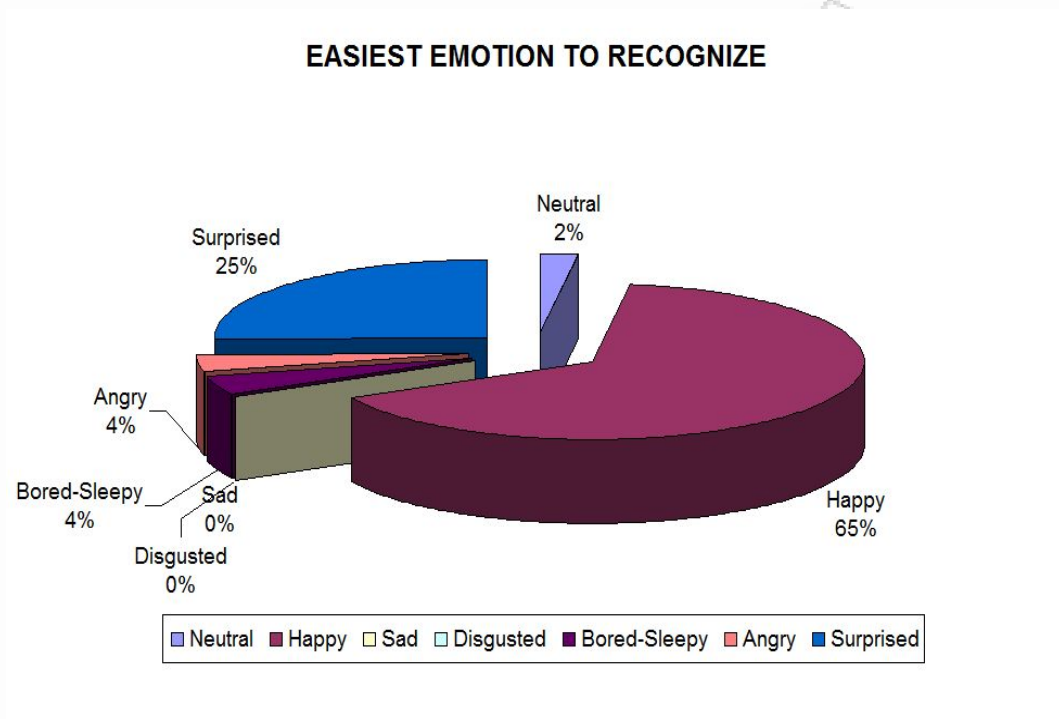


Figure 5.27 : The participants' answers regarding the most difficult emotion to recognize

recognized the expression considered only 'the eyes' as significant. (3) All the feature combinations that included the 'region between the brows' feature resulted to correct answers.

'bored-sleepy': The 'eyes' feature (alone or combined with other features) is important for the expression recognition task. Specifically: (1) The combination of 'the eyes' and 'the mouth' was used by the participants who recognized the expression correctly at a average percentage of 27,92% for the two parts of the questionnaire. (2) The 20,86% of the participants who gave the correct answers, used only 'the eyes'. (3) Finally, the combination of the three features ('the eyes', 'the mouth' and 'shape

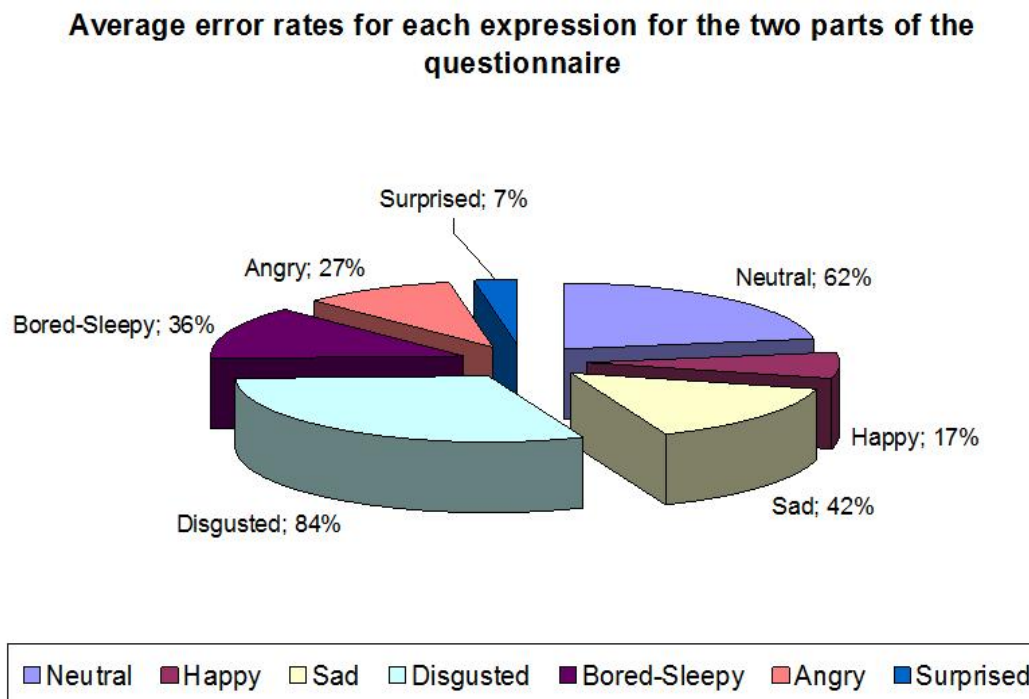


Figure 5.28 : Error rates in recognizing the expressions in our detailed questionnaire

of the face’ was used by the 15,98% of the participants who gave the correct answer.

‘disgusted’: The ‘disgusted’ expression achieved the highest error rate, 83,81% on average for the two parts of the questionnaire. Thus, it is difficult to indicate with safety the facial portions which led to good results, as only a relatively small number of the participants managed to recognize the emotion from the facial expression. Our studies concluded that: (1) The combination of ‘the eyes’ and ‘the mouth’ resulted to the best results with an average error rate of 55,56% for the two parts of the questionnaire. (2) A percentage of 16,67% of the participants who gave the correct answers used only ‘the eyes’.

‘happy’: (1) The combination of the three features (‘the eyes’, ‘the mouth’ and

Table 5.2 : Percentage to which a facial expression represents an emotion

| Percentage (%) to which an expression represents an emotion | User's answers |
|---|----------------|
| 0% | 0,00% |
| 10% | 0,00% |
| 20% | 0,76% |
| 30% | 2,27% |
| 40% | 1,52% |
| 50% | 9,85% |
| 60% | 14,39% |
| 70% | 31,06% |
| 80% | 21,97% |
| 90% | 15,91% |
| 100% | 2,27% |

‘the cheeks’) was used by the participants who recognized the expression correctly at a percentage of 35,16% for the first and 41,73% for the second part of the questionnaire.

(2) The participants who considered only ‘the eyes’ as significant for the recognition task achieved an average error rate of 24,63%, which leads us to the conclusion that the use of only ‘the eyes’ feature, cannot help to recognize the ‘happy’ expression.

‘neutral’: (1) The combination of the three features (‘the eyes’, ‘the mouth’ and ‘the cheeks’) was used by the participants who recognized the expression correctly

Table 5.3 : Error rate comparison between the two parts of the questionnaire

| Emotion | Error Rates | | Difference | P-Value |
|--------------|-------------|----------|----------------|----------------|
| | 1st Part | 2nd Part | | |
| Neutral | 61,74% | | Not applicable | Not applicable |
| Happy | 31,06% | 3,79% | 27,27% | 0,000000003747 |
| Sad | 65,91% | 17,42% | 48,48% | 0,000000000035 |
| Disgusted | 81,26% | 86,36% | -5,10% | 0,029324580032 |
| Bored-Sleepy | 49,24% | 21,97% | 27,27% | 0,000012193203 |
| Angry | 23,86% | 30,30% | -6,44% | 0,026319945845 |
| Surprised | 10,23% | 4,55% | 5,68% | 0,001390518291 |
| Other | 9,47% | 18,18% | -8,71% | |

at a average percentage of 38,45% for the two parts of the questionnaire. (2) The combination of ‘the eyes’ and ‘the mouth’ resulted to correct answers with 23,60% (3) Generally, the ‘eyes’ and the ‘cheeks’ can be considered as the most important features for the recognition task.

‘sad’: (1) The ‘mouth’ was the feature that was used from the users that classified correctly the ‘sad’ expression, at average percentage of 31,70% for the two parts of the questionnaire. (2)The combination of the ‘eyes’ and the ‘mouth’ led to

Table 5.4 : Important features for each facial expression

| | A | B | C | D | E | F | G |
|---|------|------|------|------|------|------|------|
| 1 | 66,3 | 81,6 | 63,6 | 82,6 | 77,3 | 55,7 | 83,7 |
| 2 | 84,5 | 67,8 | 76,1 | 81,1 | 79,9 | 81,4 | 88,8 |
| 3 | 10,2 | 22,7 | 4,2 | 6,1 | 4,9 | 30,9 | 46,4 |
| 4 | 20,8 | 14,4 | 31,1 | 7,6 | 14,4 | 10,0 | 11,4 |
| 5 | 18,2 | 59,5 | 8,7 | 3,0 | 4,2 | 8,9 | 23,7 |
| 6 | 46,6 | 8,1 | 30,7 | 28,8 | 60,6 | 21,4 | 5,1 |
| 7 | 0,0 | 2,5 | 3,0 | 3,0 | 3,0 | 2,3 | 1,5 |

Table 5.5 : Mapping

| | | | |
|---|---------------------------|---|--------------|
| 1 | Eyes | A | Neutral |
| 2 | Mouth | B | Angry |
| 3 | Texture of the Forehead | C | Bored-Sleepy |
| 4 | Shape of the Face | D | Disgusted |
| 5 | Texture between the brows | E | Happy |
| 6 | Texture of the cheeks | F | Sad |
| 7 | Other | G | Surprised |

good answers with average error rate of 23,16%.

‘surprised’: (1) The combination of the three features (‘the eyes’, ‘the mouth’ and ‘the region of the forehead’) was used by the participants who recognized the expression correctly at a average percentage of 63,49% for the two parts of the ques-

tionnaire. (2) The combinations of the ‘forehead’ with other features resulted to good answers. (3) The users that thought important only the ‘mouth’ feature, failed to recognize the emotion at an error rate of 12,96%.

The results are summarized in Table 5.6 and Table 5.7, where the three most important facial feature combinations are shown for each expression. This also leads to the single most important feature. Also, Table 5.6 and Table 5.7 shows the least important feature, as the one which leads to erroneous recognitions by the participants.

5.4 Summary - Conclusions

IN this Chapter, we described the two empirical studies that we conducted in order to set our error goal for our facial expression recognition system and understand how people classify an emotion. Based on the answers and the comments from the participants, we were led to the following assumptions:

1. Based on the participants’ comments and the questionnaire results, classifying an emotion of an unknown person from his/hers face image, is not a easy task. During interpersonal relationships, people usually recognize the emotion of someone they know almost instantly. However, this is not the case when they are faced with an unknown person’s image.
2. The cultural exposure increases the chances of correct recognition of facial expressions indicating cultural dependence in the ways people express themselves. This point is further strengthened from the results of our empirical studies. There is a big difference between the error rates of the first questionnaire, where

Table 5.6 : Identification of the most and least important features for each expression
- 1

| Most Important Features or Combinations of Features | | | Least Important | Single Most Important Feature |
|---|---------------------|---------------------------------|-------------------|----------------------------------|
| Angry | | | | |
| 'eyes', 'mouth' & 'rbb' | 'eyes' & 'rbb' | 'mouth' & 'rbb' | 'eyes' & mouth | 'rbb' |
| 34,37% | 14,33% | 8,94% | 8,94% | |
| Bored-Sleepy | | | | |
| 'eyes' & 'mouth' | 'eyes' | 'eyes', 'mouth' & 'shape' | 'mouth' | 'eyes' |
| 27,92% | 20,86% | 15,98% | 25,49% | 25,49% |
| Disgusted | | | | |
| 'eyes' & 'mouth' | 'eyes' | | 'cheeks' | 'eyes' |
| 16,67% | | | | 55,56% |
| Happy | | | | |
| 'eyes', 'mouth' & 'cheeks' | 'eyes' & 'mouth' | 'mouth' & 'cheeks' | 'eyes' | 'mouth' or 'cheeks' |
| 38,45% | 23,60% | 13,05% | 24,63% | |

we used images on non-Greek subjects, and the second questionnaire, where we used images from our own facial expression database. The results for the two questionnaires are summarized in Table 5.8. As we can observe, for the majority of the expressions the success rates were extremely comparable for the

Table 5.7 : Identification of the most and least important features for each expression
 - 2

| Most Important Features or Combinations of Features | | | Least Important | Single Most Important Feature |
|---|-------------------------|------------------------|--------------------|------------------------------------|
| Neutral | | | | |
| 'eyes' & 'mouth' | 'mouth' | 'eyes' | 'eyes', 'mouth' | 'mouth' or 'eyes' & 'cheeks' |
| 39,60% | 14,85% | 11,88% | 36,81% | |
| Sad | | | | |
| 'mouth' | 'eyes' & 'mouth' | | 'forehead' | 'mouth' |
| 31,70% | 23,16% | | | |
| Surprised | | | | |
| 'eyes', 'mouth' & 'forehead' | 'mouth' & 'forehead' | 'eyes' & 'forehead' | 'mouth' | 'forehead' |
| 63,49% | 7,50% | 4,49% | 12,96% | |

second questionnaire, as they achieved a difference beginning from 13% to 46%, compared to the first questionnaire. Exceptions were observed for the 'neutral' and the 'disgust' emotion.

3. In the majority of the emotions, the participants achieved better results in classifying the emotion when they were faced with parts of the subject's face rather than the entire face image, as shown in Table 5.3
4. In the majority of the expressions, the features that helped a participant to recognize the emotion were the 'eyes' and the 'mouth'. In some cases, the

Table 5.8 : Differences between the First and the Second Questionnaire

| Emotions | Average Success Rates | | Difference |
|-----------|-----------------------|-------------------|------------|
| | 1st Questionnaire | 2nd Questionnaire | |
| Neutral | 65% | 38% | -27% |
| Happiness | 70% | 83% | +13% |
| Surprise | 78% | 93% | +15% |
| Anger | 20% | 73% | +53% |
| Disgust | 37% | 16% | -21% |
| Sadness | 12% | 58% | +46% |
| Boredom | - | 64% | - |

‘texture of the cheeks’ and the ‘texture of the forehead’ were also taken into account by participants of the empirical studies.

6

Visual-Facial Emotion Recognition System

Calm down, it's only ones and zeros!

—Kathy Mar

AS stated in previous sections, expressions play a significant communicative role in human-to-human interaction and interpersonal relations because they can reveal information about the affective state, cognitive activity, personality, intention and psychological state of a person and, in fact, this information may be difficult to mask. The ability of humans to analyze facial expressions of another person is one of the objects of study of the scientific areas of pattern recognition and computer vision and the results of this study are applied in the design of interactive systems for more efficient and friendlier human-computer interfaces, multimedia services, security control systems, criminology etc.

When attempting to mimic human-to-human communication, human-computer

interaction systems must determine the psychological state of a computer user, so that the computer can react accordingly. This may be exploited in the design of advanced human-computer interfaces, which attempt to take into consideration the variations of the emotions of human users during the interaction and make the computer react accordingly. Thus, vision-based human-computer interactive systems with the ability to process computer user face images and extract information about the user's identity, state and intent would prove very effective and friendly. Similar information can also be used in multimedia interactive services, security control systems or in criminology to uncover possible criminals.

Most works in automated facial expression analysis assume that the conditions under which a facial image or an image sequence is acquired are known and controlled. Usually, the image has the face in front view and the background is fairly simple, usually uniform in color. In the majority of previous works, the location and the extend of the face is known or easily computed. However, in real environments, this is not the case. Determining the exact location of the face in a digitized facial image is a more complex problem. First, the scale and the orientation of the face can vary from image to image. Also, if the photos are taken from a fixed camera, there is no way to know a priori the size and the angle of the face. For the above reasons and in order to fully automate the procedure of facial expression recognition, a two-step task is required: (1) a face detection step in which the system determines whether or not there are any faces in an image and, if so, returns the location and extent of each face and, (2) a facial expression classification step, in which the system attempts to recognize the expression formed on a detected face.

The development of such fully automated face image analysis systems, capable of

detecting a face and classifying a person's facial expression without errors, is quite challenging. Some of the challenges that have to be addressed in developing such a system arise from the facts that faces are non-rigid and have a high degree of variability in size, shape, color and texture. Furthermore, variations in pose, image orientation and conditions add to the level of difficulty of the problem. Moreover, the variability in the ways people express themselves, depending on their culture, psychological state and habits, make it even more difficult to determine one's psychological condition through his/her face image. These facts can make the analysis of the facial expressions of another person difficult and often ambiguous.

In previous sections, we tried to understand how scientists and ordinary people interpret and understand the emotions. Our study concluded to the following assumptions:

1. There is on-going debate about how psychologists understand the emotions and the facial expressions in general. Many studies have pointed to six basic emotions, namely 'anger', 'disgust', 'fear', 'happiness', 'sadness' and 'surprise'. Despite this fact, studies have also shown that there is a cultural specificity regarding the emotion expression and understanding.
2. This assumption was further strengthened by our own studies. As stated in Chapter 4, we developed two different questionnaires. During this process, Greek people were asked to map an image to an respective emotion. The first questionnaire contained subjects of other cultures, besides the Greek, who were expressing an emotion, in contrary to the second questionnaire which contained Greek people forming an expression. The success rates for the two question-

naires are extremely different. Specifically, the average success rate for the first questionnaire is 47% in contrary to the second, which scored 60,72%. This led us to the assumption that emotions are culturally specific.

3. We also studied previous attempts towards the development of: (1) a facial expression database, (2) a face detection system and (3) a facial expression recognition system. We set the requirements for an ideal result for each of the aforementioned three occasions, respectively. Our study concluded to the fact that there are some interesting attempts but there is none that can cover all the requirements.
4. Moreover, as our aim is to build a facial expression recognition system which can be used in more advance human-computer interaction techniques, there was a need for the system to recognize expressions that are common during a human-computer interaction session. Indeed, based on our studies, facial expressions corresponding to the ‘neutral’, ‘happiness’, ‘sadness’, ‘surprise’, ‘anger’, ‘disgust’ and ‘boredom-sleepiness’ psychological states arise very commonly during a typical human-computer interaction session, as stated in Chapter 4.
5. Finally, as there was no facial expression database that could cover our requirements, we developed our own facial expression database as described in Chapter 4

Based on these assumptions, we developed our own fully automated facial expression recognition system. The system consists of two modules: (1) a face detection module that determines whether a face is present in an image [235, 236, 237, 5, 238] and, if so, estimates its location and extent, and (2) a facial expression classification

module that classifies the expression on a face that has been detected by the first module. We will present these two modules extensively in the following sections.

6.1 Face Detection

6.1.1 P. Sinha's Template

The face detection module follows a feature-based approach, which combines template matching and image invariant approaches. This approach relies on an observation that P. Sinha made while aiming at finding a model that would satisfactorily represent some basic relationships between the various regions of a human face [120]. Specifically, P. Sinha found that the *relative* brightness of different parts of a face, such as the eyes, the cheeks, the nose and the forehead, remains unchanged, even when variations in illumination change the *individual* brightness of these parts. This relative brightness between facial parts is captured by an appropriate set of pairwise brighter-darker relationships between sub-regions of the face, as in Figure 6.1.

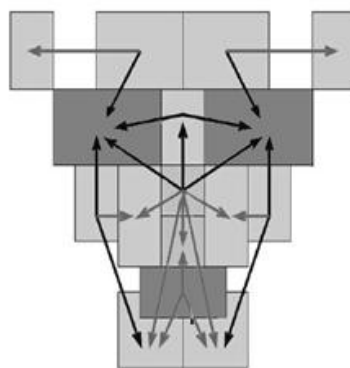


Figure 6.1 : The P. Sinha Template

6.1.2 The Face Detection Algorithm - Image Preprocessing

The face detection algorithm is built upon the P. Sinha template and preprocesses an image in order to enhance the relationships implied by the template and, subsequently, feeds the image into an artificial neural network to determine the presence and location of faces in the image. Specifically, the algorithm steps are as follows:

1. *Load image, which can be 3-dimensional (color or grayscale image)*
2. *Scan through image with a 35-by-35 pixel window. The image region contained in the window constitutes the 'window pattern', which is examined to determine whether it contains a face. The window size is gradually increased, so as to cover all the possible face sizes in the image.*
3. *Preprocess the window pattern as follows:*
 - (a) *Apply Histogram Equalization techniques to enhance the contrast within the window pattern.*
 - (b) *Compute the eigenvectors of the image using the Principal Component Analysis, a sample of the eigenfaces can be seen in Figure refeigenfaces and use the Nystrom Algorithm [239] to compute the normalized cuts.*
 - (c) *Compute three clusters of the image using the k-means algorithm and color each cluster with the average color.*
 - (d) *Convert the image from colored to grayscale (2-dimensional).*
4. *Resize image into dimensions of 20-by-20 pixels and feed it into the artificial neural network-based detectors described in the the following section.*

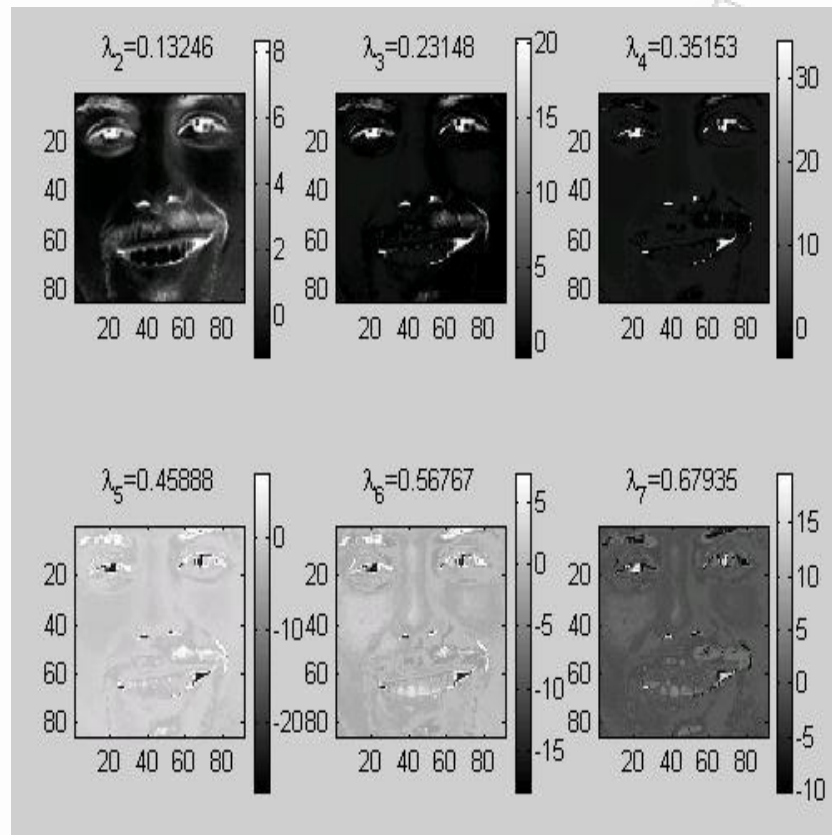


Figure 6.2 : The eigenfaces of a face image

The resulting three clusters and the the grayscale image which is fed to an artificial neural network can be seen in Table 6.1 for three face and three non-face images. Specifically, in the first column we can observe the input image, whereas, in second column we have the three clusters. Usually, each cluster of a face consists of:

1. In the first cluster we have parts of the face: such as the eyes, the mouth and the nostrils
2. In the second cluster we have parts around the facial features located in the

first cluster: regions around the eyes, the mouth and the nostrils

3. In the third cluster we have all the other face region that has been excluded from the above other two clusters

Finally, in the third column, we can observe the resulting grayscale clustered image which is fed to the artificial network classifiers.


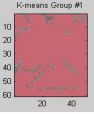
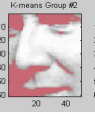
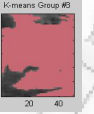


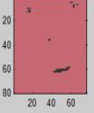
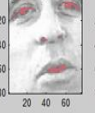



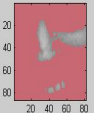







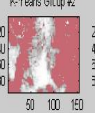
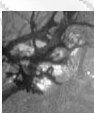
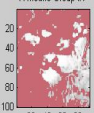

6.1.3 Artificial Neural Network-Based Face Detectors

We have developed and tested various neural network-based face detectors, two of which are presented next. To train the neural network, we used a set of 285 images of faces and non-faces, which were gathered from sources in the World Wide Web. We paid special attention to include in the training image set, non-face images that resemble human faces, such as dog, monkey and other animal images. All training images were preprocessed as described in the previous Section 6.1.2. An example of the pictures used for training the artificial neural networks are shown in Figure 6.3

A network with three hidden layers

This network consists of three hidden layers of thirty, ten and two neurons respectively, as in Figure 6.5 and 6.4. As input, it takes the entire window pattern of 20-by-20 pixels and produces a two-dimensional vector output, which classifies the window pattern as 'face' or 'non-face'. Specifically, the output vector equals $[1, 0]$, if the window pattern represents a face, and $[0, 1]$, otherwise. This implies that the output vector describes the degree of membership of the network input image in each of the 'face-image' and 'non-face-image' classes.

Table 6.1 : The computed three clusters

| Input Image | 3 clusters | Resulting image |
|---|--|---|
| Face Images | | |
|  |    |  |
|  |    |  |
|  |    |  |
| Non-face Images | | |
|  |    |  |
|  |    |  |
|  |    |  |

A network with four hidden layers

This neural network consists of four hidden layers with one, four, four and two neurons, respectively. Its input data consists of the following three types: (1) the entire window pattern (a 20-by-20 pixel image), (2) four portions of the window pattern,

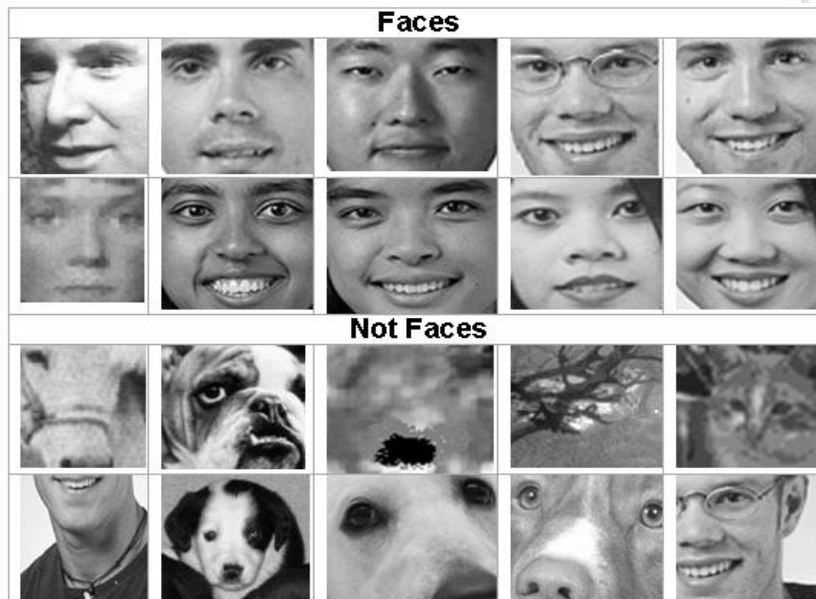


Figure 6.3 : Sample images of our Face Detection training set

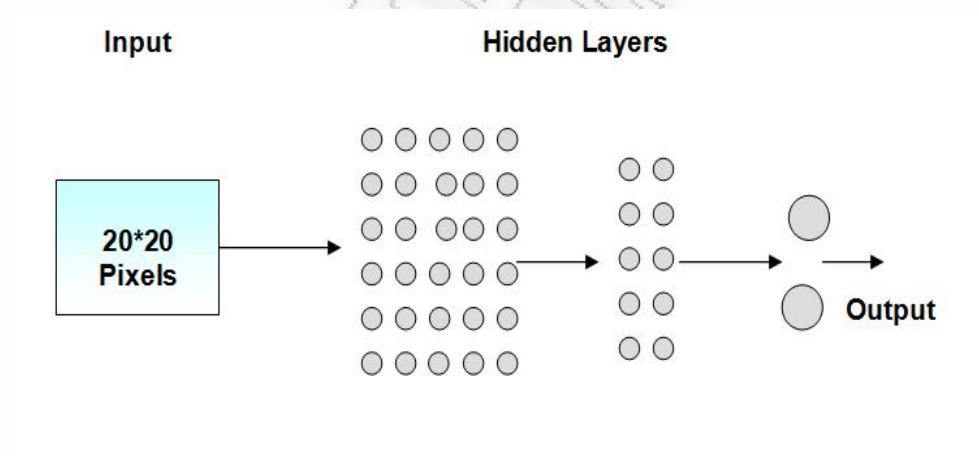


Figure 6.4 : Three Hidden Layer Network (Simple Demonstration)

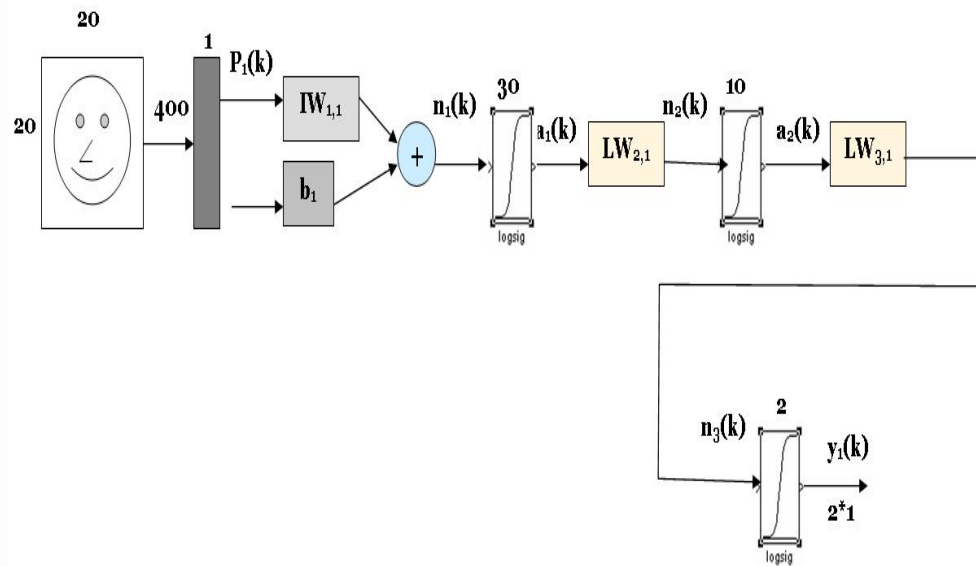


Figure 6.5 : Three Hidden Layer Network

each 10-by-10 pixels and (3) an additional four portions of the window pattern, each 5-by-20 pixels. The three input data types are fed into different hidden layers of the network. Specifically, the first, second, and third input data sets are fed into the first, second and third hidden layer, respectively. The output vector of this network is again a two-dimensional vector, which equals $[1, 0]$, if the input data represent a face, and $[0, 1]$, otherwise. Clearly, the first network consists of fewer hidden layers, but contains a higher total number of neurons and takes less input data than the second network. The network structure is shown in Figures 6.7 and 6.6

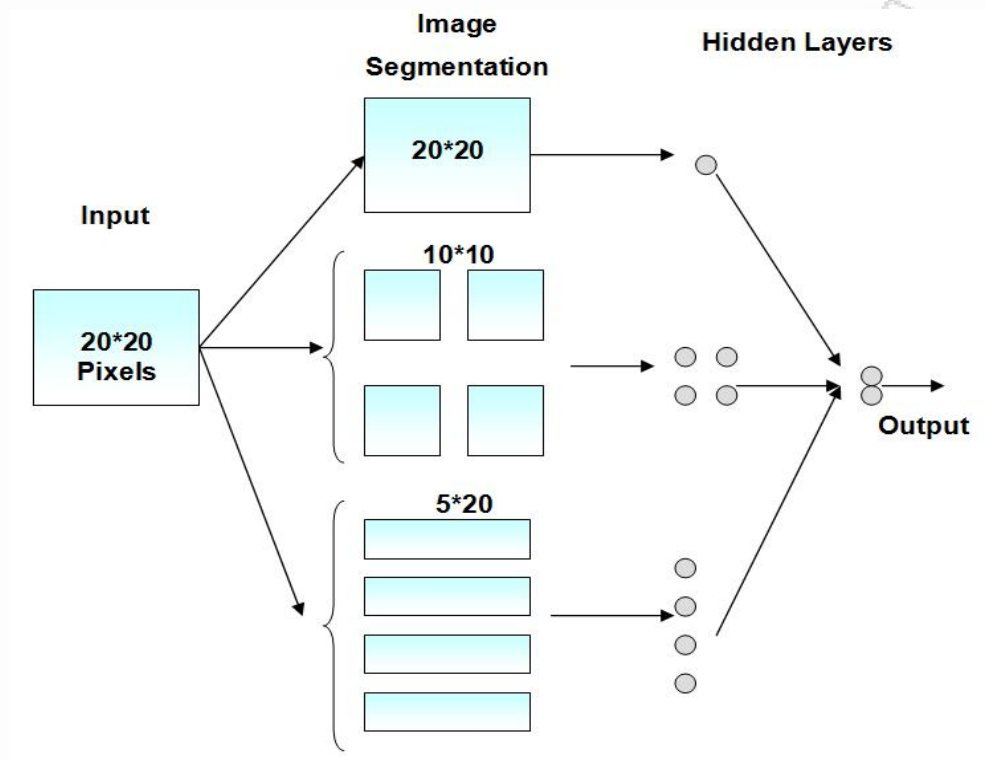


Figure 6.6 : Four Hidden Layer Network (Simple demonstration)

6.1.4 Performance Evaluation

To train these two networks, we used a common training set of 285 images of faces and non-faces, as mentioned before. During the training process, the two networks achieved error rates of 10^{-1} and 10^{-10} , respectively. The neural networks differences in terms of structure, input data and results are described briefly in Table 6.2.

Some results of the two neural networks can be seen in Figure 6.8. The first network, even though it consisted of more neurons than the second one, did not detect faces in the images to a satisfactory degree, as did the second network. On

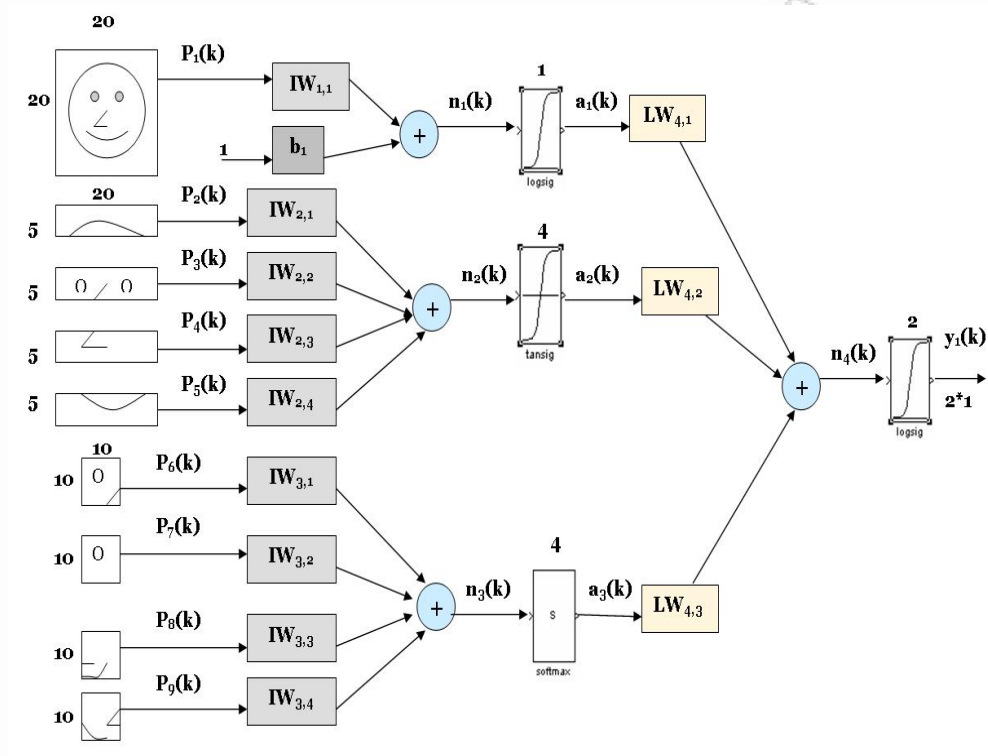


Figure 6.7 : Four Hidden Layer Network

the other hand, the execution speeds of two networks are comparable. Therefore, the second network was found superior in detecting faces in images.

To measure the performance of the second network [235, 236, 237, 5, 238] in detecting faces in images, we tested the network in four different sets of images:

1. 120 various images (60 female and 60 male photos) of different sizes and resolutions gathered from the World Wide Web and other sources (e.g. scanning old photo images)
2. 50 images (25 female and 25 male photos) from our own facial expression

Table 6.2 : Description of the two neural network classifiers

| First Neural Network | Second Neural Network |
|--|---|
| <i>Neural Network Structure</i> | |
| 3 Hidden Layers | 4 Hidden Layers |
| More Neurons in each layer | Fewer Neurons in each layer |
| <i>Input Data</i> | |
| Entire face image data as input | Entire face image data and parts of the face as input |
| <i>Training Set</i> | |
| 285 face and non face images | 285 face and non face images |
| <i>Error rates</i> | |
| 10^{-1} | 10^{-10} |

database where people may form some expression

3. 535 face images (205 female and 330 male photos) acquired in the first efforts to construct a facial expression database (low quality images)[237]
4. 50 non human face images (images of pets and animals, complex backgrounds and parts of the face and human)

The first set of images consists of images usually containing more than one faces in complex backgrounds. The faces may be partially occluded, slightly rotated or in side view. Also there are differences in the size, aspect ratio and resolution among these

















| Face Images | | | |
|---|---|---------------------------|----------------------------|
| Input Image | Preprocessed Window Pattern | First ANN's output | Second ANN's output |
|  |  | [1 ; 0] | [0.5 ; 0.5] |
|  |  | [0.947; 0.063] | [0.5 ; 0.5] |
|  |  | [1 ; 0] | [0.5 ; 0.5] |
|  |  | [0.9717; 0.0283] | [0.5 ; 0.5] |
|  |  | [1 ; 0] | [0.6 ; 0.4] |
| Not-Face Images | | | |
| Input Image | Preprocessed Window Pattern | First ANN's output | Second ANN's output |
|  |  | [0 ; 1] | [0.5 ; 0.5] |
|  |  | [0 ; 1] | [0.5 ; 0.5] |
|  |  | [0 ; 1] | [0.5 ; 0.5] |

Figure 6.8 : The face detection neural networks responses

photos. The second set of face images consists of a random selection of our own facial expression database, so the subjects may be forming one of the seven expressions: 'neutral', 'happy', 'sad', 'surprised', 'disgusted', 'angry' and 'bored-sleepy'. Finally, the third set of images consists of low quality photos acquired with web cameras. This set was constructed during our first efforts to build a facial expression database, so there is only one face in this photos and the subject may be forming facial expressions and/or images acquired in side view. The dataset included a random selection of images in front and side view and images acquired from subjects forming one of the eight expressions: 'neutral', 'happy', 'sad', 'surprised', 'angry', 'disgusted', 'screaming' and 'bored-sleepy'. The face detection results are summarized in Table 6.3.

The system managed to detect face with respective success rates of 90,83%, 94,00% and 72,89%, for the three sets. Errors (misses) occurred mostly because of overly bright illumination conditions which did not allow the extraction of facial features during k-means clustering. It was also observed that the detection of female faces was more difficult than the detection of male faces, possibly because facial features in female faces are not as tense as those in male faces. The results from these two groups are summarized in Table 6.3. Some typical results of the face detection system are depicted in Figures 6.9, 6.10, 6.11, 6.12, 6.13, and in Table 6.4. Table 6.5 for each of the four sets of images that we use to measure the performance of the Face Detection Subsystem.

In Figures 6.9 and 6.10, we observe the face as detected by our system and placed inside a green box. All images in these Figures correspond to the first test set of images, so they are images gathered from the World Wide Web or old photos scanned and used for testing our face detection system. As this is the case, we may have

Table 6.3 : Results of the Face Detection System for the three datasets

| Dataset information | # of detected faces | # of undetected faces | Success rates (%) |
|--|---------------------|-----------------------|-------------------|
| <i>1st Dataset: Images gathered from WWW and other sources</i> | | | |
| Female | 51 | 9 | 85,00% |
| Male | 58 | 2 | 96,67% |
| Sum | 109 | 11 | 90,83% |
| <i>2nd Dataset: Images from our database</i> | | | |
| Female | 23 | 2 | 92,00% |
| Male | 24 | 1 | 96,00% |
| Sum | 47 | 3 | 94,00% |
| <i>3rd Dataset: Images from our low quality database</i> | | | |
| Female | 115 | 90 | 56,09% |
| Male | 275 | 55 | 83,33% |
| Sum | 390 | 145 | 72,89% |

various faces in an image, where the subjects maybe posing during image shooting, or their faces may be partially occluded, rotated, or blurred due to noise. Specifically, in Subfigure 6.9(a) people are posing during photo shooting, so the resulting photo can be considered to be acquired in a rather controlled environment, as in these cases the photos don't usually contain faces partially occluded or rotated. Moreover the cases of bad quality images or noise because of blurring and/or subjects' movement are scarcely present in this type of images. On the contrary, in Subfigure 6.9(b), the

photo is slightly blurred because of the subjects' motion and one of the two faces are slightly rotated. The case of complex backgrounds, many subjects and different face views is shown in Subfigure 6.10(a), where there are many faces of different sizes in the image, some of which are in side view or partially occluded. Finally, in Subfigure 6.10(b) there is an old family photo which was scanned and used to test the neural network classifier. In this case, the photo can be considered of low quality because of the alterations that this photo was subjected to as time passed.



(a) People are posing during photo shooting
(b) Faces are slightly blurred or rotated because of the subjects' movement

Figure 6.9 : Face Detection results for the first set of images (images gathered from the World Wide Web) (a), (b)

As the aim of developing the face detection system was to facilitate and automate the facial expression recognition process, the system should perform well in cases where the subject is forming an expression. In order to check its performance we tested the face detection system using images from our own facial expression database, so, in this case, the subjects are posing during photo shooting and are forming some of the seven facial expressions, which correspond to 'neutral', 'angry', 'happy', 'sad',



(a) Many faces in complex backgrounds

(b) Scanned old photo

Figure 6.10 : Face Detection results for the first set of images (images gathered from the World Wide Web) (a), (b)

‘surprised’, ‘disgusted’, or ‘bored-sleepy’. Figures 6.11, 6.12 and 6.13 correspond to our facial expression database. Again, we can observe the face detected by our system and placed inside a green box.

The most challenging task of our face detection system was testing it with 535 face images (205 female and 330 male photos) acquired in the first efforts to construct a facial expression database. As mentioned in Chapter 4, these images are of low quality as they were acquired using simple web cameras. The images were of 320-by-240 pixel resolution, the size of the face varies from 170-by-220 to 70-by-80 pixels and the subjects were forming an expression during photo shooting. Specifically, the subjects maybe forming the ‘neutral’, ‘happy’, ‘sad’, ‘surprised’ ‘bored-sleepy’,

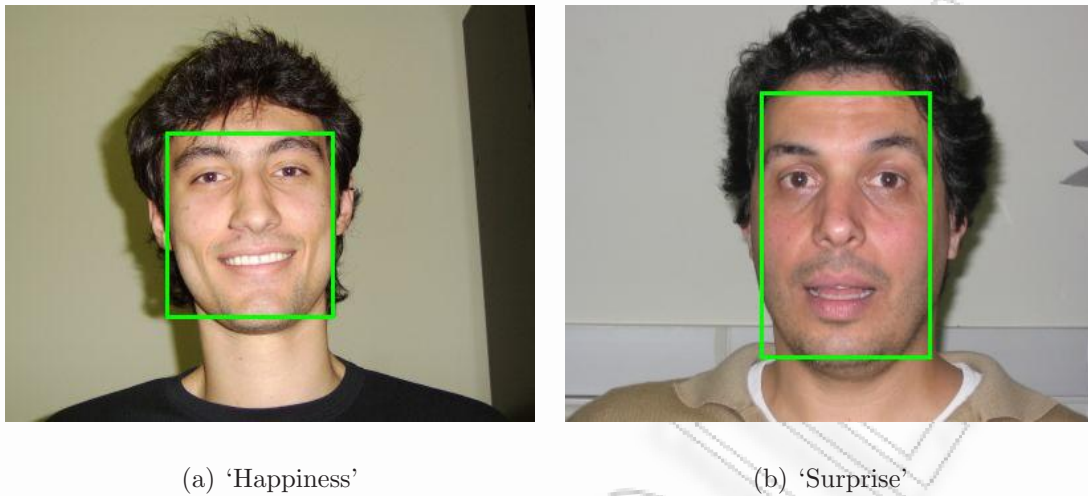


Figure 6.11 : Face Detection results for the second set of images, where the subject is forming an expression (a), (b)

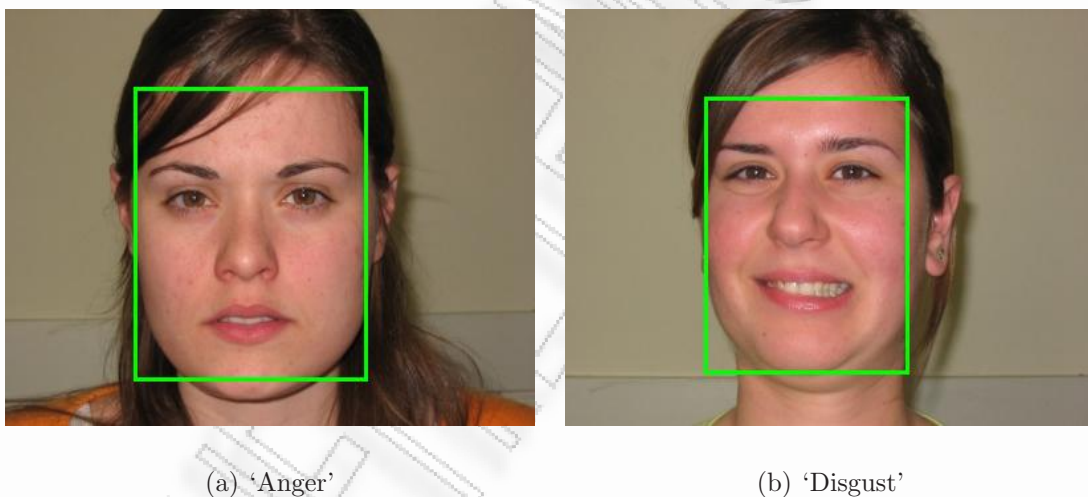
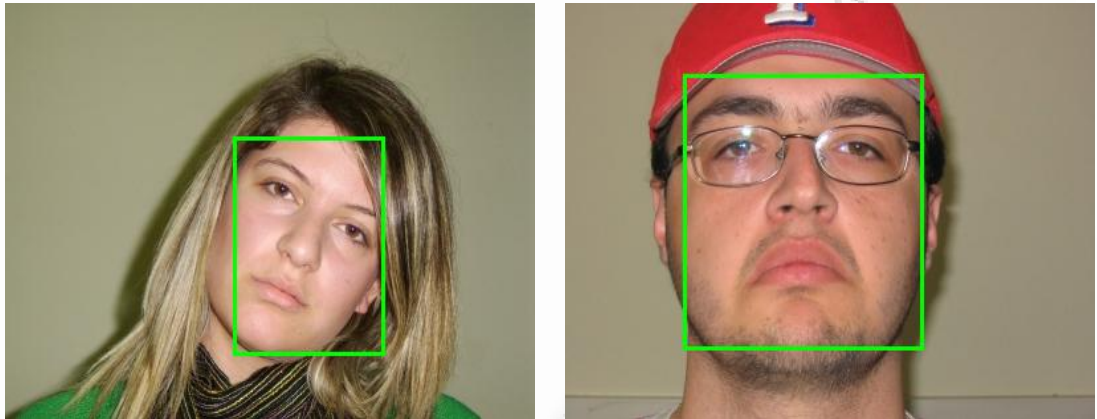


Figure 6.12 : Face Detection results for the second set of images, where the subject is forming an expression (a), (b)

'disappointed', 'screaming', 'angry', 'disgusted' and 'talk' expressions. Since we built a three-camera system to acquire the data, we also have faces in front and side view. In Table 6.4, we demonstrate some results from this third set of images. In the first



(a) 'Boredom-Sleepiness'










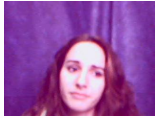


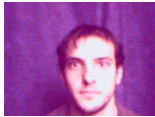


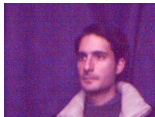


(b) 'Sadness'

Figure 6.13 : Face Detection results for the second set of images, where the subject is forming an expression (a), (b)

column we depict the image, while in the second column we demonstrate the part of the image which contains a human face. This part was used as input to the Face Detection Subsystem, so it was preprocessed (the result of preprocessing can be seen in third column) and fed to the artificial neural network. The network response is shown in the fourth column.







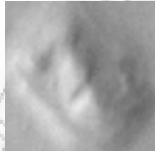





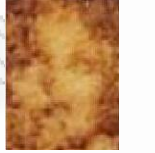

Except for the images which contained faces, the final test aimed at checking the system performance in cases where there were no faces. This test consisted of images that could confuse the neural network and lead it to a positive answer. So, in this set we tested our network with 50 images of pets and animals, complex backgrounds and parts of a face or human corpse. In Table 6.5, we demonstrate some results from this fourth set of images. Again, in the first column we depict the image, while in the second column we demonstrate the preprocessed image which is used as an input to the artificial neural network. The network's response is shown in the third column.

Table 6.4 : Sample images of our facial expression database

| Original Image | Window Pattern | Pre-processed image | Network response | Re- |
|---|---|---|------------------|-----|
|  |  |  | [1;0] | |
|  |  |  | [0,836;0,164] | |
|  |  |  | [0,9531;0,0469] | |
|  |  |  | [1;0] | |
|  |  |  | [1;0] | |
|  |  |  | [0.9865; 0.0135] | |

Finally, in terms of speed, the two networks are essentially comparable. Moreover, the first neural network has more neurons compared to the second, which explains why its speed may decrease, while the second needs more time for the segmentation

Table 6.5 : Sample images of the fourth test set - non human faces

| Image Details | Original Image | Pre-processed image | Network Response |
|---------------------------------------|---|--|------------------|
| Animal photo: Cat |  |  | [0;1] |
| Animal photo: Cat |  |  | [0;1] |
| Complex Back- ground |  |  | [0;1] |
| 'Face' on planet Mars |  |  | [0;1] |
| Animated 'Face' |  |  | [0;1] |
| 'Face' on a pumpkin |  |  | [0;1] |
| 'Face' of 'Virgin Mary' on a toast |  |  | [0;1] |

of a given image into the necessary pieces in order to produce the inputs. However, the delays, in the two cases, are negligible compared to the total time needed to pre-process the image.

6.1.5 Summary and Conclusions

We presented a neural network-based face detection algorithm and system [235, 236, 237, 5, 238]. To achieve higher face detection performance, we built two different neural network architectures and concluded to the use of the second, more complicated neural network architecture. Our system was tested using a training set of 285 face and non-face images of average to good quality, gathered from the World Wide Web. To test its performance, we used four different sets of images, namely: (1) images usually containing more than one faces (maybe partially occluded, slightly rotated or in side view) of various sizes, aspect ratios and resolution in complex backgrounds, (2) random images of our own facial expression database, in which the subjects form one of the seven expressions: 'neutral', 'happy', 'sad', 'surprised', 'disgusted', 'angry' and 'bored-sleepy', (3) low quality images acquired with a web camera and, (4) non face images of animals, human parts and complex backgrounds. Although the neural network had been trained with a set of images of higher (digital camera) quality, it was able to generalize and detected the faces in images at a quite satisfactory rate.

6.2 Introduction to our Facial Expression Recognition System

DEVELOPING a fully automated facial expression recognition system can be considered quite a challenging task as stated in the introduction of this Chapter. Towards building such a system, various tasks needed to be completed first. Specifically, in our first attempts for a facial expression recognition system, we tried to make use of the databases already available over the World Wide Web, as mentioned in Section 3.1, whereas the emotion classes, that our system would be able to recognize, had not yet been determined. These attempts are extensively presented in the following Section 6.3. As our work and study progressed, we settled in developing a facial expression recognition system that could be used for more advanced human-computer interaction techniques. This led us to the identification of the facial expressions corresponding to the ‘neutral’, ‘happy’, ‘sad’, ‘surprised’, ‘angry’, ‘disgusted’ and ‘bored-sleepy’ as the psychological states that arise very commonly during a typical human-computer interaction session. Moreover, we were able to rationalize and validate the facial features used towards this task, by making use of the questionnaires that had been collected during the empirical studies, presented in Chapter 5. This work is fairly described in Section 6.4.

6.3 First attempts for facial expression recognition

OUR first attempts for facial expression recognition can be identified by the following:

- Use of face databases gathered from World Wide Web: In our first efforts to build a facial expression recognition system, we used the AR Face Database [109] and the Cohn-Kanade AU-Coded Facial Expression Database [112]. As the system evolved and the aforementioned databases could no longer meet our needs, the development of our own facial expression database became mandatory and was finally adopted for our facial expression system.
- Different or fewer emotion classes: The use of the aforementioned databases and the fact that the emotion classes were not yet determined led us to the development of a facial expression recognition system which was adopted to the emotion classes dictated by the respective facial expression database used. As our work and study progressed, we settled in developing a facial expression recognition system that could be used for more advanced human-computer interaction techniques. This led us to the identification of the facial expressions corresponding to the ‘neutral’, ‘happiness’, ‘sadness’, ‘surprise’, ‘anger’, ‘disgust’ and ‘boredom-sleepiness’ as the psychological states that arise very commonly during a typical human-computer interaction session.
- Simpler / unsophisticated feature extraction algorithm: Another common was the use of a simpler feature extraction algorithm, based on binary images, as described in Section 6.3.1. As the study progressed, we developed our own Eye Detection Algorithm, as described in Section 6.4.2, on which was based our current feature extraction algorithm is based.

The algorithm we have developed [240, 241, 242, 243, 236, 5], as described in the following Subsection 6.3.1, was used for the majority of facial expression databases.

Some alterations were made to deal with low quality images [242, 236], as also described in this Subsection. Its performance was tested for various combinations of emotion classes, image databases and/or image qualities. The system performance and some results are shown in Subsection 6.3.4

6.3.1 The Facial Expression Classification Algorithm (1st Attempts)

In detail, our facial expression classification algorithm works as follows:

1. *We detect the front view of the face. The image region defined by the face detection step constitutes the ‘window pattern’ for our facial expression analysis module, which will be examined to determine the psychological state of the person.*
2. *We preprocess the ‘window pattern’:*
 - (a) *We apply Histogram Equalization techniques to enhance the contrast within the ‘window pattern’.*
 - (b) *We convert the image to binary and fill any holes in the binary image.*
3. *Feature Extraction: We extract basic corner points of the parts of the face and compute the Euclidean distances between them and certain specific ratios of these distances.*

Feature Extraction Algorithm

The main goal of feature extraction is to convert pixel data into a higher-level representation of shape, motion, color, texture and spatial configuration of the face

or its components. This representation is used for subsequent expression categorization. Feature extraction generally reduces the dimensionality of the input space. The reduction procedure retains essential information with high discrimination power and stability. The extracted feature vector consists of the corner points of the eyes, mouth and brows, respectively. The extracted features as well as the distances used to classify the expressions can be seen in Figure 6.14.

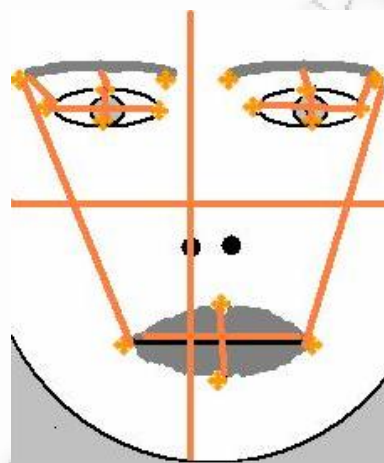


Figure 6.14 : The extracted features (orange points) and the calculated distances

Specifically, the feature extraction algorithm works as follows:

1. *We search the binary image of the face and extract each part of the face (eyes, mouth and brows) into a new image of the same size and coordinates as the original image.*
2. *In each image of a part of the face, we locate corner points using relationships between neighboring pixel values. This results in the determination of 16 points which form the feature vector. Typical results of the feature extraction algorithm*

are seen in Figure 6.15. In the first column, we can observe the ‘neutral’, ‘happy’ and ‘surprised’ facial expressions of a given person and, in the second column, the preprocessed image and the corresponding extracted features for each expression.

3. We compute the Euclidean distances between these points, depicted with orange in Figure 6.14, and certain specific ratios of these distances. The results constitute the input vector which is fed into the neural network.

Image preprocessing for low quality images

In our attempts at developing a most accurate facial expression recognition system, we tested our algorithm on low quality images. Specifically, we used our low quality database of facial expressions, which is further described in Chapter 4, Section 4.1. Most of the pictures acquired by our setup are of poor quality, mainly for three reasons:

- poor analysis of the web cameras (320-by-240 pixels)
- motion-caused blurring, which was the result of the relatively high capture time of the cameras and movements of the subject, and,
- poor lighting conditions when pictures were taken.

On purpose, we did not address any of the above problems by improving the data acquisition hardware, so as to make the image acquisition setup more realistic and closer to the operating conditions of practical human-machine interaction systems. Instead, we attempt to address these difficulties by identifying appropriate

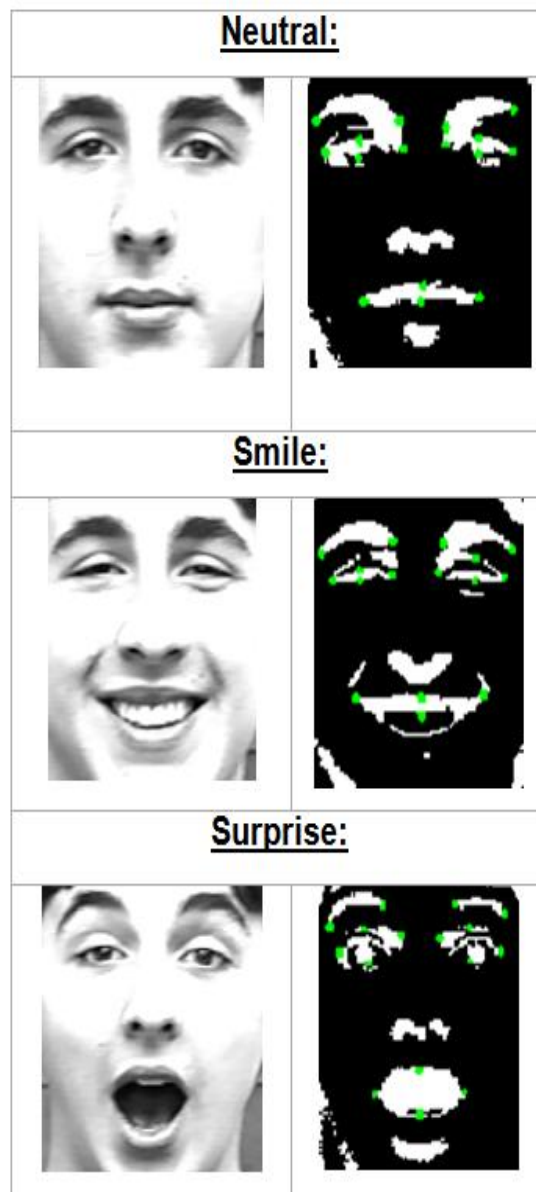


Figure 6.15 : Typical results from our feature extraction algorithm

image preprocessing algorithms which emphasize facial features, such as the eyes, the mouth and the brows, on which the face detection and facial expression classification

algorithms are based. Specifically, the preprocessing algorithms work as follows:

- 1. Load the acquired image and convert it to grayscale*
- 2. Compute a 3-by-3 unsharp/contrast enhancement filter from the negative of the Laplacian filter with parameter 0.2 and apply it to the input image*
- 3. Apply a 3-by-3 Gaussian lowpass filter to the resulting image*
- 4. Apply adaptive histogram equalization techniques to enhance the contrast between areas of the image*
- 5. Adjust image intensity values, so that 1% of data is saturated at low and high intensities of the input image. This further enhances the contrast in the resulting image.*
- 6. Apply a 3-by-3 Gaussian lowpass filter to the resulting image*
- 7. Perform two-dimensional median filtering to reduce noise and preserve edges.*

All the steps of the aforementioned algorithm are demonstrated for an example face image in Table 6.6. The resulting image is, then, converted to binary by the facial expression classification algorithm so as to extract features. As the main goal is to be able to discriminate facial features in the resulting binary image, we found that preprocessing was a necessary step to precede the face detection and facial expression classification algorithms. This becomes clear in the example image in Table 6.6, in the second and third column of the last row of which, we show the resulting binary

Table 6.6 : Demonstration of the preprocessing algorithm for low quality images










| | | |
|---|--|--|
| <p><u>Step 1:</u> Image converted to grayscale</p>  | <p><u>Step 2:</u> Unsharp Filter</p>  | <p><u>Step 3:</u> Gaussian Filter</p>  |
| <p><u>Step 4:</u> Adaptive Histogram Equalization</p>  | <p><u>Step 5:</u> Adjusted Image</p>  | <p><u>Step 16:</u> Gaussian filter</p>  |
| <p><u>Step 7:</u> Median Filter</p>  | <p>Preprocessed Resulting Binary Image</p>  | <p>Resulting Binary Image without Preprocessing</p>  |

image with and without preprocessing, respectively. The threshold level for both of these two latter images is 0.4.

6.3.2 Feature Validation (First Attempts)

The location and shape of face parts vary significantly from facial expression to facial expression. During the development of our system, various emotion classes and

facial expression databases were used. Specifically, the following systems were developed/tested:

- 2-class system for ‘neutral’ and ‘screaming’ expressions [240], using the AR Face Database [109] for training and testing.
- 3-class system for ‘neutral’, ‘happy’ and ‘surprised’ expressions [241], using Cohn-Kanade AU-Coded Facial Expression Database [112] for training and testing.
- 3-class system for ‘neutral’, ‘happy’ and ‘surprised’ expressions [242], using Cohn-Kanade AU-Coded Facial Expression Database [112] for training and our low quality facial expression database for testing.
- 3-class system for ‘neutral’, ‘happy’ and ‘surprised’ expressions [243], using Cohn-Kanade AU-Coded Facial Expression Database [112] for training and our low quality facial expression database for testing for the extension of our system for faces in side view.
- 3-class system for ‘neutral’, ‘happy’ and ‘surprised’ expressions [236], using Cohn-Kanade AU-Coded Facial Expression Database [112] for training and testing using more features.

The features among the aforementioned three main classes (‘neutral’, ‘happy’ and ‘surprised’) contain high discrimination power. The distribution of these classification features for 83 subjects forming each of the three expressions, respectively, is shown in Figures 6.17 and 6.16. Specifically in Figure 6.17, we plot the distribution of

the mouth dimension ratio for each of the examined facial expressions. Similarly in Figure 6.16, we plot the distribution of the face dimension ratio relatively to ‘neutral’ expression.

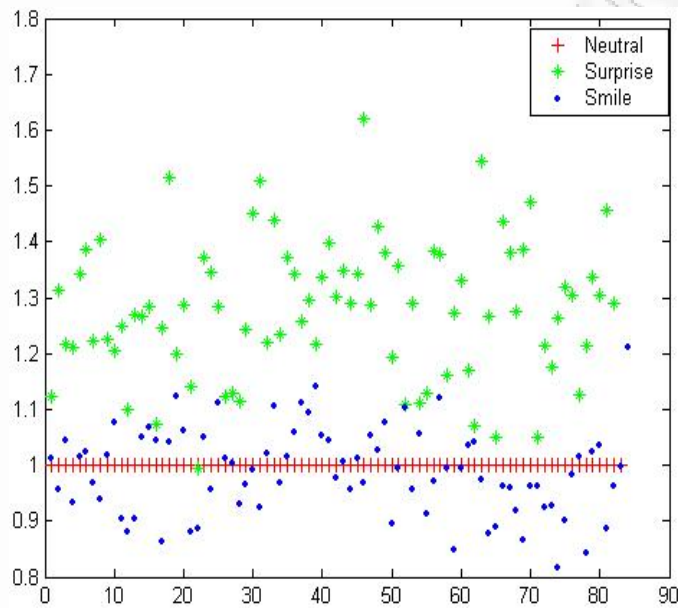


Figure 6.16 : Facial Dimension Ratio Distribution

6.3.3 Neural Network Classifiers (First Attempts)

Based on the requirements set from the databases and the emotion classes we wanted to identify, we built various neural networks. In the case of 2-class system, we built a simple neural network with 2 hidden layers, with 3 and 2 neurons, respectively. The neural network took as input the three main extracted features: (1) left eye ratio, (2) right eye ratio, and (3) mouth ratio. To train the neural network, we used a set of 250 images, 125 images for each expression. These images were contained in the AR

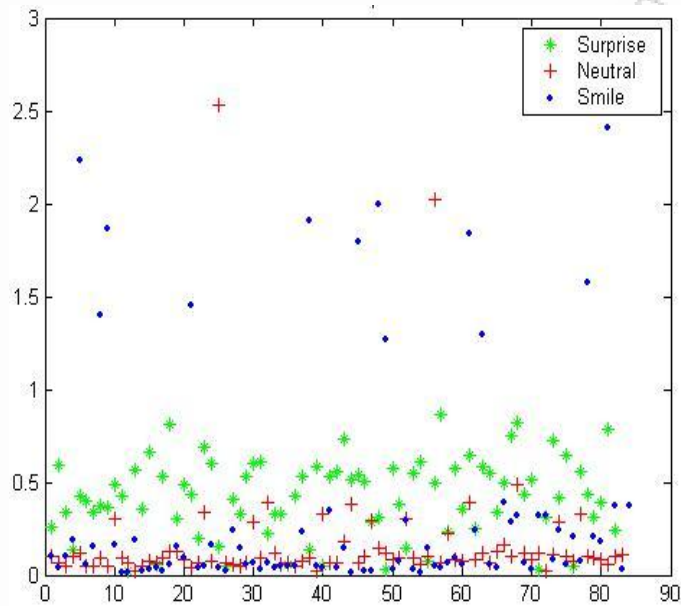


Figure 6.17 : Mouth Dimension Ratio Distribution

Face Database [109] and preprocessed before entered into the neural network. After training with this image set, the neural network achieved an error rate of 0.016. In fact, this error rate was gradually reduced as the number of the images in the training set increased. Given the relatively small size of the training set, this error rate is quite satisfying. According to the requirements set, when the window pattern represented a neutral facial expression, the neural network should produce an output value of 'one' ('1') (or something close to this value). On the other hand, if it represented a screaming face, the output value should be 'zero' ('0'). The output value can be regarded as the degree of membership of the face image in each of the 'neutral' and 'screaming' classes.

In the case of 3-classes system, we built a simple neural network with 2 hidden

layers, with 4 and 3 neurons, respectively. The neural network took as input the three main extracted features: (1) left eye ratio, (2) right eye ration, (3) mouth ratio, and (4) face size ratio. To train the neural network, we used a set of 249 images, 83 images for each expression. These images where gathered from the Cohn-Kanade AU-Coded Facial Expression Database [112] and preprocessed before entered into the neural network. After training with this image set, the neural network achieved an error rate varying between 0.01 and 0.016, depending on the neural network architecture and the extracted features fed in the network. Again, this error rate was gradually reduced as the number of images in the training set increased. Given the relatively small size of the training set, this error rate is quite satisfying.

According to the requirements set, when the window pattern represented a ‘neutral’ facial expression, the neural network should produce an output value of [1;0;0] or so. Similarly, for the ‘happy’ expression, the output must be [0;1;0] and for the ‘surprised’, [0;0;1]. The output value can be regarded as the degree of membership of the face image in each of the ‘neutral’, ‘happy’ and ‘surprised’ classes.

6.3.4 Results from neural network classifiers (First Attempts)

We tested our neural network classifiers in each case mentioned above. Some results are depicted in the following Figures 6.18, 6.19, 6.20 and 6.21.

Specifically, in Figure 6.18, we depict the results for our 2-class system for ‘neutral’ and ‘screaming’ expressions [240], using the AR Face Database [109] for training and testing. The system was tested with images of 10 subjects, some of them were already in the training database, as there were two sequences of image shooting of the same subject in the AR Face Database [109]. The result was 20 images, and the neural

network showed an accuracy of 100%.





| <u>Original Image:</u> | <u>Extracted Features:</u> | <u>Network's Response:</u> |
|--|--|----------------------------|
| <u>Scream:</u> | | |
|  |  | 0 |
| <u>Neutral:</u> | | |
|  |  | 0.867 |

Figure 6.18 : Results from our 2-class system

Moreover, in Figure 6.19, we depict the results for our 3-class system for ‘neutral’, ‘happy’ and ‘surprised’ expressions [241], using Cohn-Kanade AU-Coded Facial Expression Database [112] for training and testing. The system was tested with images of 15 subjects. The result was 45 images, and the neural network showed an accuracy of 80%.

In Figure 6.20, we depict the results for our 3-class system for ‘neutral’, ‘happy’ and ‘surprised’ expressions [242], using Cohn-Kanade AU-Coded Facial Expression Database [112] for training and our low quality facial expression database for testing. The system was tested with low quality images of 15 subjects. The result was again







| <u>Original Image:</u> | <u>Extracted Features:</u> | <u>Network's Response:</u> |
|--|--|----------------------------|
| <u>Smile:</u> | | |
|  |  | [0;0,899; 0,101] |
| <u>Neutral:</u> | | |
|  |  | [1;0;0] |
| <u>Surprise:</u> | | |
|  |  | [0;0,022; 0,978] |

Figure 6.19 : Results from our 3-class system with the Cohn-Kanade Database

45 images, and the neural network showed an accuracy of 77%.

Finally, the same system was tested for its performance on side view images of low quality. In Figure 6.21, we depict the results for our 3-class system for 'neutral', 'happy' and 'surprised' expressions [243], using Cohn-Kanade AU-Coded Facial Expression Database [112] for training and our low quality facial expression database for testing for the extension of our system for faces in side view. The system was tested with low quality images of the same 15 subjects, as above. The result was again 45 images, and the neural network showed again an accuracy of 77%.










| | Window Pattern | Binary Image | Extracted Features | Network's response |
|----------|---|---|--|---------------------|
| Neutral |  |  |  | [0.765;0.232;0.003] |
| Smile |  |  |  | [0.128;0.805;0.067] |
| Surprise |  |  |  | [0;0.186;0.814] |

Figure 6.20 : Results from our 3-class system with the Cohn-Kanade Database for training and our low quality database for testing



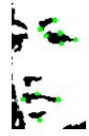





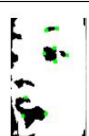
| | Window Pattern | Binary Image | Extracted Features | Network's response |
|----------|---|---|--|-----------------------|
| Neutral |  |  |  | [0.622; 0.490; 0.052] |
| Smile |  |  |  | [0.395; 0.6; 0.005] |
| Surprise |  |  |  | [0; 0.060; 0.940] |

Figure 6.21 : Results from our 3-class system with the Cohn-Kanade Database for training and our low quality, side view images for testing

6.4 Facial expression recognition system

AS our work progressed and with the results available from our empirical studies, described in previous Chapter 5, we considered additional emotion classes in our system [244, 245, 246, 247, 248]. Indeed, based on our studies, facial expressions corresponding to the ‘neutral’, ‘happiness’, ‘sadness’, ‘surprise’, ‘anger’, ‘disgust’ and ‘boredom-sleepiness’ psychological states arise very commonly during a typical human-computer interaction session, as stated in Chapter 4. Our final facial expression recognition system is identified by the following:

- More facial features are extracted from the image, such as measurements of the texture, head orientation, etc.
- We use our own facial expression database for training and/or testing, as it is more complete in terms of the classes we want to classify.
- A better, more sophisticated, algorithm to extract the facial features is used, which is based on our eye detection/extraction algorithm, described below.

6.4.1 Feature Selection

The first step in order to compute the needed features, is to identify some important facial points. These facial points are widely used in facial image processing systems and can help us in the computation of the facial features which will be used as input to an artificial neural network. The facial points are summarized in Figure 6.22.

From the collected dataset, we identified differences between the ‘neutral’ expression of a model and its deformation into other expressions, as typically high-lighted

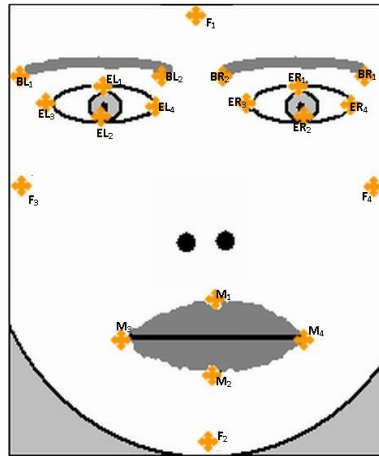


Figure 6.22 : The most important facial points which will help us in the extraction of the feature vector for the facial expression recognition task

in Table 6.7.

The changes described in Table 6.7, may show differences among different persons. From our observations we identified the following differences:

- The changes in the texture of the skin from the formation of wrinkles (in cheeks, forehead, cheeks, etc) depend of the person's skin quality and age. In younger people, wrinkle formation is not so common.
- Usually, 'bored-sleepy' people tend to slightly rotate their head, narrow their eyes and keep their mouth closed (in a away reminiscent of 'sadness'). But there is also a significant number of people who tend to yawn because they feel sleepy
- The vast majority of people, when they are angry, tend to narrow their eyes, but, there is also a number of people who tend to open their eyes wider.
- When someone is disgusted, usually he/she forms an expression as described in

Table 6.7 : Deformations of the other six expression, compared to ‘neutral’







| Variations between Facial Expressions: | | | |
|---|---|---|--|
| Happiness | | Boredom-Sleepiness | |
|  | (1)Bigger-broader mouth (2)Slightly narrower eyes (3)Changes in the texture of the cheeks (4)Occasionally, changes in the orientation of brows |  | (1)Head slightly turned downwards (2)Eyes slightly closed (3)Occasionally, wrinkles formed in the forehead and different direction of the brows (4)Occasionally, mouth opened (subject is yawning) |
| Surprise | | Sadness | |
|  | (1)Longer head (2)Bigger-wider eyes (3)Opened mouth (4)Wrinkles in the forehead (changes in the texture) (5)Changes in the orientation of eyebrows (the eyebrows are raised) |  | (1)Changes in the direction of the mouth (2) Wrinkles formed on the chin (different texture) (3)Occasionally, wrinkles formed in the forehead and different direction of the brows |
| Anger | | Disgust | |
|  | (1)Wrinkles between the eyebrows (different textures) (2)Smaller (narrower) eyes (3)Wrinkles in the chin (4)The mouth is tight (5)Occasionally, wrinkles over the eyebrows, in the forehead |  | (1)The distance between the nostrils and the eyes is shortened (2)Wrinkles between the eyebrows and on the nose (3)Wrinkles formed on the chin and the cheeks |

Figure 6.7, but, there is also a number of people who turn their head in such cases.

Based on the above observations, we can compute the effect of each facial action on the facial points we identified earlier. Also, we can identify the possible facial

expression, that each facial action may belong to. Finally, we can identify the facial features that we need to compute. All these, are summarized in Tables 6.8, 6.9 and 6.10.

Table 6.8 : Facial action and resulting Facial Features

| Facial Feature Action | Result to Facial Points | Possible expression - emotion | Identification of the needed facial feature |
|--|--|---|---|
| Broader mouth | $ M_3M_4 _{max}$ $ M_1M_2 _{min}$ | Happiness | Mouth Ratio |
| Open mouth | $ M_3M_4 _{max}$ $ M_1M_2 _{min}$ | Boredom-Sleepiness (-) Surprise | |
| The mouth is tight | $ M_1M_2 _{min}$ | Sadness Anger | |
| Changes in the direction of the mouth | mouth orientation down mouth orientation up | Sadness Boredom - Sleepiness (-) Disgust - Disapproval (-) Happiness | Mouth Orientation |
| Slightly narrower eyes , Eyes slightly Closed | $ EL_1EL_2 _{min}$ $ ER_1ER_2 _{min}$ | Happiness Boredom - Sleepiness Disgust - Disapproval | Eyes Ratio |
| Bigger, wider eyes | $ EL_1EL_2 _{max}$ $ ER_1ER_2 _{max}$ | Surprise Anger (-) | Eyes Ratio |

These observations led us to the identification of the following facial features:

Table 6.9 : Facial action and resulting Facial Features - 2

| Facial Feature Action | Result to Facial Points | Possible expression - emotion | Identification of the needed facial feature |
|--|--|---|---|
| Wrinkles in the cheeks | Changes in texture | Happiness Boredom - Sleepiness(-) Sadness(-) Disgust - Disapproval(-) | Texture of the cheeks |
| Changes in the direction of the eyebrows | brows orientation up brows orientation down | Sadness Boredom - Sleepiness (-) Surprise (-) Anger Disgust - Disapproval (-) | Brow Orientation |
| Head slightly turned downwards | head orientation changed | Boredom - Sleepiness Sadness (-) | Head Orientation |
| Longer head | $ F_1 F_2 _{max}$ | Surprise Boredom - Sleepiness (-) | Head ratio |
| Broader head | $ F_3 F_4 _{max}$ | Happiness | Head ratio |

1. Mouth Ratio:

$$\frac{||M_1 M_2|| / ||M_3 M_4||}{||M_{1Neu} M_{2Neu}|| / ||M_{3Neu} M_{4Neu}||}$$

2. **Mouth Orientation:** Measurement of the changes of the orientation of the mouth compared to 'neutral' expression

Table 6.10 : Facial action and resulting Facial Features - 3

| Facial Feature Action | Result to Facial Points | Possible expression - emotion | Identification of the needed facial feature |
|--------------------------------|-------------------------|---|---|
| Wrinkles on the forehead | Changes in texture | Anger(-) Boredom - Sleepiness(-) Sadness(-) Disgust - Disapproval(-) | Texture of the forehead |
| Wrinkles on the chin | Changes in texture | Anger(-) Sadness(-) Disgust - Disapproval(-) | Texture of the chin |
| Wrinkles between the eye-brows | Changes in texture | Anger(-) Disgust - Disapproval(-) | Texture of the region between eyebrows |
| Wrinkles on the nose | Changes in texture | Disgust - Disapproval(-) | Texture of the nose |

3. Left Eye Ratio:

$$\frac{||EL_1EL_2||/||EL_3EL_4||}{||EL_{1Neu}EL_{2Neu}||/||EL_{3Neu}EL_{4Neu}||}$$

4. Right Eye Ratio:

$$\frac{||ER_1ER_2||/||ER_3ER_4||}{||ER_{1Neu}ER_{2Neu}||/||ER_{3Neu}ER_{4Neu}||}$$

5. **Texture of the left cheek:** Measurement of the changes of the texture of the left cheek compared to 'neutral' expression

6. **Texture of the right cheek:** Measurement of the changes of the texture of the right cheek compared to 'neutral' expression

7. **Left Brow Orientation:** Measurement of the changes of the orientation of the left brow compared to ‘neutral’ expression
8. **Right Brow Orientation:** Measurement of the changes of the orientation of the right brow compared to ‘neutral’ expression
9. **Head ratio:**

$$\frac{\|F_1F_2\|/\|F_3F_4\|}{\|F_{1Neu}F_{2Neu}\|/\|F_{3Neu}F_{4Neu}\|}$$
10. **Texture of the forehead:** Measurement of the changes of the texture of the forehead compared to ‘neutral’ expression
11. **Texture of the chin:** Measurement of the changes of the texture of the chin compared to ‘neutral’ expression
12. **Texture of the region between the eyebrows:** Measurement of the changes of the texture of the region between the eyebrows compared to ‘neutral’ expression

The facial features are summarized in the Figure 6.23

6.4.2 Image Preprocessing and Feature Extraction

To convert pixel data into a higher-level representation of shape, motion, color, texture and spatial configuration of the face and its components, we locate and extract the corner points of specific regions of the face, such as the eyes, the mouth and the brows, and compute their variations in size, orientation or texture between the neutral and some other expression. This constitutes the feature extraction process and reduces the dimensionality of the input space significantly, while retaining essential

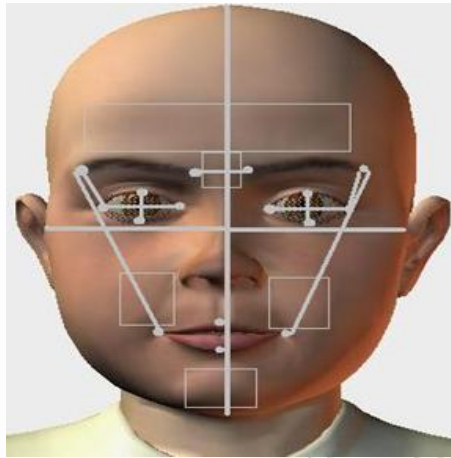


Figure 6.23 : The extracted features

information of high discrimination power and stability. The extracted features, the face regions, and the dimension ratios used to classify the expressions are summarized in Fig. 6.23.

The main algorithm is as follows:

1. *Eye Extraction*
2. *Based on the location of the eyes, extraction of the other facial features and facial regions.*
3. *Combination of all and computation of feature vector*

Specifically, the basic aim of this process is to extract the needed corner point of the facial features and the respective facial regions. In order to achieve this, we first locate the eyes of the person using a fairly complex algorithm. Based on the

location of the eyes, we then extract the rest of the facial feature points which will lead us to the computation of the resulting feature vector. We describe each algorithm separately in the following sections.

The eye extraction algorithm

We apply our eye detection algorithm [244] in the upper 60% of the detected face region. We do not process the lower 40% of the face region to decrease the complexity of the algorithm and the required computational effort. The 60% of the upper detected face is selected based on our studies so as to cover cases of face rotation. For better accuracy, the algorithm computes two different binary images, called 'skin map' and 'clustered image,' respectively, and uses them to detect the eyes. Specifically, the algorithm follows four main steps:

- 1. Skin extraction*
- 2. K-means clustering*
- 3. Combination of the resulting images and morphological processing*
- 4. Feature extraction*

Skin Extraction

The skin filter is based on the Fleck and Forsyth algorithm [249]. The input color image must be in RGB format with color intensity values ranging from 0 to 255. The algorithm, works as follows:

1. The RGB image is transformed to log-opponent values I , Rg , and By , given by the Fleck and Forsyth algorithm, as follows:

- $I = L(G)$
- $Rg = L(R) - L(G)$
- $By = L(B) - (L(G) + L(R))/2$

The $L(x)$ operation is defined as $L(x)=105*\log(x+1)$. The log transformation makes the Rg and By values, as well as differences between I values (e.g. texture amplitude), independent of illumination level.

2. After filtering the Rg and By matrices, a texture amplitude map is used to find regions of low texture information. Usually, the skin is very smooth, so the skin regions are those with little texture.
3. In these selected areas, we further select the skin region based on the measures of hue and saturation, so as their color matches that of skin. The acceptable values of hue and saturation, are $110 \leq \text{hue} \leq 180$ and $0 \leq \text{saturation} \leq 130$, respectively.
4. A binary skin map is drawn, where if the pixel in the original image is in the same coordinates as the pixel map is skin, it is represented with 1, or 0 otherwise. The skin map array can be considered as a black and white binary image with skin regions appearing as white. The resulting 'skin map' usually represents the eyes, brows, nostrils, hair and other objects on the face (e.g. glasses), with white regions and the skin with black. Some results after applying the algorithm are shown in Figure 6.24.

K-means Clustering

On the same 60% of the detected face region, we compute 3 clusters of the image, color each cluster with the corresponding average color, and, finally, convert it to binary. The resulting image is called ‘clustered face image’ and, as the ‘skin map’, represents the eyes, brows, nostrils, hair and other objects on the face (e.g. glasses), as white regions and the skin as black. Some results, corresponding to the faces in the skin extraction step, are shown in Figure 6.24.










| Eye Extraction with Skin Extraction and K-means Clustering | | | |
|---|---|---|--|
| | 60% of the Detected Face | Skin Map | K-Means Clustered Image |
| Surprise |  |  |  |
| Angry with eye glasses |  |  |  |
| Sad |  |  |  |

Figure 6.24 : Eye Extraction with Skin Extraction and K-means Clustering

Morphological Operations-Combining The Two Images

On the two resulting images, we apply morphological operations. Our aim is to remove all other objects and to end up with a binary image of only the eyes, so as to make the eye detection task simpler. The algorithm, works as follows:

- 1. We apply a window, which clears the boundary pixels in the input image (usually representing the hair and the nostrils)*
- 2. We remove areas whose size is too small. This removes some very small objects on the face (e.g. scars)*
- 3. We remove the areas whose length is larger than the 1/3 of the total row size. In this step, wide areas, e.g. the skeleton of the glasses are removed.*
- 4. Finally, we remove the areas whose length is very small to remove other small objects of the face.*

The image resulting at each step of the aforementioned algorithm is shown in Figure 6.25, in which the input image was the skin map of the second image (i.e., the image of the where the person wearing glasses). The result is an image containing only the two eyes.

The detection is done on the two images. The final location of the eyes is found based on characteristics of the detected areas in the two images, e.g. the relative position of the eyes and their size, in relation to the original image. Each eye is depicted to a binary image, where the features are extracted next.

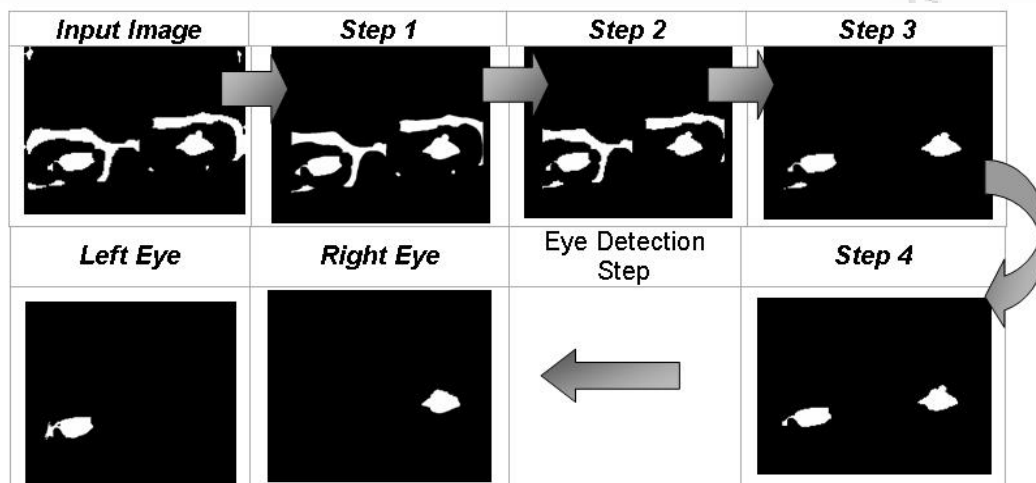


Figure 6.25 : Morphological Operations

Feature Extraction After detecting the eyes, we end up with two binary images representing the left and the right eye, respectively. First, we trace the outline of the eye area. Non-zero-valued pixels are assigned to an object and zero-valued pixels constitute the background. The curve is drawn based on these relative values. To compute edge points, first we find the minimum and the maximum coordinates of the computed contour. Then, the center points are computed. Finally, the edge points are computed based on their relationships relatively to the center points and the contour. Finally, the coordinates of the extracted features are drawn in the original (color) image of the original size. The extracted edge points, in the original image, are depicted in Figure 6.26.

Some results of our eye extraction/localization algorithm are shown in Figure 6.27.

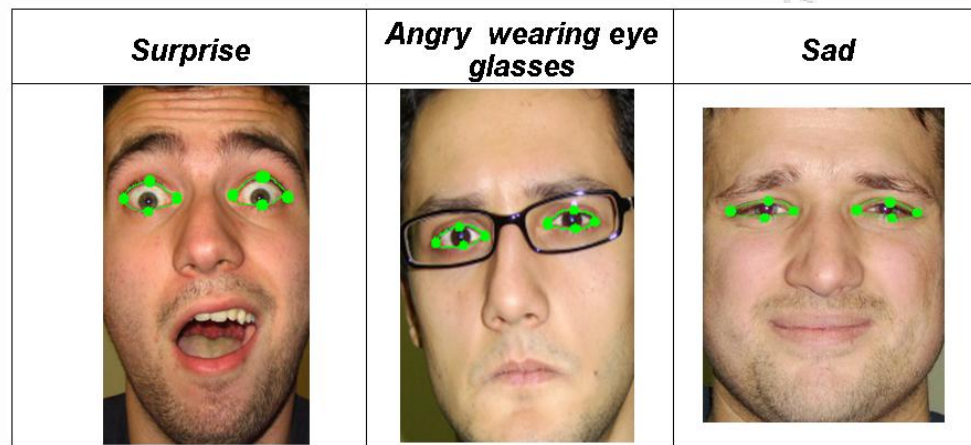


Figure 6.26 : Extracted Eye Features



Figure 6.27 : Some results of the eye extraction algorithm

6.4.3 The extraction algorithm for the rest of facial features

After successful eye detection and eye feature extraction, the rest of the needed facial portions are detected based on the given location of the eyes. Specifically, we apply

certain rules in order to locate and extract the rest of the facial features:

- Extraction of the features of the brows: (1) The brows are considered to be a light region above the eyes in the skin map image, after having applied the second step of morphological operations described above, as we can see in Figures 6.24 and 6.25. (2) In order to extract the corner points, we apply the same algorithm as the algorithm used for the extraction of the corner points of the eyes.
- Extraction of the features of the mouth: (1) The location of the mouth is computed in the remaining 40% of the face image. (2) The eyes and the mouth form a 'T' independently of the orientation of the face. The region of the mouth is located based on this rule. (3) After we have detected the location of the mouth, in order to extract the corner points, we apply the same algorithm as the algorithm used for the extraction of the corner points of the eyes.
- Extraction of the region of the chin: (1) The location of the region of the chin is computed in the remaining 40% of the face image. (2) This region starts just after the feature point M_2 of the mouth. (3) Its width is 90% of the width of the mouth and its height 18% of the height of the face.
- Extraction of the region of the cheeks: (1) The location of the region of the cheek is computed in the entire face image. (2) This region starts just after the feature points M_3 and M_4 of the mouth, for each side, respectively. (3) Its width is 18% of the width of the face and its height 18% of the height of the face.

- Extraction of the region between the brows: (1) The location of the region between the brows is computed in the entire face image. (2) This region is between the feature points BL_2 and BR_2 . (3) Its width is 10% of the width of the face and its height 10% of the height of the face.
- Extraction of the region of the forehead: (1) The location of the region between the brows is computed in the entire face image. (2) This region is above the brows. (3) Its width is 80% of the width of the face and its height 12% of the height of the face.

6.4.4 Combination of all and computation of feature vector

After successful feature point and region detection, we depict these corner points and facial regions on the original image. We compute the Euclidean distances between these points, depicted with lines in Figure 6.23, and certain specific ratios of these distances. We also compute the orientation of the brows and the mouth. Finally, we compute a measure of the texture for each of the specific regions based on the texture of the corresponding ‘neutral’ expression. All the above measurements correspond to the features described in Section ‘Feature Selection’. The computation of these features leads to the formation of a 12-by-1 feature vector with high discrimination power. This feature vector is fed to an artificial neural network in order to classify the expression.

6.4.5 Quantification of Feature Discrimination Power

The selected features contain high discrimination power and will help us towards the task of facial expression recognition. In the following Figures 6.28, 6.29, 6.30, 6.31,

6.32, 6.33 and 6.34, we demonstrate the probability density for some of the major features for 250 subjects of our Facial Expression Database, which we use in our facial expression recognition system.

Specifically, in Figure 6.28, we demonstrate the probability density for the face size ratio for each of the seven expressions. As we can observe, there is a significant difference in the values for the ‘surprise’ and ‘disgust’ emotion (green and magenta line, respectively), compared to the ‘happiness’ and ‘boredom-sleepiness’ emotion (red and dark blue line, respectively). This is done because in the former cases people tend to open their mouth and/or eye, so the face ratio become smaller. On the other hand, in the latter cases, the face becomes wider. In the ‘anger’ and ‘sadness’ emotion, the values of the face ratio are similar to ‘neutral’

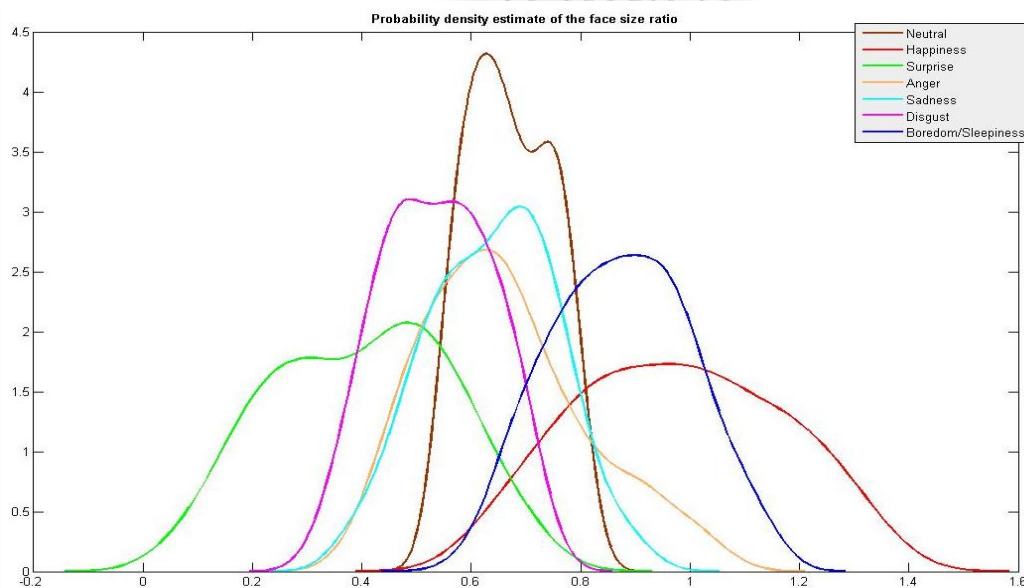


Figure 6.28 : Probability Density of the ‘Face Size Ratio’

The previous assumption is further illustrated in Figure 6.29, where, again, there

is a clear difference between the ‘happiness’ and the ‘surprise’ emotion.

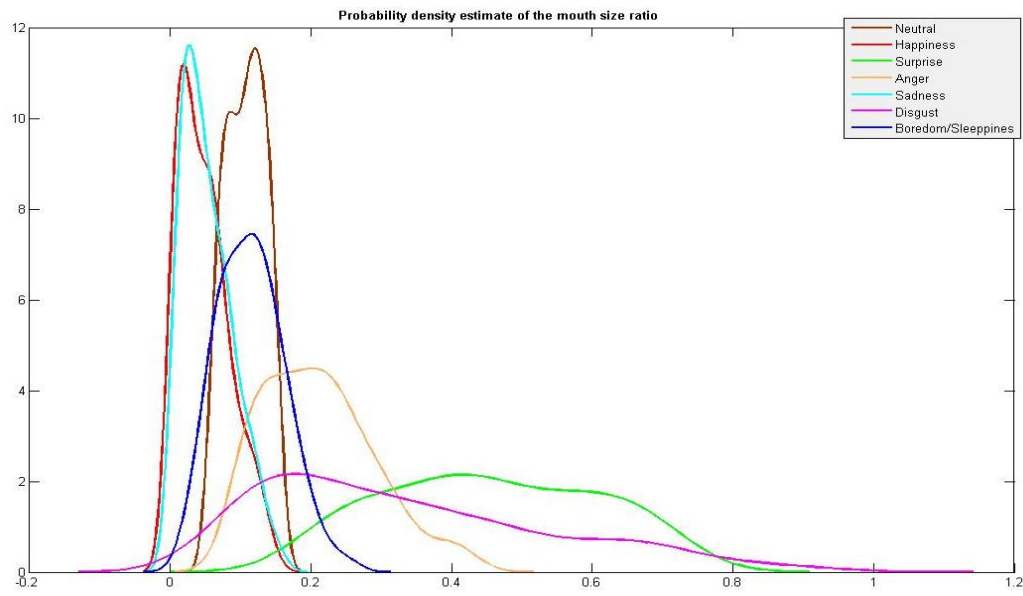


Figure 6.29 : Probability Density of the ‘Mouth Size Ratio’

In the following Figures 6.30 and 6.31, we observe the probability density estimate for the left and right eye size ratio, respectively. As we can observe, the values follow the same graph for the two eyes. This is because there is no ‘expression’ in our expression classes that would oblige the subject to treat one of his/her eyes differently from the other (e.g. to blink).

As far as the collected side view images are concerned, our study showed that formation of some expressions involves deformation of a person’s head sides and, thus, additional classification features may be derived from side view face images. In fact, features may be more evident in side view rather than front view images for certain expressions. For example, better discrimination between the ‘neutral’ and ‘happy’

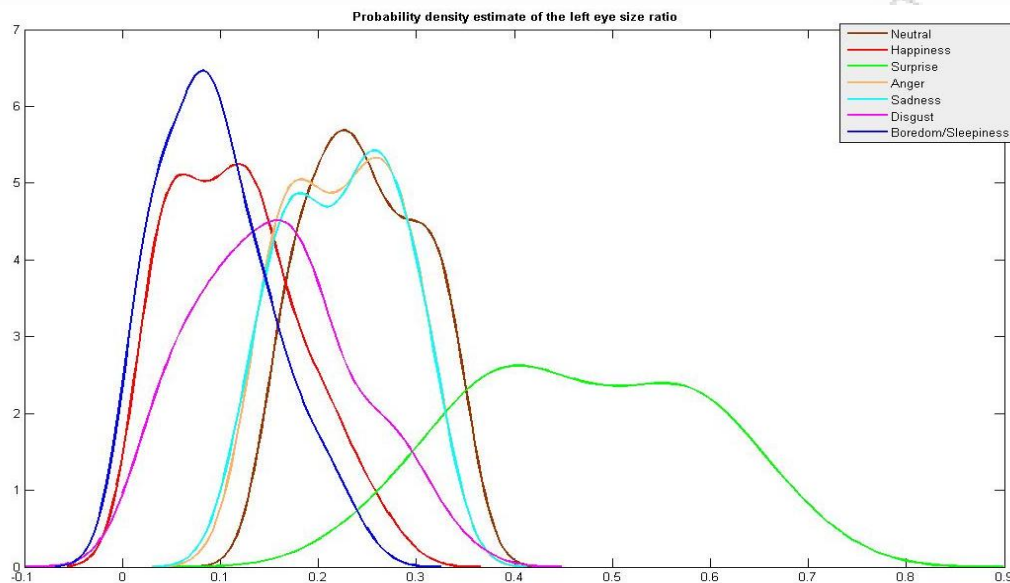


Figure 6.30 : Probability Density of the ‘Left Eye Size Ratio’

expression seems to be achieved in front view images, whereas the ‘surprised’ expression seems to be better identified in side view images. Similarly, the ‘sad’ and angry’ are better discriminated in front rather than in side view images as forehead texture, one of the corresponding classification features, is better computed in front rather than side view images. Thus, we conclude that better facial expression classification results can be achieved by using images of several views of a person’s face.

6.4.6 Classifiers for Facial Expression Classification

Towards building our facial expression recognition system, we developed several classifiers. In our first attempts, we used neural network-based classifiers. Later, we also tested the performance of facial expression recognition, by developing a more sophisticated classifiers, using the NetLab Tool of Matlab. All the classifiers are described

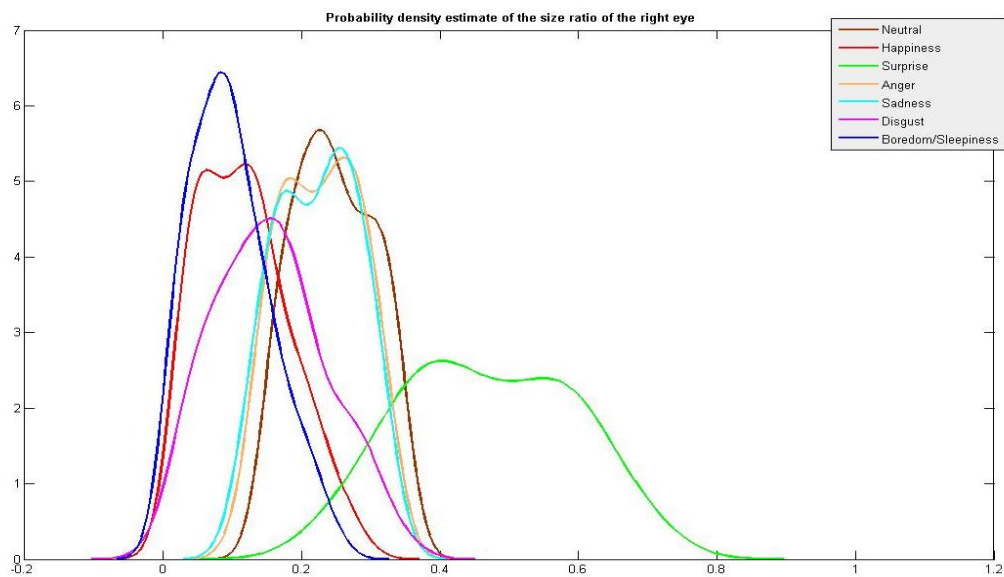


Figure 6.31 : Probability Density of the 'Right Eye Size Ratio'

and tested in the following sections.

Neural Network Architecture

In order to classify facial expressions, we developed a two layer artificial neural network which is fed with the following input data: (1) mouth dimension ratio, (2) mouth orientation, (3) left eye dimension ratio, (4) right eye dimension ratio, (5) measurement of the texture of the left cheek, (6) measurement of the texture of the right cheek, (7) left eye brow direction, (8) right eye brow direction, (9) face dimension ratio, (10) measurement of the texture of the forehead, (11) measurement of the texture of the region between the brows, and, (12) measurement of the texture of the chin. The network produces a 7-dimensional output vector which can be regarded as the degree of membership of the face image in each of the 'neutral', 'happiness', 'surprise', 'anger',

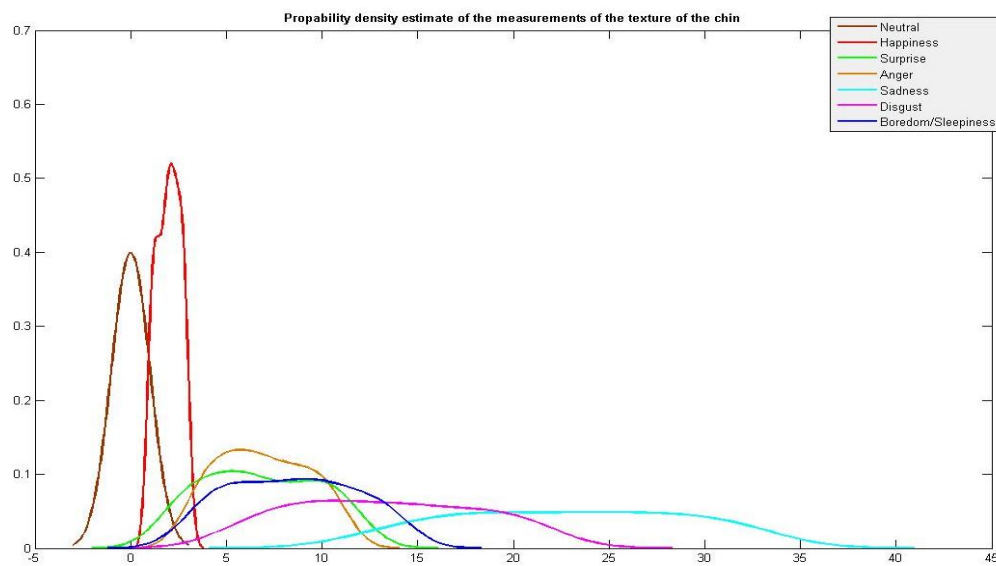


Figure 6.32 : Probability Density of the ‘Texture of the Region of the Chin’

‘disgust-disapproval’, ‘sadness’ and ‘boredom-sleepiness’ classes. An illustration of the network architecture can be seen in Figure 6.35. The neural network-based facial expression recognition system, is called **NEU-FACES** [244, 245, 246, 248].

6.4.7 Classification Performance Assessment

NEU-FACES managed to classify the emotions on a person’s face quite satisfactorily. The neural network was trained with a dataset of 230 subjects forming the 7 expression samples from all the emotion classes, i.e., a total of 1610 face images. We tested the classifier with images from 20 subjects forming the 7 facial expressions corresponding to 7 equivalent emotions, a total of 140 images. The results are summarized in Table 6.11. In the first column we demonstrate the emotion classes, while in the following three columns, we show the results of our empirical studies to humans.

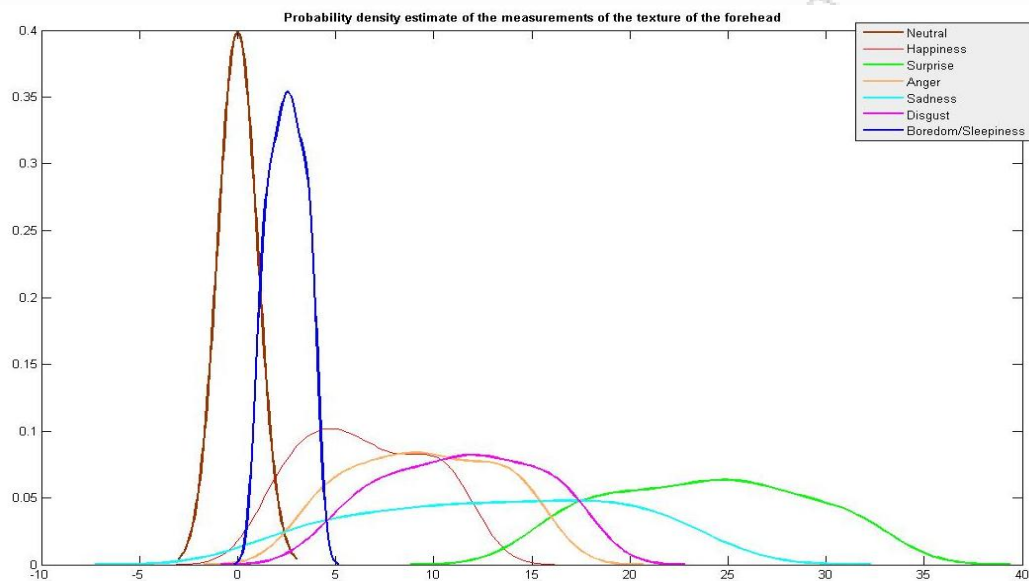


Figure 6.33 : Probability Density of the ‘Texture of the Region of the Forehead’

Specifically, results from the first part of the questionnaire are shown in the second column, results from the second part are shown in the third column, while the mean success rate is shown in the fourth. In the fifth column, we depict the success rate of our NEU-FACES recognition system for the corresponding emotion.

As we can observe, the NEU-FACES achieved higher success rates in most of the emotion compared to the success rates achieved by humans, with exception to the ‘anger’ emotion, where it achieved a success rate of only 55%. This is due mostly to pretence and to the the difficulty of humans to express such an emotion strongly. The latter is further corroborated by the fact that the majority of the face images depicting ‘anger’ that were erroneously classified by our system were misclassified as ‘neutral’. Generally, the NEU-FACES achieve very good results in positive emotions, such as ‘happiness’ and ‘surprise’, where it achieved success rates of 90% and 95%,

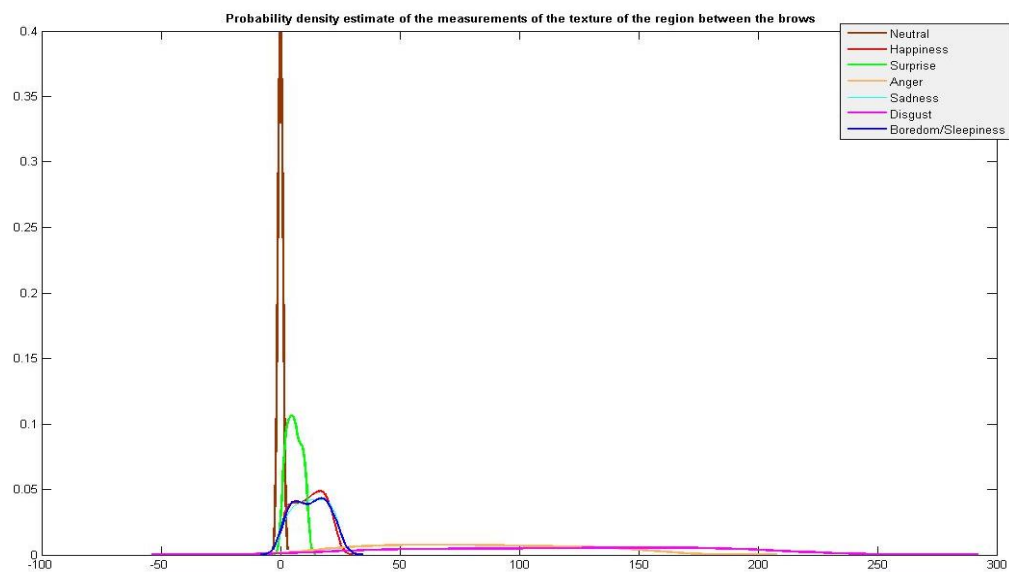


Figure 6.34 : Probability Density of the ‘Texture of the Region Between the Brows’

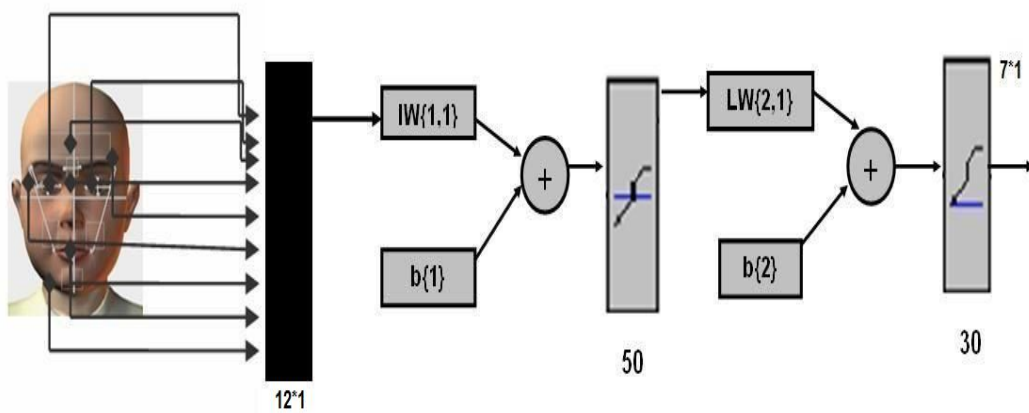


Figure 6.35 : The Facial Expression Neural Network Classifier

respectively.

Although the resulting neural network classifier achieved quite good results in

Table 6.11 : Results of the Facial Expression Classification System Compared to Human Classifiers


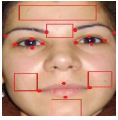

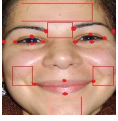
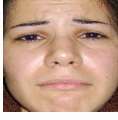
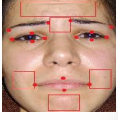
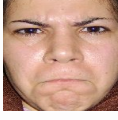
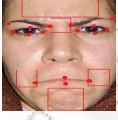
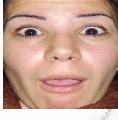
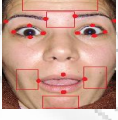
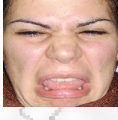
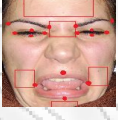
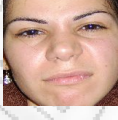
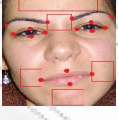
| Questionnaire results | | | | NEU-FACES System Re- sults |
|----------------------------|----------|----------|------------|----------------------------------|
| Emotions | 1st Part | 2nd Part | Mean Value | |
| Neutral | 39,25% | — | 61,74% | 100% |
| Happiness | 68,94% | 96,21% | 82,57% | 90% |
| Sadness | 34,09% | 82,58% | 58,33% | 60% |
| Disgust - Dis- approval | 18,74% | 13,64% | 16,19% | 65% |
| Boredom- Sleepiness | 50,76% | 78,03% | 64,39% | 75% |
| Anger | 76,14% | 69,7% | 72,92% | 55% |
| Surprise | 89,77% | 95,45% | 92,61% | 95% |

classifying the emotions, in certain emotion classes, especially some negative emotions such as ‘anger’ and ‘sadness’, the respective success rate was not quite satisfactory. This led us to the development of more sophisticated classifiers which are described in the following Section.

6.4.8 More Sophisticated Classifiers

In order to increase the success rates of our facial expression recognition system further, we developed more sophisticated algorithms [247, 248] and tested their clas-

Table 6.12 : Sample images of our facial expression database

| Emotions | Input Image | Extracted Features | Network's Response |
|------------------------|---|---|--------------------------|
| Neutral |  |  | [0.97;0.02;0;0;0;0.01;0] |
| Happiness |  |  | [0;0.88;0;0;0;0.12;0] |
| Sadness |  |  | [0;0;0.83;0;0.07;0.1;0] |
| Anger |  |  | [0.12;0;0.11;0.77;0;0;0] |
| Surprise |  |  | [0;0;0;0;0.89;0.11;0] |
| Disgust |  |  | [0;0.14;0;0.1;0;0.76;0] |
| Boredom- Sleepiness |  |  | [0;0;0.15;0.13;0;0;0.72] |

sification success rate. Humans are able to classify facial expressions almost instantly. Specifically, we compared the classification performance of following classifiers in facial expression classification: (1) **R**adial **B**asis **F**unctions (RBF) neural networks, (2)

K-th Nearest Neighbour (KNN) classifiers (3) **S**upport **V**ector **M**achines (SVM) and (4) **M**ultilayer **P**erceptron (MLP) neural networks. The NetLab toolbox was utilized to construct the RBF network, MLP, network and KNN classifiers, while the SVM classifier was implemented with the OSU-SVM toolbox.

In this task, we concluded to seven features which achieved good results. Namely, the extracted features are:

- Mouth Ratio
- Left Eye Ratio
- Right Eye Ratio
- Head size ratio
- Texture of the forehead: Measurement of the changes of the texture of the forehead compared to 'neutral' expression
- Texture of the chin: Measurement of the changes of the texture of the chin compared to 'neutral' expression
- Texture of the region between the eyebrows: Measurement of the changes of the texture of the region between the eyebrows compared to 'neutral' expression

SVM classifier

The support vector machine (SVM) is a supervised classification system that finds an optimal hyperplane which separates data points that will generalize best to future data. Such a hyperplane is the so called maximum margin hyperplane, which

maximizes the distance to the closest points from each class. Let

$$S = \{\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_n\} \quad (6.1)$$

where $\mathbf{s}_j \in \mathbf{R}^d$ be a set of d -dimensional feature vectors corresponding to the image files of a facial expression database. Any hyperplane separating the two data classes has the form Eq. 6.2

$$f(\mathbf{s}) = \mathbf{w} \cdot \Phi(\mathbf{s}) + b, \quad (6.2)$$

where $f : \mathbf{R}^d \rightarrow [-1, +1]$. The SVM classifier is obtained by solving a quadratic programming problem of the form:

$$\min_{w, b, \xi} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^{2n} \xi_i, \quad (6.3)$$

subject to the constraints

$$y_i (\mathbf{w} \cdot \Phi(\mathbf{s}_i) + b) \geq 1 - \xi_i, \xi_i \geq 0 \forall i \in \{1, \dots, 2n\}. \quad (6.4)$$

The optimal solution gives rise to a decision function of the following form:

$$f(\mathbf{s}) = \sum_{i=1}^{2n} y_i w_i \Phi(\mathbf{s}_i) \cdot \Phi(\mathbf{s}_j) + b \quad (6.5)$$

A significant characteristic of SVMs is that only a small fraction of the w_i coefficients are non-zero. The corresponding pairs of \mathbf{s}_i entries (known as margin support vectors) and y_i output labels fully define the decision function. Given that the training patterns appear only in dot product terms $\Phi(\mathbf{s}_i) \cdot \Phi(\mathbf{s}_j)$, we can employ a positive definite kernel function $K(\mathbf{s}_i, \mathbf{s}_j) = \Phi(\mathbf{s}_i) \cdot \Phi(\mathbf{s}_j)$ to implicitly map into a higher dimensional space and compute the dot product. Specifically, in our approach we utilize the Gaussian kernel function which is of the form $K(\mathbf{s}_i, \mathbf{s}_j) = \exp\{-\frac{\|\mathbf{s}_i - \mathbf{s}_j\|^2}{2\sigma^2}\}$.

RBF neural network classifier

For our system, we also considered radial basis function (RBF) networks for facial expression classification. RBF networks have the advantages over other classifiers that they use initially unsupervised learning methods to find clusters of facial expressions without presupposed class labels. Then, the RBF network distinguishes facial expression classes using the weights that are learnt during training when the class labels for the samples are included. Also, the RBF network can quickly classify new facial images of expressions once it has been trained. However, training can require a large amount of time because it traditionally involves finding good parameters for each basis function using gradient descent.

The input layer is determined by the dimensionality d of feature vector of each data point. Thus, the input to our RBF network is a vector $\mathbf{s}_j \in \mathbf{R}^d$.

We will choose M basis functions for our network where each function computes the distance from \mathbf{s}_i to a prototype vector \mathbf{s}_j . We use Gaussians for our basis functions: $K(\mathbf{s}_i, \mathbf{s}_j) = \exp\{-\frac{\|\mathbf{s}_i - \mathbf{s}_j\|^2}{2\sigma_j^2}\}$. The parameters \mathbf{s}_j and σ_j for each function are determined using unsupervised or supervised methods. So, the RBF network consisted of fifty basis functions (50 neurons) in the hidden layer.

The number of neurons in the output layer is determined by the number of classes we want to classify in each experiment. The equation for a single output

$$y_k(\mathbf{s}) = \sum_{i=1}^M w_{ki} \Phi_i(\mathbf{s}) + b,$$

where $b = w_{k0}$ is the weight of the bias

The network was trained with the Expectation Maximization algorithm for two hundred (200) cycles and its output estimates the degree of membership of the input

feature vector in each class. Thus, the value at each output necessarily remains between 0 and 1.

MLP neural network classifier

The Multi-layer Perceptron neural network which was constructed has two feed-forward layers. In this network, the dimensionality of input layer is d , with M hidden units and c output units. The output of the j th hidden unit is given by a weighted linear combination of the d input values:

$$a_j = \sum_{i=1}^d w_{ji}^{(1)} \mathbf{s}_i + b^{(1)}, \quad (6.6)$$

where $w_{ji}^{(1)}$ denotes a weight in the first layer going from input i to hidden unit j and $b^{(1)}$ is the bias for the first layer. Similarly, the outputs for the second layer is given in the following form:

$$a_k = \sum_{i=1}^M w_{kj}^{(2)} \mathbf{z}_i + b^{(2)}. \quad (6.7)$$

The activation of the k th output unit is obtained by transforming the linear combination using a non-linear activation function, to give:

$$y_k(\mathbf{s}) = \tilde{g}(a_k), \quad (6.8)$$

where g is the activation function.

In other words, an explicit expression for the complete function represented by our network is given in the form:

$$y_k(\mathbf{s}) = \tilde{g} \left(\sum_{i=1}^M w_{kj}^{(2)} g \left(\sum_{i=1}^d w_{ji}^{(1)} \mathbf{s}_i + b^{(1)} \right) + b^{(2)} \right) \quad (6.9)$$

The number of hidden units is five (5). The two-layer network with linear outputs

is trained by minimizing a sum-of-squares error function using the scaled conjugate gradient optimizer.

KNN classifier

The KNN classifier was based on the class label prediction of the 10 nearest neighbours.

The NetLab toolbox was utilized in order to construct the RBF network, MLP, network and KNN classifiers, while the SVM classifier was implemented with the OSU-SVM toolbox. More details for the classifiers can be found in [250, 251, 252, 253].

6.4.9 Experimental performance evaluation

Classification results were calculated using 10-fold cross-validation evaluation, where the dataset to be evaluated was iteratively partitioned so that 90% be used for training and 10% be used for testing for each class. This process was iterated with different disjoint partitions and the results were averaged. This ensured that the calculated accuracy was not biased because of the particular partitioning of training and testing.

The results have shown that the SVM classifiers achieved higher results than the other three classifiers. The results presented in Table 6.14 illustrate the SVM classifier as the most appropriate for this task. Also, based on our empirical studies, which were described in Chapter 5, we were able to measure the performance of human observers for the facial expression classification task. As we can observe, all the classifiers perform better than the human classifiers, results are also shown in Table 6.14

In Table 6.15, we observe the classification accuracy for each of the seven classes

depicting the ‘neutral’, ‘happiness’, ‘surprise’, ‘anger’, ‘sadness’, ‘boredom-sleepiness’ and ‘disgust’ expression, respectively. The results of the four classifiers are in agreement with the results from the human responses. As we can observe in Table 6.15, the expressions corresponding to ‘angry’ and ‘disgusted’ achieved the lower success rates not only from the classifiers but also from the humans.

Based on these results we consider the SVM classifier as the most appropriate classifier for this problem. Also, as we observe from the results in Table 6.15, the MLP classifier achieved the best classification rate in classifying the ‘disgust’ expression. In the following Tables 6.16, 6.17, 6.18 and 6.19, we see the confusion matrix for the 250 images for the SVM, RBF, KNN and MLP classifier, respectively.

The results in Tables 6.16, 6.17, 6.18 and 6.19 show that the misclassification for the ‘anger’ and ‘disgust’ expressions are confined to these two expressions. Especially, the MLP classifier, misclassified many images of the ‘anger’ expression as ‘disgust’. The best results for the ‘disgust’ expression are given from the MLP Classifier, but, in the same time, misclassified many of the ‘anger’ face images as ‘disgust’, so it can be trusted in using it. Based on this, we consider the SVM Classifier as most appropriate for this problem because it achieved the best results for the full set of 7 expressions.

6.5 Summary - Conclusions

In this Chapter, we described extensively the face detection and facial expression recognition system that we have developed. Face detection is based on a model proposed by P. Sinha. We preprocess the image in order to depict this model and use an artificial neural network, which classifies the image to as ‘face’ and ‘non face’. Towards

this task, we built two different artificial neural networks and decided upon using the second, which demonstrated better performance. To measure the performance of the second network in detecting faces in images, we tested the network in four different set of images: (1) various images of different sizes and resolutions gathered from the World Wide Web and other sources(e.g scanning old photo images), (2) images from our own facial expression database where people may form some expression, (3) face images acquired in the first efforts to construct a facial expression database (low quality images) and, (4) non human face images (images of pets and animals, complex backgrounds and parts of the face and human). The system managed to detect face with 90,83%, 94,00% and 72,89% success rate, for the three first sets, respectively, whereas for the ‘non-face’ images set the success rate was 100%.

Facial expression recognition can be divided in two sets of attempts. In our first attempts for a facial expression recognition system, we tried to use some of the databases already available over the World Wide Web, as mentioned in Chapter 3, Section 3.1, whereas the emotion classes, that our system would be able to recognize, were not wet been determined. We used a fairly simple feature extraction algorithm which computes specific size ration of some facial portions, such as the eyes, the mouth and the size of the face and, then, feed the computed feature vector to an artificial neural network which classifies the expression. Although the developed system showed some good results and was able to generalize in low quality face images and faces in side view, we soon developed a more sophisticated feature extraction algorithm that we finally adopt. In the newer attempts towards facial expression recognition, we use more facial features are extracted from the image, such as measurements of the texture, head orientation, etc, we use our own facial expression database for training and/or

testing, as it is more complete in terms of the classes we want to classify and, a better, more sophisticated, algorithm to extract the facial features is used, which is based on our eye detection/extraction algorithm. After successful eye detection/extraction, the rest of the features are computed based on their relative location with the eyes. The computed feature vector is, again, fed to an artificial neural network which classifies the emotion. The neural network resulted to an average success rate of 77,14% in classifying the expressions. In order to achieve better results, we developed more sophisticated classifiers, using Netlab Toolbox: (1) **R**adial **B**asis **F**unctions neural networks, (2) **K**-th **N**earest **N**eighbour classifiers (3) **S**upport **V**ector **M**achines and (4) **M**ultilayer **P**erceptron neural networks. We trained and tested the classifiers using 10-fold cross validation techniques. Finally, we concluded to the SVM Classifier as the more adequate for this problem, which achieved an accuracy of 96.97%.

Table 6.13 : Sample images of our facial expression database

| Emotions | Input Image | Extracted Features | Network's Response |
|--------------------|---|--|-----------------------------|
| Neutral |  |  | [0.88;0;0.05;0.07;0;0.01;0] |
| Happiness |  |  | [0;1;0;0;0;0;0] |
| Sadness |  |  | [0;0;0.83;0;0;0.17;0] |
| Anger |  |  | [0.14;0;0.12;0.74;0;0;0] |
| Surprise |  |  | [0.04;0;0;0;0.83;0.13;0] |
| Disgust |  |  | [0.01;0;0;0;0.12;0.74;0.13] |
| Boredom-Sleepiness |  |  | [0;0;0;0;0;0.23;0.77] |

Table 6.14 : Human versus computer classifiers

| Classifiers | Accuracy for the seven classes | Humans |
|-------------|--------------------------------|--------|
| MLP | 88.74% | 64.11% |
| RBF | 95.37% | |
| KNN | 96.11% | |
| SVM | 96.97% | |

Table 6.15 : Classification rates for each expression

| Expressions | MLP | RBF | KNN | SVM | Human responses |
|--------------|--------|--------|--------|--------|-----------------|
| Neutral | 100% | 100% | 100% | 100% | 62% |
| Happy | 100% | 100% | 100% | 100% | 83% |
| Surprised | 99.60% | 100% | 100% | 100% | 93% |
| Sad | 98.80% | 99.20% | 99.60% | 100% | 58% |
| Angry | 33.60% | 90.40% | 98.40% | 94.40% | 73% |
| Bored-Sleepy | 97.20% | 99.20% | 98.80% | 98.40% | 64% |
| Disgusted | 92.00% | 78.80% | 76.00% | 86.00% | 16% |

Table 6.16 : Confusion matrix for the SVM classifier

| | Neutral | Happy | Surprised | Sad | Angry | Bored/ Sleepy | Disgusted |
|---------------|---------|-------|-----------|-----|-------|------------------|-----------|
| Neutral | 250 | 0 | 0 | 0 | 0 | 0 | 0 |
| Happy | 0 | 250 | 0 | 0 | 0 | 0 | 0 |
| Surprised | 0 | 0 | 250 | 0 | 0 | 0 | 0 |
| Sad | 0 | 0 | 0 | 250 | 0 | 0 | 0 |
| Angry | 0 | 0 | 0 | 0 | 236 | 0 | 14 |
| Bored /Sleepy | 0 | 4 | 0 | 0 | 0 | 246 | 0 |
| Disgusted | 0 | 0 | 0 | 0 | 35 | 0 | 215 |

Table 6.17 : Confusion matrix for the RBF classifier

| | Neutral | Happy | Surprised | Sad | Angry | Bored/ Sleepy | Disgusted |
|---------------|---------|-------|-----------|-----|-------|------------------|-----------|
| Neutral | 250 | 0 | 0 | 0 | 0 | 0 | 0 |
| Happy | 0 | 250 | 0 | 0 | 0 | 0 | 0 |
| Surprised | 0 | 0 | 250 | 0 | 0 | 0 | 0 |
| Sad | 0 | 0 | 2 | 248 | 0 | 0 | 0 |
| Angry | 0 | 0 | 0 | 0 | 226 | 0 | 24 |
| Bored /Sleepy | 0 | 2 | 0 | 0 | 0 | 248 | 0 |
| Disgusted | 1 | 0 | 1 | 7 | 44 | 0 | 197 |

Table 6.18 : Confusion matrix for the KNN classifier

| | Neutral | Happy | Surprised | Sad | Angry | Bored/ Sleepy | Disgusted |
|---------------|---------|-------|-----------|-----|-------|------------------|-----------|
| Neutral | 250 | 0 | 0 | 0 | 0 | 0 | 0 |
| Happy | 0 | 250 | 0 | 0 | 0 | 0 | 0 |
| Surprised | 0 | 0 | 250 | 0 | 0 | 0 | 0 |
| Sad | 0 | 0 | 1 | 249 | 0 | 0 | 0 |
| Angry | 0 | 0 | 0 | 2 | 246 | 0 | 2 |
| Bored/ Sleepy | 0 | 3 | 0 | 0 | 0 | 247 | 0 |
| Disgusted | 0 | 0 | 0 | 2 | 58 | 0 | 190 |

Table 6.19 : Confusion matrix for the MLP classifier

| | Neutral | Happy | Surprised | Sad | Angry | Bored/ Sleepy | Disgusted |
|---------------|---------|-------|-----------|-----|-------|------------------|-----------|
| Neutral | 250 | 0 | 0 | 0 | 0 | 0 | 0 |
| Happy | 0 | 250 | 0 | 0 | 0 | 0 | 0 |
| Surprised | 0 | 0 | 249 | 1 | 0 | 0 | 0 |
| Sad | 0 | 0 | 3 | 247 | 0 | 0 | 0 |
| Angry | 0 | 0 | 0 | 5 | 84 | 0 | 161 |
| Bored/ Sleepy | 0 | 7 | 0 | 0 | 0 | 243 | 0 |
| Disgusted | 0 | 0 | 0 | 3 | 17 | 0 | 230 |

Conclusions and Future Work

All truths are easy to understand once they are discovered; the point is to discover them.

—Galileo Galilei (1564–1642)

7.1 Summary and Conclusions

DEVELOPING a fully automated facial expression system is a quite significant and challenging task. Automated face detection and expression classification in images is a prerequisite in the development of novel human-computer interaction modalities. However, the development of integrated, fully operational such detection/classification systems is known to be non-trivial, a fact that was corroborated by our own statistical results regarding expression classification by humans. Towards building such systems, in this theses, we made the following studies:

1. First of all, we studied the emotion perception from the scientific point of view.
-

Biologists and doctors, have identified some parts of the face that are considered important for human emotion expression and understanding. These include: (1) **the frontal cortex**, (2) **the superior temporal sulcus**, and (3) **the amygdala**. This fact is strengthened the opinion that the emotions are affected by the ‘evolution’ and, thus, are similar in all cultures. On the other hand, there is a strong disagreement among psychologists about emotion perception. Some studies, mainly conducted by Paul Ekman[13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29], identified some emotions called ‘**basic emotions**’, to be similar among different cultures. These emotions are, namely: ‘**anger**’, ‘**disgust**’, ‘**fear**’, ‘**happiness**’, ‘**sadness**’ and ‘**surprise**’. Besides the ‘basic emotions’, studies have shown that there are also cultural variations in the way in which humans express emotion. These studies have shown that the emotions can be varied: (1) in terms of the expression of emotion, but also (2) in terms of the intensity of the expressed emotion. This fact has further been demonstrated by our own empirical studies on humans. Also, Russell [254, 66, 67, 68, 69, 70, 71] argues that emotion in general (and facial expression of emotion in particular) can be best characterized in terms of a multidimensional affect space, rather than in terms of discrete emotion categories. More specifically, Russell claims that two dimensions, namely ‘pleasure’ and ‘arousal’, are sufficient to characterize facial affect space.

2. The next study involved understanding emotion perception from the human point of view. Towards this task, we conducted two empirical studies. The first study, as described in Section 5.1, was simpler than the second and aimed

at setting an error goal for our system. We used images from facial expression databases gathered from World Wide Web [109, 112] and asked people to map the emotion based on the subject's expression. Our second empirical study, which is described in Section 5.1, was more complicated and aimed not only at an error goal, but also, at understanding the mechanisms of facial expression recognition by humans. In this study, we used our own facial expression database [3]. The results showed us that the cultural exposure increases the chances of correct recognition of facial expressions indicating cultural dependence in the ways people express themselves. This is demonstrated by the significant difference between the error rates of the first questionnaire, where we used images on non-Greek subjects, and the second questionnaire, where we used images from our own facial expression database. Specifically, for the majority of the expressions the success rates were extremely comparable for the second questionnaire, as they achieved a difference from 13% to 46%, compared to the first questionnaire. Exceptions were observed for the 'neutral' and the 'disgust' emotion. Moreover, we were able to identify the emotions that are present during a typical human-computer interaction, so facial expressions corresponding to the **'neutral', 'happiness', 'sadness', 'surprise', 'anger', 'disgust' and 'boredom-sleepiness'** psychological states arise very commonly during human-computer interaction.

3. We also studied previous attempts towards the development of: (1) a facial expression database, (2) a face detection system and (3) a facial expression recognition system. We set the requirements for an ideal result for each of

the aforementioned three occasions, respectively. Our study concluded to the fact that there are some interesting attempts but there is none that can cover all the requirements. Moreover, the majority of the methods usually address the problem of facial expression classification to the ‘basic emotion’ classes, which they do not include the ‘boredom-sleepiness’ emotion. Finally, as there is a culturally specificity on emotion perception, the development of a facial expression recognition system for Greek people is extremely important.

4. Our study also showed that led us to the assumption that we must create our own facial expression database. This fact, led us to the creation of two different databases, namely:

(1) The database of low quality images: this database consists of many subjects, depicting many expressions, but the image quality is quite low as we used web cameras to acquire the data and,

(2) The database of high quality images: this database consists again of many subjects, depicting the expressions recognized by our system, and the image quality is quite high as we we used digital cameras to acquire the data

5. Finally, we created our own face detection and facial expression recognition system. Face detection is based on a model of the human face proposed by P. Sinha. We preprocess the image in order to depict this model and use an artificial neural network, which classifies the image to ‘face’ and ‘non-face’. Towards this task, we built two different artificial neural networks and decided upon using the second, which demonstrated better results. To measure the performance of the second network in detecting faces in images, we tested the

network in four different set of images: (1) various images of different sizes and resolutions gathered from the World Wide Web and other sources (e.g scanning old photo images), (2) images from our own facial expression database where people may form some expression, (3) face images acquired in the first efforts to construct a facial expression database (low quality images) and, (4) non-human face images (e.g., images of pets and animals, complex backgrounds and parts of the face and human). The system managed to detect face with 90,83%, 94,00% and 72,89% success rate, for the three first sets, respectively, whereas for the 'non-face' images set the success rate was 100%.

6. Facial expression recognition can be divided to two sets of attempts. In our first attempts for a facial expression recognition system, we tried to use some of the databases already available over the World Wide Web, as mentioned in Section 3.1, whereas the emotion classes, that our system would be able to recognize, were not yet been determined. We used a fairly simple feature extraction algorithm which computes specific size ration of some facial portions, such as the eyes, the mouth and the size of the face and, then, feed the computed feature vector to an artificial neural network which classifies the expression. Although the developed system showed some good results and was able to generalize in low quality face images and faces in side view, we soon developed a more sophisticated feature extraction algorithm that we finally adopt. In the newer attempts towards facial expression recognition, we use more facial features are extracted from the image, such as measurements of the texture, head orientation, etc, we use our own facial expression database for training and/or testing, as it is more

complete in terms of the classes we want to classify and, a better, more sophisticated, algorithm to extract the facial features is used, which is based on our eye detection/extraction algorithm. After successful eye detection/extraction, the rest of the features are computed based on their relative location with the eyes. The computed feature vector is, again, fed to an artificial neural network which classifies the emotion. The neural network resulted to an average success rate of 77,14% in classifying the expressions. In order to achieve better results, we developed more sophisticated classifiers, using Netlab Toolbox: (1) **R**adial **B**asis **F**unctions neural networks, (2) **K**-th **N**earest **N**eighbour classifiers (3) **S**upport **V**ector **M**achines and (4) **M**ultilayer **P**erceptron neural networks. We trained and tested the classifiers using 10-fold cross validation techniques. Finally, we concluded to the SVM Classifier as the more adequate for this problem, which achieved an accuracy of 96.97%.

7.2 Current and Future Work

7.2.1 Towards a multimodal emotion recognition system

Recently, the recognition of emotions of users while they interact with software applications has been acknowledged as an important research topic. How people feel may play an important role on their cognitive processes as well [72]. Thus the whole issue of human-computer interaction has to take into account users' feelings. Picard [73] points out that one of the major challenges in affective computers is to try to improve the accuracy of recognizing people's emotions. Improving the accuracy on emotion recognition may imply the combination of many modalities in user interfaces. Indeed,

human emotions are usually expressed in many ways. For example, as we articulate speech we usually move the head and exhibit various facial emotions [255]. Ideally evidence from many modes of interaction should be combined by a computer system so that it can generate as valid hypotheses as possible about users' emotions. This view has been supported by many researchers in the field of human-computer interaction [256, 257, 73]. However, progress in emotion recognition based on multiple modalities has been quite slow. Although several approaches have been proposed to recognize human emotions based on facial expressions or speech, relatively limited work has been done to fuse these two and other modalities to improve the accuracy and robustness of the emotion recognition system [258].

In view of the above, it is our aim to improve the accuracy of visual-facial emotion recognition by combining other modalities, namely keyboard stroke pattern information and audio-lingual information. Currently, a system that combines two modalities, namely the keyboard and the voice, has been already constructed and is described briefly in [259]. As we described in Chapter 5, towards building a facial expression recognition system, we conducted a fairly intensive empirical study. Towards combining the three modalities, we had to determine the extent to which these three different modalities can provide emotion recognition from the perspective of a human observer. Moreover, we had to specify the strengths and weaknesses of each modality.

In this way, we could determine the weights of the criteria that correspond to the respective modalities from the perspective of a human observer. Hence, for the purposes of our research we conducted empirical studies concerning emotion recognition based on two modalities: the audio-lingual, keyboard stroke patterns and the visual-facial. The above empirical studies constitute an important milestone for our

research and yield important results. Not only do they provide the basis towards the combination of modalities into the affective user modeling component of our tri-modal system, but they also give evidence for other researchers to use since, currently, there are not many results from such empirical studies in the literature. Indeed, after an extensive search of the literature, we found that there is a shortage of empirical evidence concerning the strengths and weaknesses of these modalities. The most relevant research work is that of De Silva et al. [79] who performed an empirical study and reported results on human subjects' ability to recognize emotions. However, De Silva et al. focus on the audio signals of voice concentrating on the pitch and volume of voice rather than lingual keywords that convey affective information. On the other hand, in our research we have included the lingual aspect of users' spoken words on top of the pitch and volume of voice and have compared the audio-lingual results with the results from the other two modes so that we can see which modality conveys more information for human observers. Our work has been conducted for six emotions, namely 'happiness', 'sadness', 'surprise', 'anger' and 'disgust' as well as the emotionless state which we refer to as 'neutral'. The multimodal system is shown in Figure 7.1

Currently, we have studied the possibility of combining the three modalities, based on the following empirical studies:

- Audio-lingual information and Visual-facial information [260, 261]
- Keyboard stroke pattern information and Visual-facial information [262]
- Audio-lingual information, Keyboard stroke pattern information and Visual-facial information [263, 264]

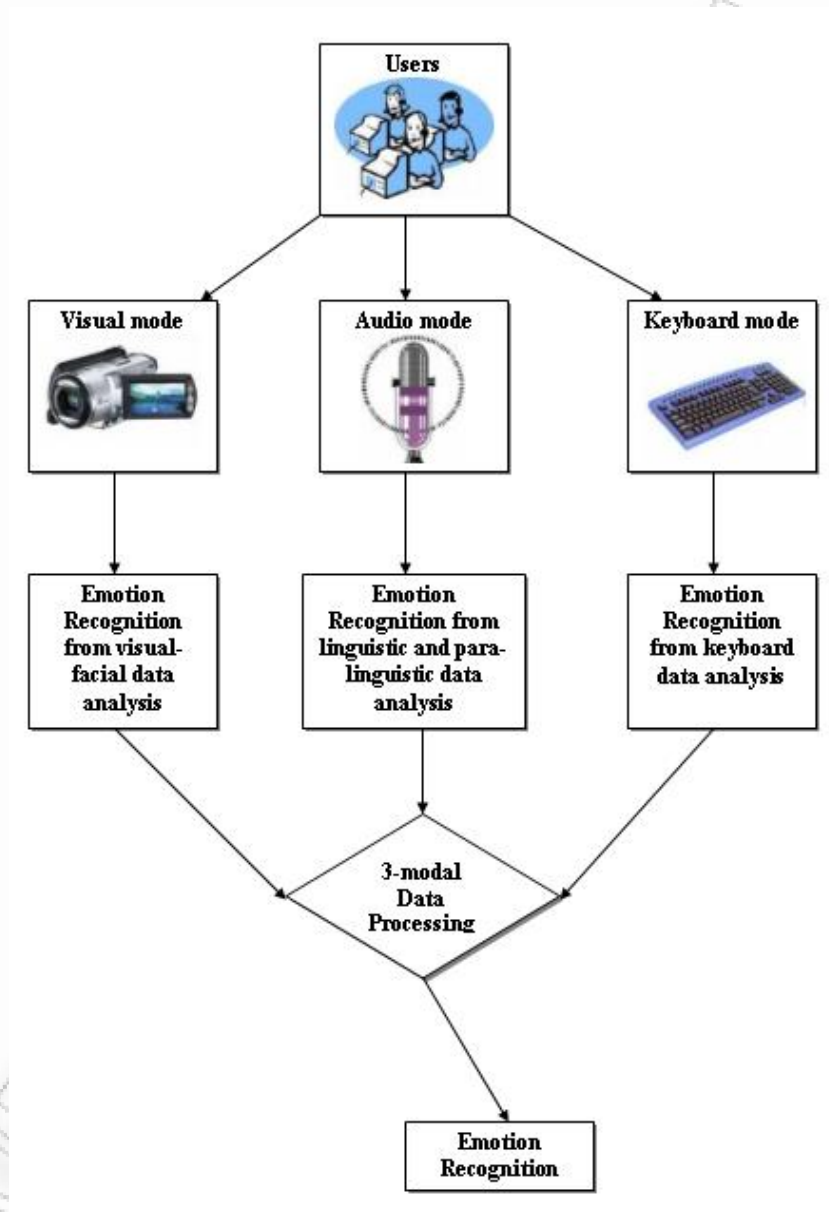


Figure 7.1 : The architecture of our multimodal emotion recognition system

Some of the results from combining the three modalities, based on their empirical studies, are shown in Figure 7.2 [264].







| <i>Visual-facial modality</i> | | <i>Keyboard-stroke pattern and audio-lingual information</i> | | |
|---|--------|--|-----------------------|------------|
| Facial Expression | (%) | (%) for keyboard-stroke patterns | (%) for audio-lingual | Mean value |
| <i>Neutral</i> | | | | |
|  | 61,74% | 65% | 18% | 41,50% |
| <i>Surprise</i> | | | | |
|  | 92,61% | 5% | 62% | 33,50% |
| <i>Anger</i> | | | | |
|  | 72,92% | 74% | 79% | 76,50% |
| <i>Happiness</i> | | | | |
|  | 82,57% | 60% | 46% | 53% |
| <i>Sadness</i> | | | | |
|  | 58,33% | 57% | 48% | 52,50% |
| <i>Disgust</i> | | | | |
|  | 16,19% | 4% | 57% | 30,50% |

Figure 7.2 : Some results from combining the three modalities

Finally, we are currently developing the combination of these modalities in terms of the unimodal systems results. But, all these are beyond the scopes of this thesei

and will be presented in future works.

7.2.2 Towards extending the visual facial expression recognition

In the future, we will extend this work in the following three directions: (1) We will improve our system by using wider training sets so as to cover a wider range of poses and cases of low quality of images. (2) We will investigate the need for classifying into more than the currently available facial expressions, so as to obtain more accurate estimates of a computer user's psychological state. In turn, this may require the extraction and tracing of additional facial points and corresponding features. (3) We plan to apply our system for the expansion of human-computer interaction techniques, such as those that arise in mobile telephony, in which the quality of the input images is too low for existing systems to operate reliably. Finally, we will also investigate the possibility of using other type of image or input data in order to classify the emotion, such as stereoscopic images, thermal images and video sequences.

Bibliography

- [1] M. Pantic and L. J. M. Rothkrantz, “Automatic analysis of facial expressions: the state of the art,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 1424–1445, 2000.
 - [2] P. Barkhuysen, E. Kraemer, and M. Swerts, “Problem detection in human-machine interactions based on facial expressions of users,” *Speech Communication*, vol. 45, pp. 343–359, 2005.
 - [3] I.-O. Stathopoulou and G. A. Tsihrintzis, “Facial Expression Classification: Specifying Requirements for an Automated System,” in *Proceedings of the 10th International Conference on Knowledge-Based Intelligent Information Engineering Systems, LNAI: Vol. 4252*. Berlin, Heidelberg: Springer-Verlag, October 2006, pp. 1128–1135.
 - [4] —, “Towards automated inferencing of Emotional State from face Images,” in *2nd International Conference on Software and Data Technologies*, Barcelona, Spain, July, 5-8 2007.
 - [5] —, *Automated Processing and Classification of Face Images for Human-Computer Interaction Applications*, ser. Studies in Computational Intelligence. Springer Berlin / Heidelberg, 2008, vol. 104, ch. Intelligent Interactive Systems in Knowledge-Based Environments, pp. 107–136.
 - [6] —, “An empirical study of facial expression classification by human observers,” in preparation.
-

- [7] A. P. Association, *Diagnostic and Statistical Manual of Mental Disorders*, 4th ed. American Psychiatric Association, 1984.
- [8] S. Iverson, I. Kupfermann, and E. R. Kandel, *Principles of Neuroscience*. New York: McGraw-Hill, 2000, ch. Emotional states and feelings, pp. 1209–1226.
- [9] M. B. Arnold, *The Nature of Emotion*. Baltimore: Penguin, 1968.
- [10] W. B. Cannon, “The James-Lange theory of emotion: a critical examination and an alternative theory,” *American Journal of Psychology*, vol. 39, pp. 106–124, 1927.
- [11] J. Panksepp, *Affective Neuroscience: The Foundations of Human and Animal Emotions*. New York: Oxford University Press, 1998.
- [12] C. Darwin, “The expression of the emotions in man and animal,” *London: J. Murray. de Haan, M., Humphreys, K. Johnson, M.H*, vol. 1872, pp. 200–212.
- [13] P. Ekman, *Cross-cultural studies of facial expression*. London: Academic Press Inc., 1973, ch. Darwin and Facial Expression, pp. 169–222.
- [14] —, *In J. Cole (Ed.), Nebraska Symposium on Motivation*. Lincoln: University of Nebraska Press, 1972, vol. 19, ch. Universals and cultural differences in facial expressions of emotion.
- [15] —, “Universal Facial Expressions in Emotion,” *Studia Psychologica*, vol. 15, no. 2, pp. 140–147, 1973.
- [16] P. Ekman and W. Friesen, *Unmasking the face: A Guide to Recognizing Emotions from Facial Expressions*. Englewood Cliffs, NJ: Prentice Hall, 1975.

- [17] —, *Manual for the Facial Action Coding System*. Palo Alto: Consulting Psychologists Press, 1977.
- [18] —, “The facial action coding system: A technique for the measurement of facial movement,” *San Diego: Consulting Psychology Press. Ekman, P., Oster, H*, vol. 30, pp. 527–54, 1978.
- [19] P. Ekman and H. Oster, “Facial Expressions of Emotion,” *Annual Review of Psychology*, vol. 30, pp. 527–554, 1979.
- [20] P. Ekman and W. Friesen, “A new pan-cultural facial expression of emotion,” *Motivation and Emotion*, vol. 10, no. 2, pp. 159–268, 1986.
- [21] P. Ekman, W. Friesen, and S. Ancoli, “Facial Signs of Emotional Experience,” *Journal of Personality and Social Psychology*, vol. 39, no. 6, pp. 1125–1134, 1980.
- [22] P. Ekman, W. Friesen, M. Osullivan, and K. Scerer, “Relative Importance of Face, Body, and Speech in Judgements of Personality and Affect,” *Journal of Personality and Social Psychology*, vol. 38, no. 2, pp. 270–277, 1980.
- [23] P. Ekman, R. J. Davidson, and W. V. Friesen, “The duchenne smile: emotional expression and brain physiology ii. journal of personality and,” *Social Psychology*, vol. 58, pp. 342–353, 1990.
- [24] P. Ekman, “Are - there basic emotions?” *Psychological Review*, vol. 99, no. 3, pp. 550–553, July 1992.

- [25] —, “Facial Expression of Emotion - New Findings, New Questions,” *Psychological Science*, vol. 3, no. 1, pp. 34–38, January 1992.
- [26] —, “Facial Expression and Emotion,” *American Psychologist*, vol. 48, no. 4, pp. 384–392, April 1993.
- [27] E. Rosenberg and P. Ekman, “Conceptual and methodological issues in the judgment of facial expressions of emotion,” *Motivation and Emotion*, vol. 19, no. 2, pp. 111–138, June 1995.
- [28] P. Ekman, *Basic Emotions*. Dalglish, T. and Power, M., 1999.
- [29] D. Matsumoto and P. Ekman, “The relationship among expressions, labels, and descriptions of contempt,” *Journal of Personality and Social Psychology*, vol. 87, no. 4, pp. 529–540, October 2004.
- [30] R. C. Solomon, “Back to basics: On the very idea of “basic emotions”,” *Journal for the Theory of Social Behaviour*, vol. 32, pp. 115–158.
- [31] G. B. Duchenne de Boulogne, *The Mechanism of Human Facial Expression*, ser. Studies in Emotion and Social Interaction. Cambridge University Press, July 1990.
- [32] A. Ortony and T. Turner, “What’s basic about basic emotions?” *Psychological Review*, vol. 97, no. 3, pp. 315–331, 1990.
- [33] J. Russell, “Negative results on a reported facial expression of contempt,” *Motivation and Emotion*, vol. 13, no. 3, pp. 281–291, 1991.

- [34] J. Haidt and D. Keltner, "Culture and facial expression: Open-ended methods find more expressions and a gradient of recognition," *Cognition and Emotion*, vol. 13, no. 3, pp. 225–266, 1999.
- [35] J. E. LeDoux, *The Emotional Brain*. New York: Simon and Schuster, 1998.
- [36] D. Matsumoto and P. Ekman, "American-japanese cultural differences in judgments of emotional expressions of different intensities," *Motivation and Emotion*, vol. 13, no. 2, pp. 143–157, January 2005.
- [37] C. Izard, "Innate and universal facial expressions: evidence from developmental and cross-cultural research," *Psychological Bulletin*, vol. 115, pp. 288–299, 1994.
- [38] H. Elfenbein and N. Ambady, "On the universality and cultural specificity of emotion recognition: a meta-analysis," *Psychological Bulletin*, vol. 128, no. 2, pp. 205–235, 2002.
- [39] P. E. Griffiths, *What Emotions Really Are: The Problem of Psychological Categories*, ser. Science and Its Conceptual Foundations. University Of Chicago Press, August 1998.
- [40] K. Nakamura, R. Kawashima, K. Ito, M. Sugiura, T. Kata, A. Nakamura, K. Hatano, S. Nagumo, K. Kubota, H. Fukuda, and S. Kojima, "Activation of the right inferior frontal cortex during assessment of facial emotion," *Journal of Neurophysiology*, vol. 82, pp. 1610–1614, 1999.
- [41] C. J. Harmer, K. V. Thilo, J. C. Rothwell, and G. M. Goodwin, "Transcranial magnetic stimulation of medial-frontal cortex impairs the processing of angry

- facial expressions,” *Nature Neuroscience*, vol. 4, pp. 17–18, 2001.
- [42] V. Gallese, C. Keysers, and G. Rizzolatti, “A unifying view of the basis of social cognition,” *Trends in Cognitive Sciences*, vol. 8, pp. 396–403, 2004.
- [43] M. Jabbi, J. Bastiaansen, and C. Keysers, “A common anterior insula representation of disgust observation, experience and imagination shows divergent functional connectivity pathways,” *PLoS ONE*, vol. 3, no. 8, pp. e29–39, August 2008.
- [44] J. S. Morris, A. Ohman, and R. J. Dolan, “Conscious and unconscious emotional learning in the human amygdala,” *Nature*, vol. 393, pp. 467–480, 1998.
- [45] J. S. Morris, B. Degelder, L. Weiskrantz, and R. J. Dolan, “Differential extrageniculostriate and amygdala responses to presentation of emotional faces in a cortically blind field,” *Brain*, vol. 124, pp. 1241–1252, 2001.
- [46] P. J. Whalen, S. L. Rauch, N. L. Etcoff, S. C. Mcinerney, M. B. Lee, and M. A. Jenike, “Masked presentations of emotional facial expressions modulate amygdala activity without explicit knowledge,” *The Journal of Neurosciences*, vol. 18, pp. 411–418, 1998.
- [47] P. Vuilleumier, J. L. Armony, J. Driver, and R. J. Dolan, “Distinct spatial frequency sensitivities for processing faces and emotional expressions,” *Nature Neuroscience*, vol. 6, pp. 624–631, 2003.
- [48] R. Saxe and N. Kanwisher, “People thinking about thinking people,” *NeuroImage*, vol. 19, pp. 1835–1842, 2003.

- [49] G. Hein and R. T. Knight, “Superior temporal sulcus—it’s my area: Or is it?” *J. Cognitive Neuroscience*, vol. 20, no. 12, pp. 2125–2136, 2008.
- [50] K. S. Labar, M. J. Crupain, J. T. Voyvodic, and G. McCarthy, “Dynamic perception of facial affect and identity in the human brain,” *Cerebral Cortex*, vol. 13, pp. 1023–1033, 2003.
- [51] P. Graf and D. Schacter, “Implicit and explicit memory for new associations in normal and amnesic subjects,” *Journal of Experimental Psychology: Learning, Memory and Cognition*, vol. 11, pp. 501–518, 1985.
- [52] L. Feldman Barrett, P. Niedenthal, M., and P. Winkielman, *Emotion and Consciousness*, 1st ed., L. Feldman Barrett, P. Niedenthal, M., and P. Winkielman, Eds. The Guilford Press, 2005.
- [53] M. L. Gorno-Tempini, S. Pradelli, M. Serafini, G. Pagnoni, P. Baraldi, C. Porro, R. Nicoletti, C. Umita, and P. Nichelli, “Explicit and incidental facial expression processing: an fmri study,” *NeuroImage*, vol. 14, pp. 465–473, 2001.
- [54] J. S. Winston, J. O’Doherty, and R. J. Dolan, “Common and distinct neural responses during direct and incidental processing of multiple facial emotions,” *NeuroImage*, vol. 20, pp. 84–97, 2003.
- [55] J. Scheuerecker, T. Frodl, N. Koutsouleris, T. Zetzsche, M. Wiesmann, A. Kleeemann, H. Bruckmann, G. Schmitt, H.-J. Mfler, and E. Meisenzahl, “Cerebral differences in explicit and implicit emotional processing – an fmri study,” *Neuropsychobiology*, vol. 56, pp. 32–39, 2007.

- [56] V. Gallese, C. Keysers, and G. Rizzolatti, "A unifying view of the basis of social cognition," *Trends in Cognitive Sciences*, vol. 8, pp. 396–403, 2004.
- [57] H. Wang, H. Prendinger, and T. Igarashi, "Communicating emotions in online chat using physiological sensors and animated text," in *CHI '04: CHI '04 extended abstracts on Human factors in computing systems*. New York, NY, USA: ACM, 2004, pp. 1171–1174. [Online]. Available: <http://dx.doi.org/10.1145/985921.986016>
- [58] C. Bartneck, "Emmu • an embodied emotional character for the ambient intelligent home," Ph.D. dissertation, Eindhoven University of Technology, Eindhoven, 2002.
- [59] D. McNair, M. Lorr, and L. Droppleman, *Manual of the Profile of Mood States*. San Diego, CA: Educational and Industrial Testing Services, 1981.
- [60] P. Ekman and W. Friesen, *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Consulting Psychologists Press, Palo Alto, 1978.
- [61] J. A. Russell, "A circumplex model of affect. journal of personality and," *Social Psychology*, vol. 39, pp. 1161–1178, 1980.
- [62] J. A. Russell and M. Bullock, "Multidimensional scaling of emotional facial expressions: similarity from preschoolers to adults. journal of personality and," *Social Psychology*, vol. 48, pp. 1290–1298, 1985.

- [63] J. A. Russell, "Is there universal recognition of emotion from facial expression?" *Psychological Bulletin*, vol. 115, pp. 102–141, 1994.
- [64] J. A. Russell and J. M. Fernandez-Dols, *The Psychology Of Facial Expression*. Cambridge University Press, 1997.
- [65] J. A. Russell, "Core affect and the psychological construction of emotion," *Psychological Review*, vol. 110, pp. 145–172, 2003.
- [66] J. Russell, "Is there Universal Recognition of Emotion from Facial Expression - A Review of the Cross-Cultural Studies," *Psychological Bulletin*, vol. 115, no. 1, pp. 102–141, January 1994.
- [67] J. Russel, "Facial Expressions of Emotion - What Lies Beyond Minimal Universability," *Psychological Bulletin*, vol. 118, no. 3, pp. 379–391, November 1995.
- [68] J. Carroll and J. Russell, "Do facial expressions signal specific emotions? Judging emotion from the face in context," *Journal of Personality and Social Psychology*, vol. 70, no. 2, pp. 205–218, February 1996.
- [69] J. Russell and L. Barrett, "Core affect, prototypical emotional episodes, and other things called Emotion: Dissecting the elephant," *JJournal of Personality and Social Psychology*, vol. 76, no. 5, pp. 805–819, May 1999.
- [70] J. Russell, "Core affect and the psychological construction of emotion," *Psychological Review*, vol. 110, no. 1, pp. 145–172, January 2003.

- [71] J. Russell, J. Bachorowski, and J. Fernandez-Dols, "Facial and vocal expressions of emotion," *Annual Review of Psychology*, vol. 54, pp. 329–349, 2003.
- [72] D. Goleman, *Emotional Intelligence*. New York, USA: Bantam Books.
- [73] R. Picard, "Affective computing: challenges," *International Journal of Human-Computer Studies*, vol. 59, no. 1-2, pp. 55–64, July 2003.
- [74] D. Goren and H. Wilson, "Quantifying facial expression recognition across viewing conditions," *Vision Research*, vol. 46, no. 8-9, pp. 1253–1262, April 2006.
- [75] C. Nass and B. Reeves, "Combining, Distinguishing, and Generating Theories in Communication - A domains of Analysis Framework," *Communication Research*, vol. 18, no. 2, pp. 240–261, April 1991.
- [76] B. Reeves and C. Nass, *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. Cambridge University Press and CSLI, New York., 1998.
- [77] T. Ganel, Y. Goshen-Gottstein, and M. Goodale, "Interactions between the processing of gaze direction and facial expression," *Vision Research*, vol. 45, pp. 1191–1200, 2005.
- [78] R. Picard, "Affective computing for future agents," *Cooperative Information Agents IV*, vol. 1860, p. 14, 2000.
- [79] L. De Silva, T. Miyasato, and R. Nakatsu, "Facial Emotion Recognition Using Multimodal Information," in *Proceedings of IEEE Int. Conf. on Information*,

- Communications and Signal Processing - ICICS*, Singapore, Thailand, September 1997.
- [80] M. Pantic and L. Rothkrantz, "Toward an affect-sensitive multimodal HCI," *Proceedings of the IEEE*, vol. 91, no. 9, pp. 1370–1390, 2003.
- [81] K. L. Schmidt and J. F. Cohn, "Human facial expressions as adaptations: Evolutionary questions in facial expression research," *American Journal of Physical Anthropology*, 533, pp. 3–24, 2001.
- [82] T. Valentine, "Face-space models of face recognition in wenger, m.j. townsend, j.t. (eds.) computational, geometric, and process perspectives on facial cognition: Contexts and challenges," *London: Lawrence Erlbaum Associates Inc. Valentine, T., Bruce, V*, vol. 15, pp. 525–535, 2001.
- [83] S. Baron-Cohen, S. Wheelwright, and T. Jolliffe, "Is there a "language of the eyes"? evidence from normal adults, and adults with autism or asperger syndrome," *Visual Cognition*, vol. 4, pp. 311–331, 1997.
- [84] R. Hassin and Y. Trope, "Facing faces: Studies on the cognitive aspects of physiognomy," *Journal of Personality and Social Psychology*, vol. 78, pp. 837–852, 2000.
- [85] J. E. Scheib, S. W. Gangestad, and R. Thornhill, "Facial attractiveness, symmetry and cues of good genes," in *Proceedings of the Royal Society of London B*, vol. 266, 1999, pp. 1913–1927.

- [86] S. V. Paunonen, K. Ewan, J. Earthy, S. Lefave, and H. Goldberg, "Facial features as personality cues," *Journal of Personality*, vol. 67, pp. 555–583, 1999.
- [87] K. Dion, E. Berscheid, and E. Waltser, "What is beautiful is good. journal of personality and," *Social Psychology*, vol. 24, pp. 207–213, 1972.
- [88] G. R. Adams and T. L. Huston, "Social perception of middle-aged persons varying in physical attractiveness," *Developmental Psychology*, vol. 11, pp. 657–658, 1975.
- [89] T. Valentine, S. Darling, and M. Donnelly, "Why are average faces attractive? the effect of view and averageness on the attractiveness of female faces," *Psychonomic Bulletin Review*, vol. 11, pp. 482–487, 2004.
- [90] G. Rhodes, L. A. Zebrowitz, A. Clark, S. M. Kalick, A. Hightower, and M. R., "Do facial averageness and symmetry signal health," *Evolution and Human Behavior*, vol. 22, pp. 31–46, 2001.
- [91] I. S. Penton-Voak and D. I. Perrett, "Female preference for male faces changes cyclically - further evidence," *Evolution and Human Behavior*, vol. 21, pp. 39–48, 2000.
- [92] U. Hess, R. B. Adams, and R. E. Kleck, "Facial appearance, gender and emotion expression," *Emotion*, vol. 4, pp. 378–388, 2004.
- [93] A. Mignault and A. Chaudhuri, "The many faces of a neutral face: head tilt and perception of dominance and emotion," *Journal of Nonverbal Behavior*, vol. 27, pp. 111–132, 2003.

- [94] M. J. Lyons, R. P. Campbell, A. Coleman, M. Kamachi, and S. Akamatsu, “The noh mask effect: vertical viewpoint dependence of facial expression perception. proceedings of the royal society of,” *London B*, vol. 267, pp. 2239–2245, 2000.
- [95] S. Harnad, “Categorical perception,” *New York: Cambridge University Press. Hassin, R. Trope, Y*, vol. 78, pp. 837–852, 1987.
- [96] M. H. Bornstein, *Categorical Perception*. New York: Cambridge University Press, 1987, ch. Perceptual categories in vision and audition.
- [97] C. L. Krumhansl, “Music psychology: Tonal structures in perception and memory,” *Annual Review of Psychology*, vol. 42, pp. 277–303, 1991.
- [98] A. W. Young, D. Rowland, A. J. Calder, N. L. Etcoff, A. Seth, and D. I. Perrett, “Facial expression megamix: Tests of dimensional and category accounts of emotion recognition,” *Cognition*, vol. 63, pp. 271–313, 1997.
- [99] D. Bimler and J. Kirkland, “Categorical perception of facial expressions of emotion: Evidence from multidimensional scaling,” *Cognition and Emotion*, vol. 15, pp. 633–658, 2001.
- [100] R. Shah and M. B. Lewis, “Locating the neutral expression in the facial-emotion space,” *Visual Cognition*, vol. 10, pp. 549–566, 2003.
- [101] M. Csikszentmihalyi, *Flow: The Psychology of Optimal Experience*. New York, USA: Harper Perennial.
- [102] J. Klein, R. Picard, and J. Riseberg, “Support for Human Emotional Needs in Human-Computer Interaction,” in *Proceedings of CHI’97 Workshop on Human*

- Needs and Social Responsibility*, 1997.
- [103] J. Mayer and P. Salovey, "The Intelligence Of Emotional Intelligence," *Intelligence*, vol. 17, no. 4, pp. 433–442, October-December 1993.
- [104] J. Mayer, M. Dipaolo, and P. Salovey, "Perceiving Affective Content in Ambitious Visual-Stimuli - A Component Of Emotional Intelligence," *Journal Of Personality Assessment*, vol. 54, no. 3-4, pp. 772–781, 1990.
- [105] J. Mayer, P. Salovey, and D. Caruso, "Emotional intelligence: Theory, findings, and implications," *Psychological Inquiry*, vol. 15, no. 3, pp. 197–215, 2004.
- [106] H. Marsh and R. Shavelson, "Self-concept: Its multifaceted, hierarchical structure," *Educational Psychologist*, vol. 20, no. FAL, pp. 107–204, 1985.
- [107] A. Ortony, G. L. Clore, and A. Collins, *The Cognitive Structure of Emotions*. New York, USA: Cambridge University Press.
- [108] M. Rosenberg, *Conceiving the Self*. New York, USA: Basic Books Inc.
- [109] A. Martinez and R. Benavente, "The AR face database," University of Wisconsin - Madison Computer Sciences Department, Tech. Rep. CVC Technical Report Num.24, June 1998.
- [110] M. J. Lyons, J. Budynek, and S. Akamatsu, "Automatic classification of single facial images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 12, pp. 1357–1362, 1999.

- [111] A. Georghiades, P. Belhumeur, and D. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. Pattern Anal. Mach. Intelligence*, vol. 23, no. 6, pp. 643–660, 2001.
- [112] T. Kanade, Y. Tian, and J. F. Cohn, "Comprehensive database for facial expression analysis," in *FG '00: Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition 2000*. Washington, DC, USA: IEEE Computer Society, 2000, p. 46.
- [113] M. Pantic, M. Valstar, R. Rademaker, and L. Maat, "Web-based database for facial expression analysis," in *IEEE Int'l Conf. on Multimedia and Expo 2005*, June 2005, pp. 317–321. [Online]. Available: <http://pubs.doc.ic.ac.uk/Pantic-ICME05-2/>
- [114] A. Colmenarez and T. Huang, "Face detection with information-based maximum discrimination," in *CVPR '97: Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition (CVPR '97)*. Washington, DC, USA: IEEE Computer Society, 1997, p. 782.
- [115] G. Yang and T. Huang, "Human face detection in a complex background," *Pattern Recognition*, vol. 27, no. 1, pp. 53–63, 1994.
- [116] S. Lee, Y. Ham, and R. Park, "Recognition of human front faces using knowledge-based feature extraction and neuro-fuzzy algorithm," *Pattern Recognition*, vol. 29, no. 11, pp. 1863–1876, 1996.
- [117] T. K. Leung, M. C. Burl, and P. Perona, "Finding faces in cluttered scenes using random labeled graph matching," in *ICCV '95: Proceedings of the Fifth*

- International Conference on Computer Vision*. Washington, DC, USA: IEEE Computer Society, 1995, p. 637.
- [118] H. A. Rowley, S. Baluja, and T. Kanade, "Rotation invariant neural network-based face detection," in *CVPR '98: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Washington, DC, USA: IEEE Computer Society, 1998, p. 38.
- [119] —, "Neural network-based face detection," *IEEE Transactions On Pattern Analysis and Machine intelligence*, vol. 20, pp. 23–38, 1998.
- [120] M.-H. Yang and N. Ahuja, *Face Detection and Gesture Recognition for Human-Computer Interaction*. Norwell, MA, USA: Kluwer Academic Publishers, 2001.
- [121] P. Juell and R. Marsh, "A hierarchical neural network for human face detection," *Pattern Recognition*, vol. 29, no. 5, pp. 781–787, 1996.
- [122] C. Lin and K. Fan, "Triangle-based approach to the detection of human face," *Pattern Recognition*, vol. 34, no. 5, pp. 1271–1284, 2001.
- [123] K.-K. Sung and T. Poggio, "Example-based learning for view-based human face detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 1, pp. 39–51, 1998.
- [124] L.-L. Huang and A. Shimizu, "A multi-expert approach for robust face detection," *Pattern Recogn.*, vol. 39, no. 9, pp. 1695–1703, 2006.
- [125] M. Castrillón, O. Déniz, C. Guerra, and M. Hernández, "ENCARA2: Real-time detection of multiple faces at different resolutions in video streams," *J.*

- Vis. Comun. Image Represent.*, vol. 18, no. 2, pp. 130–140, 2007.
- [126] S. Phimoltares, C. Lursinsap, and K. Chamnongthai, “Face detection and facial feature localization without considering the appearance of image context,” *Image Vision Comput.*, vol. 25, no. 5, pp. 741–753, 2007.
- [127] S. Kadoury and M. D. Levine, “Face detection in gray scale images using locally linear embeddings,” *Comput. Vis. Image Underst.*, vol. 105, no. 1, pp. 1–20, 2007.
- [128] C. Kotropoulos and I. Pitas, “Rule-based face detection in frontal views,” in *ICASSP '97: Proceedings of the 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '97) - Volume 4*. Washington, DC, USA: IEEE Computer Society, 1997, p. 2537.
- [129] C. P. Papageorgiou, M. Oren, and T. Poggio, “A general framework for object detection,” in *Computer Vision, 1998. Sixth International Conference on*, 1998, pp. 555–562. [Online]. Available: <http://dx.doi.org/10.1109/ICCV.1998.710772>
- [130] P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” 2001, pp. 511–518.
- [131] M. Jones and P. Viola, “Fast multi-view face detection,” Mitsubishi Electric Research Laboratories, Tech. Rep., 2003.
- [132] C. Liu, “A bayesian discriminating features method for face detection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, pp. 725–740, 2003.

- [133] H. Schneiderman and T. Kanade, "A statistical approach to 3d object detection applied to faces and cars," 2000, pp. 0–6.
- [134] S. Z. Li and Z. Zhang, "Floatboost learning and statistical face detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, pp. 1–12, 2004.
- [135] C. Huang, H. Ai, Y. Li, and S. Lao, "Vector boosting for rotation invariant multi-view face detection," in *ICCV '05: Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*. Washington, DC, USA: IEEE Computer Society, 2005, pp. 446–453.
- [136] B. Wu, H. Ai, C. Huang, and S. Lao, "Fast rotation invariant multi-view face detection based on real adaboost," in *Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on*, 2004, pp. 79–84. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1301512
- [137] M.-H. Yang, D. 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.
- [138] E. Hjelmås and B. Lowu, "Face detection: a survey," *Comput. Vision Image Understand.*, vol. 83, no. 3, pp. 236–274, 2001.
- [139] Y. Dai and Y. Nakano, "Face-texture model based on sgld and its application in face detection in a color scene," *Pattern Recognition*, vol. 29, pp. 1007–1017, 1996.

- [140] A. Tankus, "Face detection by direct convexity estimation," *Pattern Recognition Letters*, vol. 18, pp. 913–922, 1997.
- [141] 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*, pp. 669–680, 1998.
- [142] S.-H. Jeng, H. Y. Liao, C. C. Han, M. Y. Chern, and Y. T. Liu, "Facial feature detection using geometrical face model: An efficient approach," *Pattern Recognition*, vol. 31, no. 3, pp. 273–282, March 1998. [Online]. Available: [http://dx.doi.org/10.1016/S0031-3203\(97\)00048-4](http://dx.doi.org/10.1016/S0031-3203(97)00048-4)
- [143] J.-G. Wang and E. Sung, "Frontal-view face detection and facial feature extraction using color and morphological operations," *Pattern Recognition Letters*, vol. 20, pp. 1053–1068, 1999.
- [144] G. Wei and I. K. Sethi, "Face detection for image annotation," *Pattern Recognition Letters*, vol. 20, pp. 1313–1321, 1999.
- [145] C.-C. Han, H.-Y. 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.
- [146] J. Wang and T. Tan, "A new face detection method based on shape information," *Pattern Recognition Letters*, vol. 21, pp. 463–471, 2000.
- [147] C. Chen, C.-W. Hsu, and T.-L. Lin, "Image prediction using face detection and triangulation," *Pattern Recognition Letters*, vol. 22, pp. 1347–1357, 2001.

- [148] H. Yao and W. Gao, "Face detection and location based on skin chrominance and lip chrominance transformation from color images," *Electronic Engineering*, vol. 34, pp. 1555–1564, 2001.
- [149] Y. Wang and B. Yuan, "A novel approach for human face detection from color images under complex background," *Pattern Recognition*, vol. 34, pp. 1983–1992, 2001.
- [150] O. Ayinde and Y.-H. Yang, "Region-based face detection," *Pattern Recognition*, vol. 35, pp. 2095 – 2107, 2002.
- [151] J. Zhou, D. Zhang, and C.-y. Wu, "Orientation analysis for rotated human face detection," *Image and Vision Computing*, vol. 20, 2002.
- [152] L. Hock Koh, S. Ranganath, and Y. V. Venkatesh, "An integrated automatic face detection and recognition system," *Pattern Recognition*, vol. 35, pp. 1259–1273, 2002.
- [153] I.-S. Hsieh, K.-C. Fan, and C. Lin, "A statistic approach to the detection of human faces in color nature scene," *Pattern Recognition*, vol. 35, pp. 1583–1596, 2002.
- [154] L.-l. Huang, A. Shimizu, Y. Hagihara, and H. Kobatake, "Face detection from cluttered images using a polynomial neural network," *Test*, vol. 51, pp. 197 – 211, 2003.
- [155] J. Wu and Z.-H. Zhou, "Efficient face candidates selector for face detection," *Pattern Recognition*, vol. 36, pp. 1175 – 1186, 2003.

- [156] K.-W. Wong, K.-M. Lam, and W.-C. Siu, “A robust scheme for live detection of human faces in color images,” *Signal Processing*, vol. 18, pp. 103–114, 2003.
- [157] —, “An efficient algorithm for human face detection and facial feature extraction under different conditions,” *Pattern Recognition*, vol. 34, pp. 1993–2004, 2004.
- [158] H. Bae and S. Kim, “Real-time face detection and recognition using hybrid-information extracted from face space and facial features,” *Image and Vision Computing*, vol. 23, pp. 1181–1191, 2005.
- [159] C. Kubleck and A. Ernst, “Face detection and tracking in video sequences using the modified census transformation,” *International Journal of Computer Vision*, vol. 24, pp. 564–572, 2006.
- [160] P. Shih and C. Liu, “Face detection using discriminating feature analysis and support vector machine,” *Pattern Recognition*, vol. 39, pp. 260 – 276, 2006.
- [161] T. Kondo and H. Yan, “Automatic human face detection and recognition under non-uniform illumination,” *Pattern Recognition*, vol. 32, 2006.
- [162] P. Wang and Q. Ji, “Multi-view face and eye detection using discriminant features,” *Computer Vision and Image Understanding*, vol. 105, pp. 99–111, 2007.
- [163] C. Lin, “Face detection in complicated backgrounds and different illumination conditions by using ycbcr color space and neural network,” *Pattern Recognition Letters*, vol. 28, pp. 2190–2200, 2007.

- [164] S. Phimoltares, C. Lursinsap, and K. Chamnongthai, "Face detection and facial feature localization without considering the appearance of image context," *Image (Rochester, N.Y.)*, vol. 25, pp. 741–753, 2007.
- [165] J. Meynet, V. Popovici, and J.-P. Thiran, "Face detection with boosted gaussian features," *Pattern Recognition*, vol. 40, pp. 2283 – 2291, 2007.
- [166] S. Kadoury and M. D. Levine, "Face detection in gray scale images using locally linear embeddings," *Computer Vision and Image Understanding*, vol. 105, pp. 1–20, 2007.
- [167] M. Castrillon, O. Deniz, C. Guerra, and M. Hernandez, "Encara2: Real-time detection of multiple faces at different resolutions in video streams /," *Journal of Visual Communication and Image Representation*, vol. 18, pp. 130–140, 2007.
- [168] C.-F. Juang and S.-J. Shiu, "Using self-organizing fuzzy network with support vector learning for face detection in color images," *Electrical Engineering*, vol. 71, pp. 3409–3420, 2008.
- [169] M. S. Bartlett, J. C. Hager, P. Ekman, and T. J. Sejnowski, "Measuring facial expressions by computer image analysis," *Psychophysiology*, vol. 36, pp. 253–263, 1999.
- [170] M. S. Bartlett, G. Littlewort, I. Fasel, and J. R. Movellan, "Real time face detection and facial expression recognition: Development and application to human computer interaction," in *In CVPR Workshop on CVPR for HCI*, 2003, pp. 139–157.

- [171] G. D. Kearney and S. Mckenzie, "Machine interpretation of emotion: Design of memory based expert system for interpreting facial expressions in terms of signaled emotions (janus)," *Cognitive Science*, vol. 17, pp. 589–622, 1993.
- [172] J. Cohn and T. Kanade, "Use of automated facial image analysis for measurement of emotion expression," in *The handbook of emotion elicitation and assessment. Oxford University Press Series in Affective Science*, J. A. Coan and J. B. Allen, Eds., 2006, (In Press).
- [173] A. Goneid and R. El Kaliouby, "Facial feature analysis of spontaneous facial expression," in *Proceedings of the 10th International AI Applications Conference*, 2002.
- [174] J. Lien, T. Kanade, J. Cohn, and C. Li, "Detection, tracking and classification of action units in facial expression," *Journal of Robotics and Autonomous Systems*.
- [175] I. Cohen, N. Sebe, L. Chen, A. Garg, and T. S. Huang, "Facial expression recognition from video sequences: Temporal and static modeling," in *Computer Vision and Image Understanding*, 2003, pp. 160–187.
- [176] D. Terzopoulos and K. Waters, "Analysis and synthesis of facial image sequences using physical and anatomical models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, no. 6, pp. 569–579, 1993.
- [177] I. A. Essa and A. P. Pentland, "Coding, analysis, interpretation, and recognition of facial expressions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 757–763, 1997.

- [178] 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 *FG '98: Proceedings of the 3rd. International Conference on Face & Gesture Recognition*. Washington, DC, USA: IEEE Computer Society, 1998, p. 454.
- [179] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Comput. Surv.*, vol. 35, no. 4, pp. 399–458, 2003.
- [180] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "Coding facial expressions with gabor wavelets," in *FG '98: Proceedings of the 3rd. International Conference on Face & Gesture Recognition*. Washington, DC, USA: IEEE Computer Society, 1998, p. 200.
- [181] M. J. Black and Y. Yacoob, "Tracking and recognizing rigid and non-rigid facial motions using local parametric models of image motion," in *In ICCV*, 1995, pp. 374–381.
- [182] M. N. Dailey, G. W. Cottrell, and R. Adolphs, "A six-unit network is all you need to discover happiness," in *In TwentySecond Annual Conference of the Cognitive Science Society*. Erlbaum, 2000, pp. 101–106.
- [183] Y. Y. Rosenblum and L. Davis, "Human expression recognition from motion using a radial basis function network architecture," *IEEE Transactions on Neural Networks*, vol. 7, no. 5, pp. 1121–1138, 1996.
- [184] I. Kotsia, I. Buciu, and I. Pitas, "An analysis of facial expression recognition under partial facial image occlusion," *Image Vision Comput.*, vol. 26, no. 7, pp.

1052–1067, 2008.

- [185] B. Hernández, G. Olague, R. Hammoud, L. Trujillo, and E. Romero, “Visual learning of texture descriptors for facial expression recognition in thermal imagery,” *Comput. Vis. Image Underst.*, vol. 106, no. 2-3, pp. 258–269, 2007.
- [186] Z. Hammal, L. Couvreur, A. Caplier, and M. Rombaut, “Facial expression classification: An approach based on the fusion of facial deformations using the transferable belief model,” *Int. J. Approx. Reasoning*, vol. 46, no. 3, pp. 542–567, 2007.
- [187] D. H. Kim, S. U. Jung, and M. J. Chung, “Extension of cascaded simple feature based face detection to facial expression recognition,” *Pattern Recogn. Lett.*, vol. 29, no. 11, pp. 1621–1631, 2008.
- [188] D. Liang, J. Yang, Z. Zheng, and Y. Chang, “A facial expression recognition system based on supervised locally linear embedding,” *Pattern Recogn. Lett.*, vol. 26, no. 15, pp. 2374–2389, 2005.
- [189] T. Xiang, M. K. H. Leung, and S. Y. Cho, “Expression recognition using fuzzy spatio-temporal modeling,” *Pattern Recogn.*, vol. 41, no. 1, pp. 204–216, 2008.
- [190] M. Pantic and L. Rothkrantz, “Expert system for automatic analysis of Facial Expression,” *Image and Vision Computing Journal*, vol. 18, no. 11, pp. 881–905, July 2000. [Online]. Available: <http://pubs.doc.ic.ac.uk/Pantic-IVCJ00/>
- [191] C.-L. Huang and Y.-M. Huang, “Facial expression recognition using model-based feature extraction and action parameters classification,” *Journal of Visual*

- Communication and Image Representation*, vol. 8, no. 3, pp. 278–290, 1997.
- [192] M. Pantic and L. Rothkrantz, “Automatic analysis of facial expressions: The state of the art,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp. 1424–1445, December 2000.
- [193] G. W. Cottrell and J. Metcalfe, *EMPATH: Face, Emotion, Gender Recognition Using Holons*, 1991, vol. ed, pp. 564–571.
- [194] A. Rahardja, A. Sowmya, and W. H. Wilson, “A neural network approach to component versus holistic recognition of facial expressions in images,” *SPIE, Intelligent Robots and Computer Vision X: Algorithms and Techniques*, vol. 1607, pp. 62–70, 1991.
- [195] K. Matsuno, C. W. Lee, and S. Tsuji, “Recognition of facial expression with potential net,” in *Proc. Asian Conf. Computer Vision*, 1993, pp. 504–507.
- [196] K. Mase, “Recognition of facial expression from optical flow,” *IEICE Trans.*, vol. 74, pp. 3474–3483, 1991.
- [197] Y. Moses, D. Reynard, and A. Blake, “Determining facial expressions in real time,” in *Proc. Int’l Conf. Automatic Face and Gesture Recognition*, vol. pp, 1995, pp. 332–337.
- [198] M. Rosenblum, Y. Yacoob, and L. Davis, “Human emotion recognition from motion using a radial basis function network architecture,” in *Proc. IEEE Workshop on Motion of Non-Rigid and Articulated Objects*, vol. pp, 1994, pp. 43–49.

- [199] Y. Yacoob and L. Davis, "Recognizing facial expressions by spatio-temporal analysis," *Proc. Int'l Conf. Pattern Recognition*, vol. 1, pp. 747–749, 1994.
- [200] H. Kobayashi and F. Hara, "Recognition of six basic facial expressions and their strength by neural network," in *Proc. Int'l Workshop Robot and Human Comm.*, 1992, pp. 381–386.
- [201] H. Ushida, T. Takagi, and T. Yamaguchi, "Recognition of facial expressions using conceptual fuzzy sets," *Proc. Conf. Fuzzy Systems*, vol. 1, pp. 594–599, 1993.
- [202] P. Vanger, R. Honlinger, and H. Haken, "Applications of synergetics in decoding facial expression of emotion," in *Proc. Int'l Conf. Automatic Face and Gesture Recognition*, vol. pp, 1995, pp. 24–29.
- [203] G. J. Edwards, T. F. Cootes, and C. J. Taylor, "Face recognition using active appearance models," in *ECCV '98: Proceedings of the 5th European Conference on Computer Vision-Volume II*. London, UK: Springer-Verlag, 1998, pp. 581–595.
- [204] H. Hong, H. Neven, and C. Von Der Malsburg, "Online facial expression recognition based on personalized galleries," in *FG '98: Proceedings of the 3rd. International Conference on Face & Gesture Recognition*. Washington, DC, USA: IEEE Computer Society, 1998, p. 354.
- [205] C. L. Huang and Y. M. Huang, "Facial expression recognition using model-based feature extraction and action parameters classification," *J. Visual Comm. and Image Representation*, vol. 8, pp. 278–290.

- [206] C. Padgett and G. W. Cottrell, "Representing face images for emotion classification," in *Proc. Conf. Advances in Neural Information Processing Systems*, vol. pp, 1996, pp. 894–900.
- [207] M. Yoneyama, Y. Iwano, A. Ohtake, and K. Shirai, "Facial expressions recognition using discrete hopfield neural networks," *Proc. Int'l Conf. Information Processing*, vol. 3, pp. 117–120, 1997.
- [208] Z. Zhang, M. Lyons, M. Schuster, and S. Akamatsu, "Comparison between geometrybased and gabor wavelets-based facial expression recognition using multi-layer perceptron," in *Proc. Int'l Conf. Automatic Face and Gesture Recognition*, vol. pp, 1998, pp. 454–459.
- [209] F. Hara and H. Kobayashi, "State of the art in component development for interactive communication with humans," *Advanced Robotics*, vol. 11, pp. 585–604, 1997.
- [210] J. Zhao and G. Kearney, "Classifying facial emotions by backpropagation neural networks with fuzzy inputs," *Proc. Conf. Neural Information Processing*, vol. 1, pp. 454–457, 1996.
- [211] M. J. Black and Y. Yacoob, "Recognizing facial expressions in image sequences using local parameterized models of image motion," *Int'l J. Computer Vision*, vol. 25, pp. 23–48, 1997.
- [212] I. Essa and A. Pentland, "Coding, analysis interpretation, recognition of facial expressions," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, pp. 757–763, July 1997.

- [213] S. Kimura and M. Yachida, "Facial expression recognition and its degree estimation," in *Proc. Computer Vision and Pattern Recognition*, 1997, pp. 295–300.
- [214] T. Otsuka and J. Ohya, "Spotting segments displaying facial expression from image sequences using hmm," in *Proc. Int'l Conf. Automatic Face and Gesture Recognition*, vol. pp, 1998, pp. 442–447.
- [215] M. Wang, Y. Iwai, and M. Yachida, "Expression recognition from time-sequential facial images by use of expression change model," in *Proc. Int'l Conf. Automatic Face and Gesture Recognition*, 1998, pp. 324–329.
- [216] J. F. Cohn, A. J. Zlochower, J. J. Lien, and T. Kanade, "Feature-point tracking by optical flow discriminates subtle differences in facial expression," in *Proc. Int'l Conf. Automatic Face and Gesture Recognition*, vol. pp, 1998, pp. 396–401.
- [217] M. J. Lyons, J. Budynek, and S. Akamatsu, "Automatic classification of single facial images," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 21, pp. 1357–1362, 1999.
- [218] M. J. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "Coding facial expressions with gabor wavelets," in *Proc. Int'l Conf. Automatic Face and Gesture Recognition*, vol. pp, 1998, pp. 200–205.
- [219] B. Moghaddam and A. Pentland, "Face recognition using view-based and modular eigenspaces," in *In Automatic Systems for the Identification and Inspection of Humans, SPIE*, 1994, pp. 12–21.

- [220] E. P. Simoncelli, "Distributed representation and analysis of visual motion," Ph.D. dissertation, Cambridge, MA, USA, 1993.
- [221] J. Wang and E. Adelson, "Layered representation for motion analysis," June 1993, pp. 361–366.
- [222] T. Otsuka and J. Ohya, "Recognition of facial expressions using hmm with continuous output probabilities," in *Proc. Int'l Workshop Robot and Human Comm.*, vol. pp, 1996, pp. 323–328.
- [223] X. Zhou, X. Huang, and Y. Wang, "Real-time facial expression recognition in the interactive game based on embedded hidden markov model," in *CGIV '04: Proceedings of the International Conference on Computer Graphics, Imaging and Visualization*. Washington, DC, USA: IEEE Computer Society, 2004, pp. 144–148.
- [224] M. Pardas, A. Bonafonte, and J. Landabaso, "Emotion recognition based on mpeg-4 facial animation parameters," vol. 4, 2002, pp. IV–3624–IV–3627 vol.4.
- [225] Y. Wang, H. Ai, B. Wu, and C. Huang, "Real time facial expression recognition with adaboost," in *ICPR '04: Proceedings of the Pattern Recognition, 17th International Conference on (ICPR'04) Volume 3*. Washington, DC, USA: IEEE Computer Society, 2004, pp. 926–929.
- [226] L. Ma and K. Khorasani, "Facial expression recognition using constructive feed-forward neural networks," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 34, no. 3, pp. 1588–1595, June 2004.

- [227] H. Tao and T. S. Huang, "Connected vibrations: A modal analysis approach to non-rigid motion tracking," in *In CVPR*, 1998, pp. 735–740.
- [228] T. Wang, H. Ai, and G. Huang, "A two-stage approach to automatic face alignment," H. Lu and T. Zhang, Eds., vol. 5286, no. 1. SPIE, 2003, pp. 558–563. [Online]. Available: <http://link.aip.org/link/?PSI/5286/558/1>
- [229] A. V. Nefian and H. H. I. Monson, "Face recognition using an embedded hmm," in *IEEE Conference on Audio and Video-based Biometric Person Authentication*, 1999.
- [230] N. Karayiannis and J. Bezdek, "An integrated approach to fuzzy learning vector quantization and fuzzy c-means clustering," *Fuzzy Systems, IEEE Transactions on*, vol. 5, no. 4, pp. 622–628, November 1997.
- [231] T. Kobayashi, Y. Ogawa, K. Kato, and K. Yamamoto, "Learning system of human facial expression for a family robot," May 2004, pp. 481–486.
- [232] B. Abboud and F. Davoine, "Appearance factorization based facial expression recognition and synthesis," vol. 4, August 2004, pp. 163–166 Vol.4.
- [233] M. Yuki, W. W. Maddux, and 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," *Journal of Experimental Social Psychology*, vol. 43, pp. 301–311, 2007.
- [234] H. Elfenbein and N. Ambady, "When familiarity breeds accuracy: cultural exposure and facial emotion recognition." *Journal of Personality and Social Psy-*

- chology*, vol. 85, no. 2, pp. 276–290, 2003.
- [235] I.-O. Stathopoulou and G. A. Tsihrintzis, “A new neural network-based method for face detection in images and applications in bioinformatics,” in *6th International Workshop on Mathematical Methods in Scattering Theory and Biomedical Technology*, Tsepelovo, Greece, September, 15-18 2003.
- [236] —, “Detection and Expression Classification Systems for Face Images (FADECS),” in *Proceedings of the IEEE Workshop on Signal Processing Systems (SiPS’05)*, Athens, Greece, November 2005.
- [237] —, “A neural network-based system for face detection in low quality web camera images,” in *International Conference on Signal Processing and Multimedia Applications*, Barcelona, Spain, July, 28-31 2007.
- [238] —, “Appearance - based face detection with artificial neural networks,” *Journal of Intelligent Decision Technologies*, IOS Press, 2009, to appear.
- [239] C. Fowlkes, S. Belongie, and J. Malik, “Efficient spatiotemporal grouping using the nystrom method,” *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, vol. 1, pp. I-231–I-238 vol.1, 2001.
- [240] I.-O. Stathopoulou and G. A. Tsihrintzis, “A neural network-based facial analysis system,” in *Proceedings of the 5th International Workshop on Image Analysis for Multimedia Interactive Services*, Lisboa, Portugal, April 2004.
- [241] —, “An Improved Neural Network-Based Face Detection and Facial Ex-

- pression Classification System,” in *IEEE International Conference on Systems, Man, and Cybernetics*, The Hague, Netherlands, October 2004.
- [242] —, “Pre-processing and expression classification in low quality face images,” in *Proceedings of 5th EURASIP Conference on Speech and Image Processing, Multimedia Communications and Services*, July 2005.
- [243] —, “Evaluation of the Discrimination Power of Features Extracted from 2-D and 3-D Facial Images for Facial Expression Analysis,” in *Proceedings of the 13th European Signal Processing Conference*, Antalya, Turkey, September 2005.
- [244] —, “An Accurate Method for eye detection and feature extraction in face color images,” in *Proceedings of the 13th International Conference on Signals, Systems, and Image Processing*, Budapest, Hungary, September 2006.
- [245] —, “Neu-faces: A neural network-based face image analysis system,” in *ICANN'07: Proceedings of the 8th international conference on Adaptive and Natural Computing Algorithms, Part II, LNCS: Vol. 4432*. Berlin, Heidelberg: Springer-Verlag, 2007, pp. 449–456.
- [246] —, “Comparative performance evaluation of artificial neural network-based vs. human facial expression classifiers for facial expression recognition,” in *KES-IMSS 2008: 1st International Symposium on Intelligent Interactive Multimedia Systems and Services, SCI: Vol. 142*. Berlin, Heidelberg: Springer-Verlag, 2008, pp. 55–65.
- [247] A. S. Lampropoulos, I.-O. Stathopoulou, and G. A. Tsihrintzis, “Comparative performance evaluation of classifiers for facial expression recognition,” in *KES-*

- IMSS 2009: 2nd International Symposium on Intelligent Interactive Multimedia Systems and Services*, 2009, to appear.
- [248] I.-O. Stathopoulou and G. A. Tsihrintzis, “A face detection and visual-facial emotion recognition system,” in preparation.
- [249] D. A. Forsyth and M. Fleck, “Finding naked people,” in *In European Conference on Computer Vision*. Springer-Verlag, 1996, pp. 593–602.
- [250] C. M. Bishop, *Neural Networks for Pattern Recognition*, 1st ed. Oxford University Press, 1995.
- [251] S. Haykin, *Neural Networks*, 2nd ed. Prentice Hall, 1999.
- [252] R. O. Duda, P. E. Hart, and D. G. Strock, *Pattern Classification*, 2nd ed. John Wiley, 2000.
- [253] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, 1st ed. Academic Press, 1998.
- [254] I. Russel, “Emotion and Memory,” *Behavioural Brain Research*, vol. 58, no. 1-2, p. R9, December 1993.
- [255] H. Graf, E. Cosatto, V. Strom, and F. Huang, “Visual prosody: Facial movements accompanying speech,” in *5th IEEE International Conference on Automatic Face and Gesture Recognition*, 2002, pp. 381–386.
- [256] L. S. Chen, T. S. Huang, T. Miyasato, and R. Nakatsu, “Multimodal human emotion/expression recognition,” in *Proc. Int’l Conf. Automatic Face and Gesture Recognition*, vol. pp, 1998, pp. 366–371.

- [257] M. Pantic and L. J. M. Rothkrantz, "Toward an affect-sensitive multimodal human-computer interaction," *Proceedings of the IEEE*, vol. 91, no. 9, pp. 1370–1390, September 2003.
- [258] C. Busso, Z. Deng, S. Yildirim, M. Bulut, C. M. Lee, A. Kazemzadeh, S. Lee, U. Neumann, and S. Narayanan, "Analysis of emotion recognition using facial expressions, speech and multimodal information," in *ICMI '04: Proceedings of the 6th international conference on Multimodal interfaces*. New York, NY, USA: ACM, 2004, pp. 205–211.
- [259] E. Alepis, M. Virvou, and K. Kabassi, "Affective student modeling based on microphone and keyboard user actions," in *ICALT '06: Proceedings of the Sixth IEEE International Conference on Advanced Learning Technologies*. Washington, DC, USA: IEEE Computer Society, 2006, pp. 139–141.
- [260] M. Virvou, G. A. Tsihrintzis, E. Alepis, I. O. Stathopoulou, and K. Kabassi, "Combining empirical studies of audio-lingual and visual-facial modalities for emotion recognition," in *KES '07: Knowledge-Based Intelligent Information and Engineering Systems and the XVII Italian Workshop on Neural Networks on Proceedings of the 11th International Conference, LNAI: Vol. 4693*. Berlin, Heidelberg: Springer-Verlag, 2007, pp. 1130–1137.
- [261] M. Virvou, G. A. Tsihrintzis, E. Alepis, I.-O. Stathopoulou, and K. Kabassi, "On combining audio-lingual and visual-facial modalities for emotion recognition: Empirical studies," in preparation.
- [262] G. A. Tsihrintzis, M. Virvou, E. Alepis, and I.-O. Stathopoulou, "Towards

- improving visual-facial emotion recognition through use of complementary keyboard-stroke pattern information,” in *ITNG '08: Proceedings of the Fifth International Conference on Information Technology: New Generations*. Washington, DC, USA: IEEE Computer Society, 2008, pp. 32–37.
- [263] M. Virvou, G. A. Tsihrintzis, E. Alepis, and I.-O. Stathopoulou, “Designing a multi-modal affective knowledge-based user interface: combining empirical studies,” in *JCKBSE: 8th Joint Conference on Knowledge-Based Software Engineering*, 2008, pp. 250–259.
- [264] G. A. Tsihrintzis, M. Virvou, I.-O. Stathopoulou, and E. Alepis, “On improving visual-facial emotion recognition with audio-lingual and keyboard stroke pattern information,” in *WI-IAT '08: Proceedings of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*. Washington, DC, USA: IEEE Computer Society, 2008, pp. 810–816.

Publications which have resulted from this research up to the instant of submission of this Thesis

Proceedings in International Peer Reviewed Conferences:

1. I.-O. Stathopoulou and G. A. Tsihrintzis, 'A new neural network-based method for face detection in images and applications in bioinformatics', In Proc. 6th International Workshop on Mathematical Methods in Scattering Theory and Biomedical Technology, Tsepelovo, Greece, September, 15-18, 2003.
 2. I.-O. Stathopoulou and G. A. Tsihrintzis, 'A neural network-based facial analysis system', In Proceedings of the 5th International Workshop on Image Analysis for Multimedia Interactive Services, Lisboa, Portugal, April 2004.
 3. I.-O. Stathopoulou and G. A. Tsihrintzis, 'An Improved Neural Network-Based Face Detection and Facial Expression Classification System', In Proc. of IEEE International Conference on Systems, Man, and Cybernetics, The Hague, Netherlands, October 2004.
 4. I.-O. Stathopoulou and G. A. Tsihrintzis, 'Pre-processing and expression classification in low quality face images', In Proceedings of 5th EURASIP Conference on Speech and Image Processing, Multimedia Communications and Services, July 2005.
-

5. I.-O. Stathopoulou and G. A. Tsihrintzis, 'Evaluation of the Discrimination Power of Features Extracted from 2-D and 3-D Facial Images for Facial Expression Analysis', In Proceedings of the 13th European Signal Processing Conference, Antalya, Turkey, September 2005.
6. I.-O. Stathopoulou and G. A. Tsihrintzis, 'Detection and Expression Classification Systems for Face Images (FADECS)', In Proceedings of the IEEE Workshop on Signal Processing Systems (SiPS05), Athens, Greece, November 2005.
7. I.-O. Stathopoulou and G. A. Tsihrintzis, 'An Accurate Method for eye detection and feature extraction in face color images', In Proceedings of the 13th International Conference on Signals, Systems, and Image Processing, Budapest, Hungary, September 2006.
8. I.-O. Stathopoulou and G. A. Tsihrintzis, 'Towards automated inferencing of Emotional State from face Images', In Proc. 2nd International Conference on Software and Data Technologies, Barcelona, Spain, July, 5-8, 2007.
9. I.-O. Stathopoulou and G. A. Tsihrintzis, 'A neural network-based system for face detection in low quality web camera images', In Proc. of International Conference on Signal Processing and Multimedia applications, Barcelona, Spain, July, 28-31 2007.
10. G. A. Tsihrintzis, M. Virvou, E. Alepis, and I.-O. Stathopoulou, 'Towards improving visual-facial emotion recognition through use of complementary keyboard-stroke pattern information', In ITNG'08: Proceedings of the Fifth International Conference on Information Technology: New Generations. Washington, DC,

USA: IEEE Computer Society, 2008, pp. 32-37.

11. G. A. Tsihrintzis, M. Virvou, I.-O. Stathopoulou, and E. Alepis, 'On improving visual-facial emotion recognition with audio-lingual and keyboard stroke pattern information', In WI-IAT '08: Proceedings of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology. Washington, DC, USA: IEEE Computer Society, 2008, pp. 810-816.

Lecture Notes:

1. I.-O. Stathopoulou and G. A. Tsihrintzis, 'Automated Processing and Classification of Face Images for Human Computer Interaction Applications, ser. Studies in Computational Intelligence. Springer Berlin/Heidelberg, 2008, vol. 104, ch. Intelligent Interactive Systems in Knowledge-Based Environments, pp. 107-136.
2. I.-O. Stathopoulou and G. A. Tsihrintzis, 'NEU-FACES: A neural network-based face image analysis system', In ICANNGA '07: Proceedings of the 8th international conference on Adaptive and Natural Computing Algorithms, Part II, LNCS: Vol. 4432, Berlin, Heidelberg: Springer/Verlag, 2007, pp. 449-456.
3. M. Virvou, G. A. Tsihrintzis, E. Alepis, I. O. Stathopoulou, and K. Kabassi, 'Combining empirical studies of audio-lingual and visual-facial modalities for emotion recognition', In KES '07: Knowledge-Based Intelligent Information and Engineering Systems and the XVII Italian Workshop on Neural Networks on Proceedings of the 11th International Conference, LNAI: Vol. 4693. Berlin, Heidelberg: Springer/Verlag, 2007, pp. 1130-1137.

4. I.-O. Stathopoulou and G. A. Tsihrintzis, 'Comparative performance evaluation of artificial neural network-based vs. human facial expression classifiers for facial expression recognition', In KES-IMSS 2008: 1st International Symposium on Intelligent Interactive Multimedia Systems and Services, SCI: Vol. 142. Berlin, Heidelberg: Springer-Verlag, 2008, pp. 55-65.
5. M. Virvou, G. A. Tsihrintzis, E. Alepis, and I.-O. Stathopoulou, 'Designing a multi-modal affective knowledge-based user interface: combining empirical studies', In JCKBSE: 8th Joint Conference on Knowledge-Based Software Engineering, 2008, pp. 250-259.
6. A. S. Lampropoulos, I.-O. Stathopoulou, and G. A. Tsihrintzis, 'Comparative performance evaluation of classifiers for facial expression recognition', In KES-IMSS 2009: 2nd International Symposium on Intelligent Interactive Multimedia Systems and Services, 2009, (To appear).

Book Chapters:

1. I.-O. Stathopoulou and G. A. Tsihrintzis, 'Automated Processing and Classification of Face Images for Human Computer Interaction Applications', ser. Studies in Computational Intelligence. Springer Berlin / Heidelberg, 2008, vol. 104, ch. Intelligent Interactive Systems in Knowledge-Based Environments, pp. 107-136.

Journal Papers:

1. I.-O. Stathopoulou and G. A. Tsihrintzis, 'Appearance - based face detection with artificial neural networks', *Journal of Intelligent Decision Technologies*, IOS Press, 2009 (to appear).
2. I.-O. Stathopoulou and G. A. Tsihrintzis, 'An empirical study of facial expression classification by human observers' (in preparation).
3. I.-O. Stathopoulou and G. A. Tsihrintzis, 'A face detection and visual-facial emotion recognition system', (in preparation).
4. M. Virvou, G. A. Tsihrintzis, E. Alepis, I.-O. Stathopoulou, and K. Kabassi, 'On combining audio-lingual and visual-facial modalities for emotion recognition: Empirical studies', (in preparation).