

Recognizing Emotions Conveyed through Facial Expressions

Teena Chakkalayil Hassan

Publisher: Dean Prof. Dr. Wolfgang Heiden

University of Applied Sciences Bonn-Rhein-Sieg,
Department of Computer Science

Sankt Augustin, Germany

May 2017

Technical Report 04-2017



**Hochschule
Bonn-Rhein-Sieg**
University of Applied Sciences

ISSN 1869-5272

ISBN 978-3-96043-047-6

Copyright © 2017, by the author(s). All rights reserved. Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission.

Das Urheberrecht des Autors bzw. der Autoren ist unveräußerlich. Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Das Werk kann innerhalb der engen Grenzen des Urheberrechtsgesetzes (UrhG), *German copyright law*, genutzt werden. Jede weitergehende Nutzung regelt obiger englischsprachiger Copyright-Vermerk. Die Nutzung des Werkes außerhalb des UrhG und des obigen Copyright-Vermerks ist unzulässig und strafbar.

Digital Object Identifier **doi:10.18418/978-3-96043-047-6**
DOI-Resolver **<http://dx.doi.org/>**

Report on R&D 2

Recognizing Emotions Conveyed through Facial Expressions

Teena Chakkalayil Hassan*

B-IT Master Studies Autonomous Systems

University of Applied Sciences Bonn-Rhein-Sieg

Advisors: Prof. Dr. Erwin Prassler** and Matthias Füller††

Submitted on September 30, 2013

Minor Revisions for Tech Report: February 28, 2017

*teena.hassan@smail.inf.fh-bonn-rhein-sieg.de

**erwin.prassler@h-brs.de

††matthias.fueller@h-brs.de

Abstract

Emotional communication is a key element of habilitation care of persons with dementia. It is, therefore, highly preferable for assistive robots that are used to supplement human care provided to persons with dementia, to possess the ability to recognize and respond to emotions expressed by those who are being cared-for. Facial expressions are one of the key modalities through which emotions are conveyed. This work focuses on computer vision-based recognition of facial expressions of emotions conveyed by the elderly.

Although there has been much work on automatic facial expression recognition, the algorithms have been experimentally validated primarily on young faces. The facial expressions on older faces has been totally excluded. This is due to the fact that the facial expression databases that were available and that have been used in facial expression recognition research so far do not contain images of facial expressions of people above the age of 65 years. To overcome this problem, we adopt a recently published database, namely, the FACES database, which was developed to address exactly the same problem in the area of human behavioural research. The FACES database contains 2052 images of six different facial expressions, with almost identical and systematic representation of the young, middle-aged and older age-groups.

In this work, we evaluate and compare the performance of two of the existing image-based approaches for facial expression recognition, over a broad spectrum of age ranging from 19 to 80 years. The evaluated systems use Gabor filters and uniform local binary patterns (LBP) for feature extraction, and AdaBoost.MH with multi-threshold stump learner for expression classification. We have experimentally validated the hypotheses that facial expression recognition systems trained only on young faces perform poorly on middle-aged and older faces, and that such systems confuse ageing-related facial features on neutral faces with other expressions of emotions. We also identified that, among the three age-groups, the middle-aged group provides the best generalization performance across the entire age spectrum. The performance of the systems was also compared to the performance of humans in recognizing facial expressions of emotions. Some similarities were observed, such as, difficulty in recognizing the expressions on older faces, and difficulty in recognizing the expression of sadness.

The findings of our work establish the need for developing approaches for facial expression recognition that are robust to the effects of ageing on the face. The scientific results of our work can be used as a basis to guide future research in this direction.

List of Abbreviations and Acronyms

AAM	A ctive A ppearance M odel
AU	A ction U nit
CK	C ohn- K anade
CV	C ross- V alidation
DBN	D ynamic B ayesian N etwork
DWT	D iscrete W avelet T ransform
ELBP	E xtended L ocal B inary P attern
EMFACS	E motion F acial A ction C oding S ystem
FACS	F acial A ction C oding S ystem
FEED	F acial E xpressions and E motion D atabase
FER	F acial E xpression R ecognition
FFD	F ree- F orm D eformation
HMM	H idden M arkov M odel
ICA	I ndependent C omponent A nalysis
JAFFE	J apanese F emale F acial E xpressions
KCCA	K ernel C anonical C orrelation A nalysis
LBP	L ocal B inary P attern
LBP-TOP	L ocal B inary P attern- T hree O rthogonal P lanes
LDA	L inear D iscriminant A nalysis
LDP	L ocal D irectional P attern
LG	L abeled G raph
LOIO	L eave O ne I mage O ut
LOSO	L eave O ne S ubject O ut
LP	L inear P rogramming
LPQ	L ocal P hase Q uantization
LPQ-TOP	L ocal P hase Q uantization- T hree O rthogonal P lanes
MHI	M otion H istory I mages
MLP	M ulti- L ayer P erceptron
MTSL	M ulti- T hreshold S tump L earner
NN	N earest N eighbor
PCA	P rincipal C omponent A nalysis
POFA	P ictures of F acial A ffect
RBF	R adial B asis F unction
RBFN	R adial B asis F unction N etwork
ROC	R eciever- O perator C haracteristic
SVM	S upport V ector M achine

Contents

1	Introduction	6
2	Background	7
2.1	Emotions and Facial Expressions	7
2.2	Facial Expression Analysis	8
2.2.1	Overview	8
2.2.2	System Architecture	8
2.2.3	Face Descriptors	11
2.3	Facial Expression Databases	14
2.3.1	Japanese Female Facial Expressions Database	15
2.3.2	Pictures of Facial Affect	15
2.3.3	MMI Facial Expression Database	16
2.3.4	Facial Expressions and Emotion Database	16
2.3.5	Extended Cohn-Kanade Dataset	17
2.3.6	FACES	18
2.3.7	Others	21
2.3.8	Open Issues and Challenges	21
2.4	Neurological, Psychological and Behavioural Research	22
3	Related Work	24
3.1	Facial Expression Recognition	25
3.2	Facial Action Unit Recognition	32
3.3	Multimodal Emotion Recognition	34
3.4	Deficits	34
4	Problem Formulation	37
4.1	Hypotheses	39
5	Validation of Hypotheses	39
5.1	Database Selection	39
5.2	Human Performance on Selected Databases	41
5.3	Experiments	42
5.3.1	Cross-Database Experiments	42
5.3.2	Intra-age-group Experiments	51
5.3.3	Inter-age-group Experiments	61
5.4	Scientific Results	70
5.5	Enhancements: Suggestions, Ideas	71

5.5.1	Improvements to Evaluated Systems	71
5.5.2	Proposed Approach for Facial Expression Recognition across Age- groups	73
6	Comparison of Human and System Performance	75
7	Summary	78
8	Future Work	78
	Appendices	92
A	Facial Action Coding System (FACS)	92
B	General Resources	93

1 Introduction

¹Statistics [67, 60] show that people, especially women, belonging to the 65+ age-group have higher risks of developing dementia. With population ageing², a significant increase in the number of persons with dementia is expected in the coming decades. According to estimates [98], there would be roughly 65.7 million people living with dementia in the world by the year 2030 (almost double the figure in 2010) and 115.4 million by the year 2050 (almost double the projected figure for 2030).

People with dementia have been found to retain some ability to recognise and respond to emotions conveyed through facial expressions [9]. Such evidence has encouraged the use of non-verbal, emotional communication in dementia care [9, 24]. With assistive robots being considered as a means to supplement assistance and care provided by humans, enhancing robots with emotional intelligence would enable them to stimulate emotions in persons with dementia, as well as provide effective companionship. Such robots should possess the ability to recognise and respond to emotions expressed by the persons with dementia, who, predominantly, belong to the age-group of 65 years and above.

In this work, one of the dimensions of this problem is studied, namely, the recognition of emotions conveyed through facial expressions by the elderly. There are numerous methods for facial expression recognition which have been experimentally validated [94, 44, 23]. However, these have been validated on images and image sequences of young faces. The usefulness of such systems in recognizing facial expressions on older faces has not been verified. The objective of this work is to identify an approach that would be suitable for recognizing facial expressions on older faces. In order to do this, we draw in on the knowledge about human performance in recognizing facial expressions of elderly, evaluate and compare the performance of existing approaches for facial expression recognition on older faces, and search for cues from age estimation methods.

In section 2, the reader is provided an overview of facial expression analysis, facial expression recognition systems and facial expression databases. Some of the findings of neurological, psychological and behavioural research are also discussed. In section 3, some of the main works in the field of facial expression analysis are reviewed and deficits are identified. Section 4 describes the problem that is studied in this work. The validation of the hypotheses postulated in section 4 is elaborated in section 5. The results of compari-

¹Certain parts of the introduction section have been submitted as part of the proposal, and as part of the Master Seminar assignment- "Proposal for R&D2"

²Demographic projections estimate that, by 2050, people aged above 60 years will constitute 22% of the world population and 34% of the European population [83].

son of facial expression recognition performance of humans and systems are summarized in section 6.

The readers of this report are assumed to have basic knowledge in the areas of computer vision, image processing and machine learning.

2 Background

In this section, some of the fundamental elements of research on facial expression analysis are discussed briefly. Key insights from neurological, psychological and behavioural research on facial expression recognition in humans are also discussed.

2.1 Emotions and Facial Expressions

Emotions play a key role in the interaction between humans, and in social communication. Emotions not only convey the mood and behavioural intent of the expresser, but also influence the emotional state and behavioural response of the perceivers [63]. Emotions are conveyed through facial expressions, vocal intonations, body postures, body movements, etc. [64].

Different sets of emotions have been identified as basic emotions by researchers, based on different criteria [89]. In 1971, Ekman and Friesen [40] identified **six basic or prototypic facial expressions of emotions** that are universal across different cultures and ethnicities. These expressions correspond to the emotions of **anger, disgust, fear, joy, sadness and surprise**. In [106], Shaver et al. proposed a three-level hierarchical classification of emotions, in which, love, joy, surprise, anger, sadness and fear are identified as primary emotions. In [10], Adolphs identified happiness, fear, anger, disgust, sadness, surprise and contempt as basic emotions. Facial expression recognition studies primarily attempt to recognize the six prototypic expressions of emotions identified by Ekman and Friesen in [40]. The same set is adopted in our work as well.

The significance of facial expressions in face-to-face communication has been established. The study performed by Mehrabian [80] revealed that facial expressions contribute 55% to how liked the listener feels. The contribution of vocal intonations is 38% and that of the actual words spoken is only 7%. The expressive facial actions displayed in everyday life often involve only movements of certain parts of the face, for example, tightening or lowering of lips [30]. Facial expressive movements not only convey emotions, but are also used for non-verbal communication of messages and their meanings [38].

2.2 Facial Expression Analysis

2.2.1 Overview

Within the realm of facial expression analysis, two different research directions have been pursued. One of these deals with the recognition of a fixed set of expressions of emotions (facial affect), especially the six prototypic expressions, namely, anger, disgust, fear, joy, sadness and surprise, the neutral expression and the expression of contempt. Such systems, however, cannot recognize any arbitrary, non-prototypic facial expressions, such as the expression of boredom or interest, that it has not been trained on. Moreover, the prototypic expressions occur less frequently in social interactions, compared to facial actions such as raising of eyebrows, tightening of lips, etc. [30, 110]. Therefore, the other research direction deals with the recognition of **facial action units** (AUs) based on the **Facial Action Coding System**, FACS (see Appendix A), which was developed by Ekman et al. [42] originally for the manual analysis of faces.

Facial AUs represent atomic facial muscle movements. The facial AU recognition systems, therefore, identify the atomic expressive movements of different parts of the face, such as eyebrows, nose, lips, cheeks, etc. As a result, these systems are not limited to the recognition of a few prototypic expressions of emotions. Moreover, by analysing the recognized facial AUs, and by using the mapping of facial AUs to basic emotions provided in EMFACS (**E**motion **F**ACS [42]), the basic emotions can also be identified through facial AU recognition.

Facial expression and facial AU recognition systems find application, primarily, in the fields of human-computer and human-robot interaction. They provide an intuitive communication interface, and help in making robots more sociable. Consumer research, patient monitoring, behaviour analysis, animation, etc. are some of the other application areas of facial expression analysis. These systems also help in accelerating research in the fields of behavioral science and psychology, by automating the process of FACS and emotion labelling of images and video sequences [65]³.

2.2.2 System Architecture

Automatic facial expression recognition involves three main steps, namely, *face detection*, *extraction of facial expression information*, and *classification of facial expression* [94]. Face detection and facial expression information extraction in still images are referred to

³Manual labelling of facial AUs in images and videos is a very time-consuming and tedious process [93]

as **face and feature localization**, and the same processes when performed on image sequences are referred to as **face and feature tracking** [94].

2.2.2.1 Face and Facial Feature Detection

The facial features to be detected depend on the type of descriptors used to describe the face. There are two types of descriptors, namely, geometric and appearance-based. When **geometric descriptors** are used, it is necessary to localize or track a set of facial fiducial points. The facial expression or action unit recognition is performed on the basis of the relative positions of these points and the changes in their positions in subsequent frames [55]. **Appearance-based descriptors**, on the other hand, code the textural details of the face, such as, wrinkles, folds, bulges, etc. When appearance-based descriptors are used, it is often necessary to estimate the whole face region. However, some methods (for example, [82]) also require sub-regions of the face, such as, the eyes, nose and mouth, to be located. There are also methods that use a combination of geometric and appearance-based descriptors (for example, [131]), requiring the detection of the face as well as facial fiducial points.

For automatic face detection, methods based on the Viola-Jones object detection framework [120] are commonly used. Various public classifiers based on this framework, used for detecting face and facial features (such as, eyes, nose, mouth, head, and shoulders), are listed and compared in [31]. The Computer Expression Recognition Toolbox (CERT), a toolbox for automatic real-time recognition of facial expressions and facial action units, developed by Littlewort et al. [66], also uses Viola-Jones framework-based method to detect the face. It uses the facial feature detection approach based on context-dependent inference (CDI), used in [37], for detection of 10 facial fiducial points within the face region. After the detection of the fiducial points, the face region is re-estimated on the basis of the location of these points.

In some of the works (for example, [105]), manual input was used for reliable face region estimation. The eye positions were manually labelled and a face model was applied to estimate the face region. However, these systems are no longer automatic.

The estimated face regions are often normalized to the same resolution, with eye positions aligned and distances between the eyes fixed.

2.2.2.2 Feature Extraction

After detecting the face and facial features, facial expression information is extracted using appropriate face descriptor(s). Face descriptors are of two types, namely, geometric and appearance-based. Accordingly, the features extracted are called geometric and appearance-based features. Geometric and appearance-based face descriptors have been used for both facial expression recognition and facial AU recognition⁴.

Geometric features describe changes in the shape and geometrical properties of different facial features, for example, stretched lips, raised eyebrows, etc. Feature extraction methods that use geometric features, first localize (in the case of static images) or track (in the case of image sequences or video streams) the positions of facial fiducial points that describe the position and shape of facial features such as, eyes, nose, eyebrows, lips, etc. Then, they compute the relative positions of these points, the ratio of distances between these points, the changes in the positions and distances across frames, etc. These features are used as parameters for rule-based approaches, or as inputs to various machine learning methods, to recognize the corresponding facial expression or facial AU.

Facial expressions cause appearance changes on the face, for example, wrinkles appear on the nose when expressing disgust, folds appear in the mouth region when smiling, etc. Features describing the facial texture or facial appearance have, therefore, been used for facial expression recognition. **Appearance-based feature** extraction methods extract information about textural patterns, such as, wrinkles, furrows, bulges, folds, edges, etc. from the entire face region or its sub-regions. Descriptors used for classification of textures have also been used to describe the appearance of faces. Examples of appearance-based face descriptors, include, Gabor filters, local binary patterns (LBP), local directional patterns (LDP), local ternary patterns (LTP), local phase quantization (LPQ), etc.

While geometric features were found to perform poorly compared to Gabor wavelet coefficients for the recognition of prototypic expressions [131], geometric features have been found to perform as good as appearance-based features for facial action unit recognition [116]. However, geometric features are difficult to extract at lower resolutions [109], making them unreliable for use on low resolution images and videos. Appearance-based feature descriptors, such as Gabor wavelets, LBP and their variants, and LDP and their variants, are available even at low image resolutions [109, 105, 53, 57, 56]. However, when using appearance-based face descriptors for facial expression or facial AU recognition tasks, the alignment of estimated face regions in input images and frames is crucial [109].

⁴See section 2.2.3 for a discussion on face descriptors.

2.2.2.3 Classification of Facial Expression

The next step in facial expression recognition is the classification of the extracted features into appropriate emotion or action unit categories. A number of different machine learning models and algorithms have been used for this purpose. For example, Multi-Layer Perceptrons (MLPs), Hidden Markov Models (HMMs), Dynamic Bayesian Networks (DBN), Principal Component Analysis (PCA), Independent Component Analysis (ICA), Linear Discriminant Analysis (LDA), Template Matching (TM), Support Vector Machines (SVM), Linear Programming (LP), AdaBoost, etc. SVMs have been shown to perform better than TM or LDA for facial expression recognition [65, 105, 53, 57]. HMMs and DBNs are used in approaches based on image sequences that exploit the spatial and temporal characteristics of facial expressions or actions (for example, [111, 61]).

Sometimes, a feature selection step precedes the classification step. During the training phase, the feature selection step selects the top ‘N’ feature dimensions that contain maximum information to distinguish between the different classes. Only the values of these selected feature dimensions are given as input to the classifier during the training and classification phases. The feature selection step reduces the dimensionality of the feature vector and thereby speeds up the training and classification phases. PCA and AdaBoost have been commonly used for feature selection. However, Littlewort et al. [65] found that AdaBoost was a better feature selection method than PCA, for facial expression recognition. The use of AdaBoost for feature selection and SVM for classification was found to yield better facial expression recognition rates than SVM or AdaBoost used alone with Gabor filters [65] or LBP [105].

2.2.3 Face Descriptors

Face descriptors describe the shape and/or appearance of the face and facial features⁵. Different face descriptors have been used to extract features from images and videos, for use in facial analysis tasks, such as, face recognition, facial expression recognition, facial action unit recognition, age estimation, gender classification, etc. While geometric face descriptors describe face and facial feature geometry, appearance-based face descriptors capture the textural details of the face, such as, wrinkles, furrows, edges, bulges, etc. In this section, two of the widely used appearance-based face descriptors, namely Gabor filters and Local Binary Patterns, are discussed⁶.

⁵There are two types of facial features [110]- *permanent facial features*, such as, the eyes, eyebrows, nose, mouth, ageing-related facial wrinkles and folds, etc., and *transient facial features*, such as, the wrinkles and folds that appear on the face only when displaying emotions, and that vanish when displaying a neutral expression.

⁶Gabor filters and LBP have been used in the experiments performed in this work.

2.2.3.1 Gabor Filters

Gabor filters were introduced by Dennis Gabor in 1946 [45]. A **Gabor filter** is obtained by modulating a complex sinusoidal wave with a Gaussian envelope. A Gabor filter kernel is characterised by the spatial frequency (a.k.a scale), orientation and phase shift of the underlying sinusoidal wave, and by the width and ellipticity of the Gaussian envelope [74]. 2D Gabor filters have been used in image processing to extract spatially localized textural features that match the scale and orientation of the filter kernel. The image is first convolved with the Gabor filter to obtain the Gabor features in complex form. The real and imaginary parts of the Gabor features are then combined to generate a Gabor magnitude representation, which is later used for analysis. The resolution of the Gabor magnitude image is identical to the input image.

2D Gabor filters have been extensively used in facial expression analysis (for example, [131, 71, 65, 132, 19, 21, 111]). Usually, a bank of Gabor filters of different scales and orientations are chosen. Each Gabor filter is applied to the input facial image and the magnitudes of the Gabor features are computed. The Gabor magnitude representations obtained using all the filters are concatenated to obtain the feature vector that is later used as input for a classifier. Each element in this feature vector is also called **Gabor wavelet coefficient**. Gabor feature vectors thus formed have very high dimensionality, making facial expression analysis based on Gabor filters memory- and time-intensive [105]. However, in certain works, such as, [131, 71], only the Gabor features at certain locations or regions are considered, which greatly reduces the dimensionality of the feature vector.

A mathematical description of Gabor filters, and a pictorial representation of Gabor filters and Gabor magnitude images have been provided in [74].

2.2.3.2 Local Binary Patterns

In 1996, Ojala et al. [87] proposed **Local Binary Patterns** (LBP) as a two-level version of Wang and He's [123] texture descriptor. For each 3x3 grid of pixels in the grayscale image, a binary pattern is computed and the equivalent decimal code is assigned to the center pixel. In order to compute the binary pattern, the intensity value of each of the eight border pixels in the grid is compared to the intensity value of the center pixel. If the former is greater than the latter, then '1' is assigned to the corresponding border pixel. Otherwise, '0' is assigned to it. On assigning '0' or '1' to the eight border pixels, we get an 8-bit binary pattern, which is then converted into a decimal number and assigned to the pixel at the center of the grid. The decimal numbers so obtained are called **LBP codes**.

For a 3x3 neighbourhood, $2^8 = 256$ different binary patterns and LBP codes are possible. LBP is grayscale-invariant [87] and immune to monotonic illumination variations [97].

Later on, a rotation-invariant version of LBP was proposed by Ojala [85, 97]. All patterns obtained by rotating an 8-bit binary pattern in clockwise or counter-clockwise direction are grouped together and represented by the binary pattern having the lowest decimal value in the set. For a 3x3 neighbourhood, there exists 36 such groups, and consequently 36 distinct patterns. Each of these patterns are arranged in ascending order of their decimal values and assigned an index or label ranging from 0 to 35. A **rotation-invariant LBP** operator, denoted by LBP^{ri} , maps an 8-bit binary pattern to the index range [0, 35].

A 3x3 grid cannot capture large-scale textural structures. Ojala et al. [86] later extended LBP to circular neighbourhoods. A circular neighbourhood (P, R) is defined by the radius (R) and the number of samples (P) to be considered. For example, an (8,2) circular neighbourhood is formed by 8 pixels located on a circle having radius equal to 2 pixel-width (see [86, 74] for a pictorial representation). The computation of the binary pattern and decimal code for the center pixel are identical to conventional LBP. The extended operator based on circular neighbourhoods is called **extended LBP**, and is denoted by $LBP_{P,R}$. A rotation-invariant version of eLBP, denoted by $LBP_{P,R}^{ri}$, is obtained in a similar fashion as the rotation-invariant version of conventional LBP. For a neighbourhood consisting of 8 samples, there are 36 distinct rotation-invariant binary patterns. $LBP_{8,R}^{ri}$ maps the local binary patterns to the index range [0, 35].

Ojala et al. [86] found that certain local binary patterns occur more frequently than others, and constitute more than 90% of all conventional LBP patterns⁷ extracted from texture images. These patterns were found to contain very few bit changes from 1 to 0 or 0 to 1, and represented bright or dark spots, edges of different curvatures, etc. The binary patterns containing at most two bit changes were defined as **uniform patterns** [86]. For example, 00011100 is a uniform pattern that contains two bit transitions, 00001111 is a uniform pattern that contains only one bit transition, and 00000000 is a uniform pattern that contains zero bit transitions. For a neighbourhood of 8 samples, there are 58 uniform binary patterns and 9 rotation-invariant uniform binary patterns. $LBP_{P,R}^{u2}$ and $LBP_{P,R}^{riu2}$ are LBP operators that are based on uniform binary patterns and rotation-invariant uniform binary patterns, respectively, for a given neighbourhood (P,R). These operators group all non-uniform binary patterns into a single set. Therefore, $LBP_{8,R}^{u2}$ operator maps local binary patterns to the index range [0, 58] and $LBP_{8,R}^{riu2}$ operator maps them to the index range [0, 9].

⁷for a 3x3 neighbourhood

There exists numerous other variants of LBP. In [50], Huang et al. have surveyed the different variants of LBP, and their use in facial image analysis tasks such as face recognition, facial expression recognition, gender classification, head-pose estimation, age and ethnicity classification, etc. Extraction of LBP features is much less memory- and time-intensive than extraction of Gabor wavelet coefficients [105]. LBP has also been found to yield slightly better performance than Gabor filters in facial expression recognition [105].

2.2.3.3 Others

Local Directional Patterns (LDP), an appearance-based face descriptor proposed by Jabid et al. [52], use local directional responses computed using Kirsch masks to extract textural features. It is robust to non-monotonic illumination variations, unlike LBP. LDP and its variants have been found to perform better than LBP in face recognition [52], facial expression recognition [53, 57], and gender classification [54]. Other face descriptors which have been used in facial analysis include Local Ternary Patterns (LTP), Local Phase Quantization (LPQ), scale-invariant feature transforms (SIFT), etc.

LPQ was proposed as a texture descriptor that is robust to blurring by Ojansivu and Heikkilä in 2008 [88]. They found that LPQ performed better than LBP and Gabor filters in classifying blurred as well as non-blurred texture images. For facial AU recognition, LPQ was found to perform better than LBP on still images, and LPQ from Three Orthogonal Planes (LPQ-TOP) was found to perform better than LBP from Three Orthogonal Planes (LBP-TOP) on image sequences [55].

Combinations of various appearance-based face descriptors have also been used for facial image analysis, and these have been found to perform better than the individual descriptors. For example, a combination of Gabor filters and LBP has been used in face recognition [130], head pose estimation [73] and facial action unit recognition [124]. A similar combination of Gabor filters and LDP has been used for age and gender classification [49]. Appearance-based descriptors have also been used together with geometric descriptors for facial expression analysis [131]. The face descriptors thus obtained are known as **hybrid face descriptors**.

2.3 Facial Expression Databases

A survey and comparison of the databases that are commonly used in automatic facial expression recognition research can be found in [128], [15] and [23]. [46] is an older survey, published in the year 2005, of face databases that have been used in various face-related

research such as face recognition, face detection and facial expression analysis. In the following sub-sections, six of the databases used in facial expression analysis are discussed in detail. The section is closed with a discussion of the open issues and challenges.

2.3.1 Japanese Female Facial Expressions Database

The Japanese Female Facial Expressions (JAFPE) database [70, 72] is a small database of frontal images showing 10 Japanese females posing the 6 basic expressions and the neutral expression. There are 213 images in the database. 2 to 4 examples of each expression posed by each expresser is included in the database. The age-group of the models is not mentioned, but appear to belong to the young age-group. The images are provided in the .tiff format and have a resolution of 256×256 . The images were taken under controlled lighting conditions. The illumination conditions and the apparatus used to capture the images are discussed in [70]. The database is available for download at [7].

Psychological experiments were conducted using the JAFPE database to establish the ground truth [70]. 60 Japanese female students were asked to rate the degree to which the 6 basic expressions are present in each of the images in the database [3]. This was based on the reasoning that expressions are not pure, but contain traces of other expressions. The ratings were done on a scale of 1 to 5 and the average ratings are available at [3]. The experiment was repeated by excluding the images labelled as “fear”. However, only 30 female students participated in this experiment and the ratings were done for the presence of 5 basic expressions (excluding fear). The average ratings are available at [3].

The JAFPE database has been used in a number of facial expression analysis studies, and recognition rates of upto 95.71% have been reported [107].

2.3.2 Pictures of Facial Affect

The Pictures of Facial Affect (POFA) database [39], compiled by Ekman and Friesen, contains 110 black-and-white frontal images showing expressers displaying the 6 basic expressions and the neutral expression [6]. Semantic ratings obtained through psychological experiments conducted using these pictures are also available [132]. Although information on the age-group of the expressers could not be found, the demo images available at [6] reveal that at least a few of the expressers possibly belong to the middle-aged group. The database must be purchased, and currently costs \$175.00. Recognition rates of 81-84% have been reported on the Pictures of Facial Affect database by studies, such as, [71], [91] and [132].

2.3.3 MMI Facial Expression Database

The MMI⁸ facial expression database [96, 118] database was developed by Pantic et al. at the Delft University of Technology in the Netherlands. It contains both images and videos of persons displaying single AU, combinations of AUs and basic facial expressions [96]. The database is organized into five parts [118]. Part I contains 1767 video clips showing 20 participants expressing different AUs, action descriptors and affective states such as sleepy, happy and bored, in both frontal and profile views. Part II contains 238 video clips showing frontal and profile views of 28 participants displaying the six basic expressions of emotions. Part III contains 484 static images showing 5 subjects displaying the AUs and the six basic facial expressions in frontal and profile views. Part IV contains video clips of spontaneous expressions of happiness, disgust and surprise, elicited from 25 participants using audio-visual stimuli. The video clips have been annotated by FACS experts to indicate the AUs, and wherever applicable, the basic expressions of emotions. Part V contains annotated audio-video clips of voiced and unvoiced laughter. Some of the video clips and images in the database show participants wearing glasses and/or having facial hair. The age of the participants ranges from 19 to 62 years [96]. However, the age-group-wise categorization and labelling of images and videos have not been done. The age distribution is also not available.

The MMI facial expression database has been used in research on automatic facial AU recognition, and analysis of temporal characteristics of facial expressions ([93, 117, 114, 55, 115]). Access to the database is free for academic research purposes, and is obtained after the signing and submission of an end-user license agreement. The agreement is available online at [4].

2.3.4 Facial Expressions and Emotion Database

The Facial Expressions and Emotion Database (FEED) [122] was developed at the Technical University Munich, as part of the Face and Gesture Recognition Research Network (FG-NET) project of the European Union. It contains image sequences of spontaneous expressions of emotions elicited through audio-visual stimuli. The image sequences start with a neutral expression, progress towards a peak facial expression of emotion, and end with a neutral expression. Head movements were allowed. The database includes the six basic expressions of emotions displayed by 19 participants, with three samples per expression per participant. There are also three samples per participant, in which, the participant displays a neutral face. Thus, there are 399 image sequences in all.

⁸originally stood for M&M's Initiative, where the Ms are the initials of Pantic and Valstar [96]

The image sequences were captured under identical illumination settings, and have the same background. The images are in the JPEG format and have a resolution of 320 x 240 pixels. The first frame of each image sequence is labelled with the positions of the eyes, the nose and the mouth. Metadata about the image sequences include the emotion expressed, the person counter, the sequence counter, and the indices of the start, apex, hold and end frames [79]. The database is usable free-of-charge for research purposes, upon request and submission of a form available online at [122].

2.3.5 Extended Cohn-Kanade Dataset

The Extended Cohn-Kanade Dataset (CK+) [68] is an extension of the Cohn-Kanade (CK) database [58], and was released in the year 2010. It contains 593 image sequences, in which, 123 subjects display single AUs or combinations of AUs. Each image sequence starts with a neutral expression and ends with the peak display of the AU(s). The length of each sequence varies between 10 and 60 frames. The peak frames are FACS coded by FACS experts. Out of the 593 image sequences, 327 were selected through a multi-step process, and labelled with one of the 7 expression labels, namely, anger, contempt, disgust, fear, happiness, sadness and surprise. These 327 image sequences are from 118 human subjects. The age of the human subjects posing the AU(s) range between 18 and 50 years. However, no systematic age-group-wise categorization and labelling of image sequences have been done on the CK+ database.

The key differences between CK+ and CK databases include addition of more image sequences, inclusion of more human subjects, establishment of ground truth for discrete expressions, and inclusion of non-posed smiles of different types [13]. Baselineing of automatic AU recognition and expression recognition on CK+ database has been done using active appearance model (AAM) and SVM. The results are reported in [68]. The benchmarking of the evaluation protocol and performance metric has also been done to facilitate comparability of performance of state-of-the-art methods in AU recognition and expression recognition. The leave-one-subject-out (LOSO) scheme for cross-validation, and the area under the receiver-operator characteristic (ROC) curve as performance metric, have been recommended for reporting results of AU recognition on the CK+ database. LOSO for cross-validation, and the confusion matrix as the performance metric, have been recommended for reporting the results of facial expression recognition on the CK+ database. Comparison of the performance of automatic facial expression recognition systems with the performance of human observers has also been recommended by the authors.

The CK database has been widely used in research on facial expression analysis. Recognition rates above 90% have been reported on the CK database by various researchers [128, 23]. However, it was noted that researchers selected different sets of images from the CK database for training and testing facial expression recognition systems [68]. They used different cross-validation schemes and different performance metrics. These, together with the fact that validated emotion labels are not distributed with the CK database, affect the comparability of the results across different research works. Measures have been taken to address these issues in the new CK+ distribution, as described in the previous paragraph.

Information about CK and CK+ databases, and the database download procedure are available online at [1]. The databases are available free-of-charge for research or non-commercial use.

2.3.6 FACES

The FACES database [36], from the Max Planck Institute for Human Development, Berlin, Germany, contains 2052 frontal images, posed by face models belonging to three different age-groups, namely, young, middle-aged and older. It was created to meet the need for a database of images in which both the age of the expressers and the displayed expressions are systematically varied. The FACES database includes images of 171 face models, each one posing 6 expressions, namely, anger, disgust, fear, joy, sadness and neutral. Of the 171 face models, 58 belonged to the young age-group (19-31 years), 56 to the middle-aged group (39-55 years) and 57 to the older age-group (69-80 years). The mean age of the three age-groups are 24.2 years, 49 years and 73.2 years, respectively. Roughly half of the face models were females. The ethnicity of all the face models is Caucasian.

The face models were given training on how to pose the expressions. Pictures and vidoes were also used to induce emotions and appropriate facial expressions of those emotions. The method and equipments used to capture the images, the method used to select images for the final database, and the editing and standardization performed on the images are elaborated in [36]. The database contains two images per face model per expression, which were selected based on the ratings given by trained raters. These images are organized into two sets A and B. Each image in the database is labelled with the code assigned to the face model, the age-group and gender of the face model, the posed expression, and the picture set A or B, to which it belongs.

The final FACES database was validated by another group of 154 human subjects, each

belonging either to the young (20-31 years), middle-aged (44-55 years) or older (70-81 years) age-groups [36]. Of the 154 raters, 52 belonged to the young age-group, and 51 each to the middle-aged and older age-groups. All the subjects were Caucasian. A summary of the overall recognition rates for each facial expression, averaged over all subjects, is provided in table 1. It can be seen that the recognition rate was highest for “joy” and lowest for “disgust”.

Facial Expression	Average Recognition Rate (%)
Anger	81
Disgust	68
Fear	81
Joy	96
Neutral	87
Sadness	73

Table 1: Recognition rate achieved for each facial expression during the validation of FACES database by human raters [36].

Appendix A in [36] provides the overall recognition rate for each expression, for different age and gender categories of human raters and face models. These are summarized in tables 2 and 3. Table 2 reveals the effect of the age of the face model on the recognition of each facial expression. The recognition rates shown in the table have been averaged over all human raters. The authors found that the recognition rate for all expressions except

Face model age-group	Anger	Disgust	Fear	Joy	Neutral	Sadness	Overall
Young	85.58%	73.92%	80.50%	97.08%	92.08%	78.83%	84.67%
Middle-aged	84.25%	68.67%	81.08%	96.75%	86.83%	72.08%	81.61%
Older	73.75%	60.58%	78.92%	95.50%	81.00%	67.00%	76.13%
Overall	81.19%	67.72%	80.17%	96.44%	86.64%	72.64%	80.80%

Table 2: Overall recognition rates achieved by human raters for each facial expression in the FACES database for different age categories of face models. (Computed based on the data published in [36].)

“fear” was low for the older age-group compared to the young and middle-aged groups [36]. In the case of “fear”, no significant variation was found in the recognition rates across age-groups of face models. When compared to the young age-group, the recognition rate for the middle-aged group showed significant drop for the expressions disgust, neutral and

sadness.

Table 3 reveals the facial expression recognition performance for different age-groups of human raters. The authors found that young raters performed better than older raters, especially in recognizing anger, disgust and sadness [36]. The young raters also performed much better than the middle-aged group in recognizing sadness. The performance of human raters were more or less identical for the expressions of joy, fear and neutral.

Young raters							
Face model age-group	Anger	Disgust	Fear	Joy	Neutral	Sadness	Overall
Young	90.25%	79.50%	85.25%	97.75%	93.50%	82.25%	88.08%
Middle-aged	86.75%	76.50%	85.00%	97.00%	88.00%	78.50%	85.29%
Older	79.25%	66.25%	82.75%	95.75%	82.75%	74.25%	80.17%
Overall	85.42%	74.08%	84.33%	96.83%	88.08%	78.33%	84.51%
Middle-aged raters							
Face model age-group	Anger	Disgust	Fear	Joy	Neutral	Sadness	Overall
Young	89.00%	73.50%	79.75%	98.25%	92.75%	78.00%	85.21%
Middle-aged	89.25%	67.75%	81.50%	98.50%	88.25%	69.75%	82.50%
Older	76.50%	60.00%	78.50%	97.50%	81.25%	63.75%	76.25%
Overall	84.92%	67.08%	79.92%	98.08%	87.42%	70.50%	81.32%
Older raters							
Face model age-group	Anger	Disgust	Fear	Joy	Neutral	Sadness	Overall
Young	77.50%	68.75%	76.50%	95.25%	90.00%	76.25%	80.71%
Middle-aged	76.75%	61.75%	76.75%	94.75%	84.25%	68.00%	77.04%
Older	65.50%	55.50%	75.50%	93.25%	79.00%	63.00%	71.96%
Overall	73.25%	62.00%	76.25%	94.42%	84.42%	69.08%	76.57%

Table 3: Recognition rates for each facial expression in the FACES database, for different age categories of human raters and face models. (Computed based on the data published in [36].)

The images in the FACES database are distributed in three different resolutions, namely,

111×139 , 335×419 and 2835×3543 pixels. Registration is required to get access to the database. The registration procedure, download instructions and release agreement are available online at [2].

The FACES database has been mainly used in behavioural, psychological and developmental research on emotion. It was used in computer vision-based studies for the first time by Guo and Wang [47], for estimating the age of a person under different facial expressions.

2.3.7 Others

Apart from the six facial expression databases discussed in the earlier sub-sections, there are also a number of other databases that are used in facial expression recognition research. While some of these are publicly available, some are not. While some contain posed expressions, some contain spontaneous expressions. The database developed by Sebe et al. in [103] contains spontaneous expressions, but is not public. The BU-3DFE database [127] contains 3D posed facial expressions, and is publicly accessible. The AR database [77] is yet another publicly available database that contains posed expressions with facial occlusion and variations in illumination conditions. For more information on facial expression databases, the surveys [128], [15] and [23] can be referred.

2.3.8 Open Issues and Challenges

There are many aspects associated with images and videos of human facial expressions of emotion, variations in which affect the performance of facial expression recognition systems. Some of the variations that are encountered are listed below:

- Differences in face and facial feature characteristics such as size and shape of head, eyes and mouth, skin texture, skin color, etc. caused by variations in age, gender, ethnicity and personal history.
- Facial occlusions caused by hairdo, glasses, beard, moustache, jewellery such as nose piercings, head and neck scarfs, etc.
- Variations in head pose
- Variations in intensity of displayed expressions
- Differences between posed and spontaneous expressions
- Variations in resolution and quality of images and image sequences
- Variations in illumination conditions

There is no single database that systematically varies all the afore-mentioned aspects over the six basic emotions, and 44 facial action units and their combinations. The development of such a benchmark database would enable performance benchmarking and comparison of facial expression recognition approaches [23]. In order to develop such a database, the existing databases should be fully validated, filtered, systematized, and extended. Image sequences in the database should contain the three temporal phases, namely, onset, apex and offset. Images and image sequences should be FACS coded and also labeled with prototypic expressions, wherever applicable. The development of such a database can be realized only through the concerted effort of research labs across the world.

However, there are a number of challenges in developing such a benchmark database. The difficulty in capturing spontaneous expressions of emotions is one of the challenges. This has been discussed by Sebe et al. in [103]. They observed that spontaneous expressions are more subtle and have interpersonal variations in intensity. Some emotions, such as, fear and sadness, were found to be difficult to elicit. Systematic variation of other parameters, such as, age, illumination conditions, etc. over spontaneous expressions is also difficult to realize.

Labeling images and videos in the database with FACS codes and emotion labels requires considerable human effort. Manual FACS coding of a one-minute video clip or 100 facial images takes about one hour [55]. The process is also expensive and requires the coders to be trained in FACS. Database validation and establishment of ground truth by measuring performance of untrained human observers is also time-consuming.

A discussion of some of the aspects of the problem space for facial expression database is available in [58].

2.4 Neurological, Psychological and Behavioural Research

In this section, some of the neurological, psychological and behavioural research findings on human performance in face and facial expression recognition are discussed.

Adolphs [10] found that humans easily confuse fear with surprise. Palermo and Coltheart [92] studied the effect of intensity of expressions on the recognition rate. They found that the expression of “joy” could be identified correctly even at low intensity. The accuracy in recognizing “anger”, “disgust” and “sadness” increased as the intensity with which it was expressed increased. Calvo and Lundqvist [28] measured the time taken to recognize facial expressions. They found that identification of “joy” took the least time, and identification

of “fear” took the most time. Irrespective of the age-group of the expresser, happiness or joy is easier for humans to identify, and disgust is more difficult to identify than any other expression [36]. The accuracy of facial expression recognition was higher on young faces than on middle-aged and older faces [36]. The accuracy was the lowest on older faces [36].

Ruffman et al. [100] studied the age-related differences in the recognition of emotions expressed through different modalities. Older adults were found to have significant difficulty in recognizing anger, fear and sadness conveyed through facial expressions, when compared to young adults. The recognition rate for the facial expressions of joy and surprise, were also lower than young adults, although the difference was small. The observed age differences in emotion recognition were attributed to the neuropsychological changes associated with adult ageing.

Multiple parts of the brain have been found to be involved in emotion recognition. Some of these are the amygdala, the fusiform gyrus, the occipitotemporal cortices, the orbitofrontal cortex, basal ganglia, insula and right parietal cortices [10, 100]. The amygdala has been found to be crucial in recognizing the facial expression of fear, as well as in identifying blended emotions in facial expressions [11]. A review of the neural structures involved in multi-modal human emotion recognition is provided by Ruffman et al. in [100]. The sensory modalities for communicating emotions include vision, audition, olfaction and touch [10]. Visual cues of emotions include facial expressions, body posture and gestures [94, 26]. Auditory cues include speech and vocal intonation [10, 94].

Mather, Carstensen and Charles [78, 32] found that older adults remembered and recognized positive faces better than negative faces. Isaacowitz et al. [51] found that older adults paid less attention to negative faces than to happy and neutral faces. An “own-age bias” in face recognition has been evidenced [14]. An “own-age bias” was also found in the recognition of facial expressions [76]. People were found to perform better in recognizing facial expressions displayed by persons belonging to the same age-group, which could be due to the increased social contact with the same age-group [76].

Hess et al. [48] recently studied the impact of wrinkles and folds, which appear on the human face due to the process of ageing, on the human perception of emotion and behavioural intent. They conducted experiments involving young subjects, who were asked to rate the facial expressions of emotions on young and older faces. The expressions of joy, sadness and anger were morphed on to the neutral faces of young and old persons using FaceGenModeller [8] software. This was done to ensure that the expressions on the faces had identical characteristics, for example, intensity. The young subjects, who were

selected to participate in the study, were asked to rate, on a scale of 1 to 6, the intensity of expressions that they recognized on the faces shown. They were also asked to rate the intensity of behavioural intentions, namely, dominance and affiliation (or sociability), on scales of 1 to 7.

From the experiments, Hess et al. [48] found that facial expressions, when expressed by older persons, were perceived by young persons to be less intense than the same expressions on young faces. In addition, other (inaccurate) emotions were perceived to be present on the older faces at greater intensities than on young faces. When neutral faces were shown, the young subjects perceived more non-neutral emotions on the older faces than on young faces. This replicated the finding of Malatesta et al. [75] in 1987, and ***showed that wrinkles and folds resemble facial expressions of emotions***. The second finding was that older faces were rated as less dominant and less sociable than young faces, for all the expressions displayed. The results, put together, showed that ***wrinkles and folds on the faces due to ageing not only reduce the clarity of emotions, but also affect the behavioural intentions communicated through facial expressions***, which, in turn, negatively affect the social interactions between young and older individuals.

3 Related Work

In this section, some of the important works in the field of computer vision-based facial expression recognition are discussed. A brief overview of facial action unit recognition and multimodal emotion recognition systems is also provided. The state-of-the art surveys [94] and [44] summarize the research on automatic facial expression recognition and facial action unit recognition until a decade ago. [128] and [23] (still in pre-print stage) are the more recent surveys that cover the developments in the past decade.

Table 4 presents the classifications of the approaches for facial expression and facial action unit recognition. Depending on the type of descriptors used to extract facial expression information, the approaches for facial expression and facial action unit recognition can be categorized into geometric feature-based and appearance-based approaches [61]. In some works, a combination of geometric and appearance-based features is used, and these can be categorized as hybrid feature-based approaches. Depending on whether the temporal characteristics, such as timing, length of onset, apex and offset phases, etc., of expressions are considered along with the spatial characteristics, the approaches can be categorized into spatial and spatio-temporal approaches. Depending on whether static images or image sequences are used, the approaches can be categorized into static and dynamic approaches

[33].

Criteria	Classification
Feature type [61]	geometric, appearance-based, hybrid
Expression characteristics	spatial, spatio-temporal
Input type [33]	static, dynamic

Table 4: Classification of approaches for facial expression analysis.

3.1 Facial Expression Recognition

Facial expression recognition from static images is the primary focus of this work. In this section, some of the important works in this area are discussed, so as to provide the reader an overview of the different approaches that have been taken to recognize the six prototypic facial expressions of emotions, namely, anger, disgust, fear, joy, sadness and surprise. In most of the works, the neutral expression has also been considered.

In [131], Zhang et al. identified 34 fiducial points on the 213 images in the JAFFE database, and extracted complex Gabor wavelet coefficients at these points using 18 different kernels. 2-layer perceptrons were trained using the fiducial points and the magnitudes of the complex Gabor wavelet coefficients independently, as well as together. Resilient propagation learning algorithm was used to train the 2-layer perceptrons, and 10-fold cross-validation scheme was used to evaluate the performance. For each case, the training and evaluation were repeated 10 times to account for the influence of the randomly chosen initial weights. Gabor wavelet coefficients outperformed the geometric positions. A recognition rate of 90.1% was obtained with magnitudes of Gabor wavelet coefficients using 7 hidden layer neurons. With fear excluded, the recognition rate was 92.2%, against the 73.3% with geometric positions alone. The combination of the two feature descriptors improved the performance only when the number of hidden layer units was less than 5.

In [71], Lyons et al. describe an approach to extract facial expression information based on Gabor filters, Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA). First, Gabor filters are applied to the image under consideration. Then, a fiducial grid is superimposed and the Gabor wavelet coefficients on the grid are sampled. The amplitudes of these sampled coefficients constitute the Labeled-Graph (LG) vector. Dimensionality reduction of the LG vectors computed for all images in the training set is then performed using PCA. The set of resulting vectors are then clustered by performing 2-class LDA. A separate discriminant vector is computed for each of the six basic emotions. The discriminant vectors discriminate one expression from all the rest (for example, joy v/s

other). To classify new input vectors, they are projected along each of the six discriminant vectors computed during the training phase. The distances to centers of clusters are computed and normalized for each case. The nearest cluster is identified from each of the six binary classifications. If two or more expressions are identified as being present, then the closest cluster is chosen. If none of the expressions are identified as being present, then the input is classified as belonging to the “neutral” category. The approach was tested on the JAFFE database. A recognition rate of 92% was obtained using 10-fold cross-validation. This is similar to the recognition rates reported in [131]. An average recognition rate of 75% was obtained over expresser identity using the leave-one-subject-out cross-validation strategy.

Zheng et al. [132] also extracted the LG vector, in the same way as described in [131] and [71]. The JAFFE and POFA databases were used for the study. Images showing neutral expression were excluded. Semantic ratings are available for all the images in these databases. These ratings were compiled through psychological experiments, where subjects were asked to rate the presence of each of the prototypic expressions in the images. A semantic vector was constructed for each of the training images based on these semantic ratings. The correlation between the LG vector and the corresponding normalized semantic vector was learned using Kernel Canonical Correlation Analysis (KCCA). During the testing phase, the semantic vector was estimated and the most prominent basic expression was identified. Leave-one-image-out (LOIO) and leave-one-subject-out (LOSO) cross-validation strategies were employed for testing. Best recognition rates were obtained when using KCCA with a Gaussian kernel. For the JAFFE database, LOIO cross-validation produced a recognition rate of 85.79% and LOSO cross-validation produced a recognition rate of 74.32%. For the POFA database, the recognition rates were 81.25% and 79.17% respectively. When using the expression class label, instead of the semantic rating, to construct the semantic vector, the best recognition rate (98.36%) on the JAFFE database was obtained using KCCA with degree-2 polynomial kernel and LOIO cross-validation scheme. The LOSO scheme yielded a recognition rate of 77.05% using KCCA with a Gaussian kernel. For the POFA database, the recognition rate was 77.08% using KCCA with Gaussian kernel (for both the cross-validation schemes).

In [107], Shih et al. have evaluated the performance of different feature extraction and classification methods for expression recognition on the JAFFE database, and compared the performance with that of the existing systems. The feature extraction methods that were studied include Discrete Wavelet Transform (DWT), together with Independent Component Analysis (ICA), PCA, 2D-PCA, LDA or 2D-LDA. The studied classifiers include Radial Basis Function Network (RBFN) and Support Vector Machine (SVM). Different

kernel functions were chosen for SVM, namely, linear, polynomial (degrees 2, 3, 4) and radial basis functions. The experimental procedure consisted of three main steps. First, the images in the JAFFE database were pre-processed to remove the background and to neutralize the illumination effects (using histogram equalization). This was followed by feature extraction and classification steps. The best performance was obtained when using DWT together with 2D-LDA for feature extraction, and tree-based one-against-one linear SVM for classification. The 10-fold cross-validation yielded a recognition rate of 94.13%. The study also revealed that RBFN yielded poor recognition rates (as low as 25.24%).

In [105], Shan et al. empirically evaluated the performance of Local Binary Patterns (LBP) as face descriptors for automatic facial expression recognition. Various machine learning methods, such as, template matching, SVM, LDA and linear programming (LP), were examined, and different databases, such as, MMI, JAFFE and Cohn-Kanade database, were used in the study. A description of the experimental setup and the performance of the different combinations of features and machine learning methods are given in table 5. Each method was evaluated for two types of recognition tasks- one in which all 7 expressions had to be recognized, and the other in which only the 6 basic expressions had to be recognized. As seen from table 5, for all the evaluated methods, the generalization performance for the 6-class recognition task was higher than that for the 7-class recognition task. The LBP + SVM combination performed slightly better than the Gabor + SVM combination for all the three types of SVM kernels considered. Among all the methods, the LBP + SVM with RBF kernel yielded the best overall recognition rate for both the 7-class as well as 6-class recognition tasks.

Shan et al. [105] also evaluated the performance of $LBP_{4,1}$ operator on images of different resolutions. The evaluation was performed for the 6-class recognition task on the Cohn-Kanade database, using SVM with RBF kernel as the classifier. Six resolutions were examined- 110×150 , 55×75 , 36×48 , 27×37 , 18×24 and 14×19 . The performance was compared with that of Gabor wavelet coefficients extracted using a bank of 40 Gabor filters having 8 different orientations and 5 different scales. The performances of these systems were compared with the results published by Tian in [109]⁹. The comparison of the different systems revealed that, since facial component localization was difficult at low resolutions, geometric features are unreliable in such cases. However, the appearance-based feature descriptors, namely, Gabor and LBP, performed reliably even at low resolutions. Among these, LBP features yielded slightly better performance than Gabor wavelet co-

⁹Tian had examined the performance of geometric features as well as Gabor wavelet coefficients on images of different resolutions, using neural networks as the classifier for the 6-class recognition task. The images had been taken from the Cohn-Kanade database.

Operator	Classifier	7-class recognition rate (in %)	6-class recognition rate (in %)
LBP	Template Matching	79.1 ^a	84.5
LBP	SVM (linear kernel)	88.1	91.5
Gabor ^b	SVM (linear kernel)	86.6	89.4
LBP	SVM (polynomial ^c kernel)	88.1	91.5
Gabor	SVM (polynomial kernel)	86.6	89.4
LBP	SVM (RBF ^d kernel)	88.9 ^e	92.6 ^f
Gabor	SVM (RBF kernel)	86.8	89.8
LBP + PCA	LDA + NN ^g	73.4	79.2
LBP + PCA	SVM (linear kernel)	80.2	87.7
LBP ^h	LP	82.3	89.6
LBP ^h	SVM (linear kernel)	86.0	90.4

^asurprise (92.4), joy (90.4), disgust (85), sadness (72.4), neutral (70.3), fear (61.7), anger (58.7)

^bGabor magnitude images were created using a bank of 40 Gabor filters of 8 different orientations and 5 different scales, and each convolved image was downsampled by factor 16.

^cdegree = 1

^d $\sigma = 2^{13}$ for 7-class; $\sigma = 2^{11}$ for 6-class

^esurprise (98.2), joy (94.7), disgust (97.5), sadness (69.5), neutral (90), fear (68), anger (85)

^fsurprise (98.7), joy (97.9), disgust (97.5), sadness (83.5), fear (73), anger (89.7)

^gwith Euclidean distance measure

^hDimensions having occurrence frequency less than 5 were dropped.

Table 5: Results of empirical evaluation of LBP as face descriptor for facial expression recognition, published in [105]. Experimental settings: 1280 images were selected from the Cohn-Kanade database. Face regions of size 110×150 pixels were extracted based on manually-located eye positions. The 59-bin LBP_{8,2}^{u2} operator was used and the face region was divided into 6×7 blocks for constructing the histogram sequence. Generalization performance was evaluated using the 10-fold cross-validation scheme.

efficients for all the six resolutions examined. The evaluation results are summarized in table 6. The table shows that the recognition rate decreased as the resolution was lowered.

Shan et al. [105] also proposed the use of boosted LBP histograms for improving the performance of facial expression recognition. This was inspired by the work of Zhang et al. [129], who used AdaBoost to learn discriminative LBP-based histograms for face recognition. To avoid the dependence of LBP features on the size and location of blocks, shifting and scaling of a sub-window is suggested. In the experiments conducted, 16,640 LBP histograms were created by shifting and scaling a sub-window. AdaBoost, with histogram-based template matching as the weak classifier, was used to learn a few tens of the most discriminative histograms (or, equivalently, sub-regions) out of the 16,640 histograms extracted. One AdaBoost learner was trained for each expression using the one-against-rest technique. The discriminative sub-regions selected by AdaBoost revealed

Image resolution	With LBP features (%)	With Gabor wavelets (%)
110×150	92.6	89.8
55×75	89.9	89.2
36×48	87.3	86.4
27×37	84.3	83.0
18×24	79.6	78.2
14×19	76.9	75.1

Table 6: Recognition rates obtained using LBP features and Gabor wavelets for different image resolutions [105]. SVM with RBF kernel was used as the classifier, and the task was to recognize the 6 prototypic expressions. The images used for the evaluation were taken from the Cohn-Kanade database. The face regions of resolution 110×150 were extracted from the original images in the database, based on manually-identified eye positions. The images of lower resolutions were created through downsampling.

that the regions around the eyes and mouth mainly contained discriminative information about facial expressions. It was also found that, for each expression, a different set of sub-regions were selected as discriminative. For example, for the expression of disgust, the region around the nose was crucial. The experiments were conducted using Boosted-LBP features for facial expression recognition showed that Boosted-LBP features yielded better recognition rates than LBP. The recognition rates obtained using Boosted-LBP are listed in table 7.

Classifier	7-class recognition rate (in %)	6-class recognition rate (in %)
LDA	77.6	84.2
AdaBoost	85	89.8
SVM (linear kernel)	91.1	95.0
SVM (polynomial kernel)	91.1	95.0
SVM (RBF kernel)	91.4	95.1

Table 7: Results of evaluation of Boosted-LBP features for facial expression recognition, published in [105]. The recognition rates reported are based on 10-fold cross-validation.

Recently, Nagi et al. [82] proposed an approach for automatic facial expression recognition that considers only the regions of the face that contain discriminative information about facial expressions of emotion. The regions around the eyes, eyebrows, nose and mouth were considered. In the proposed approach, first of all, the face is detected in the image using Viola-Jones algorithm, and the detected face region is cropped and scaled to 120×120 pixels. To eliminate false positives, eye detection is done on the face re-

gion. Facial features, namely, eyes, eyebrows, nose and mouth, are then detected using Haar-like feature-based cascade classifiers. After extracting the different facial features, they are represented using the extended Local Binary Pattern (eLBP) feature descriptor. Only the 58 uniform patterns of the (8,1) neighbourhood are considered. A histogram is constructed for each facial feature. The histograms are then concatenated to form the feature vector. A one-against-rest SVM with Radial Basis Function (RBF) as the kernel is used for facial expression classification. The JAFEE and Cohn-Kanade databases were used for validating the approach. Recognition rates of 82.79% and 77.93% were obtained on the Cohn-Kanade and JAFFE databases respectively. Expressions showing neutrality and surprise were recognized more accurately on the Cohn-Kanade database, and expressions showing neutrality and joy were recognized more accurately on the JAFFE database. The approach was compared to the traditional approach (originally proposed in [12]) that divides the whole face region into blocks, computes histograms for each block, and concatenates the histograms to form the feature vector. The proposed approach was found to perform better and faster than the traditional approach on the Cohn-Kanade database. Experiments conducted to identify the most discriminative facial features, revealed that eyes, eyebrows and mouth were the most crucial for recognizing facial expressions. The results discussed above were obtained using a pre-defined set of training and test images from the databases. However, when using the 10-fold cross-validation scheme, the overall recognition rate on the Cohn-Kanade database was only 77.81%. The 10-fold cross-validation results for JAFFE database are not reported.

In the work discussed above, Nagi et al. [82] had developed their own open-mouth classifier using Haar-like features, since precise location of mouth region is crucial for recognizing the expressions of fear and surprise. 5000 positive images and 3000 negative images were used to train a 20-stage cascade classifier for detecting an open mouth. The scanning window-size for detecting the Haar-like features was set to 24×24 pixels. AdaBoost algorithm was used for feature selection, and it was run for 82 iterations (i.e., 82 weak classifiers were learned). A success rate of 93% was obtained for open-mouth detection on the Cohn-Kanade database. The proposed classifier for open-mouth detection was also shown to be better than the one developed by Castrillon-Santana et al. [101].

The works discussed in the above paragraphs are representative examples illustrating the key approaches in the area of static facial expression recognition. These, and other similar works, sorted in the order of year of publication, are summarized in the table 8.

Reference	Method	Validation	Database	Results
Zhang et al. [131] Year: 1998	Gabor at fiducial points + MLP ^a	10-fold	JAFFE	7-class: 90.1% w/o Fear: 92.2%
Lyons et al. [71] Year: 1999	Gabor at fiducial points + PCA + 2-class LDA	10-fold	JAFFE	7-class: 92%
		LOSO	JAFFE	7-class: 75%
Zheng et al. [132] Year: 2006	Gabor at fiducial points + KCCA (w/ semantic rating)	LOIO	JAFFE	6-class: 85.79%
		LOSO	JAFFE	6-class: 74.32%
		LOIO	POFA	6-class: 81.25%
		LOSO	POFA	6-class: 79.17%
	Gabor at fiducial points + KCCA (w/ expression label)	LOIO	JAFFE	6-class: 98.36%
		LOSO	JAFFE	6-class: 77.05%
		LOIO	POFA	6-class: 77.08%
		LOSO	POFA	6-class: 77.08%
Bartlett et al. [65] Year: 2006	Gabor: 9 scales, 8 orientations + AdaSVM	LOSO	CK	7-class: 93.3%
			POFA	7-class: 97.3%
			CK + POFA	7-class: 93.1%
Shih et al. [107] Year: 2008	DWT + 2D-LDA + SVM	10-fold	JAFFE	7-class: 94.13%
		LOSO	JAFFE	7-class: 95.71%
Shan et al. [105] Year: 2009	Boosted-LBP + SVM (RBF)	10-fold	CK	7-class: 91.4% 6-class: 95.1%
	LBP _{8,2} ^{u2} + SVM(RBF)	10-fold	CK	7-class: 88.9% 6-class: 92.6%
Jabid et al. [53] Year: 2010	LDP + SVM	10-fold ^b	CK	7-class: 93.4% 6-class: 96.4%
Kabir et al. [57] Year: 2010	LDPv + SVM ^c	7-fold	CK	7-class: 93.1% 6-class: 96.7%
	LDP + SVM	7-fold	CK	7-class: 91.3% 6-class: 94.5%
Xu et al. [125] Year: 2011	wLDP ^d w/ σ^e + PCA + NN	LOSO	JAFFE	7-class: 92.57%
Nagi et al. [82] Year: 2013	Region-based LBP _{8,1} ^{u2} + SVM	10-fold	CK	7-class: 77.81%
		Single test set	CK	7-class: 82.79%
			JAFFE	7-class: 77.93%

^aMLP- Multi-layer perceptron

^bperson-independent

^cWhen PCA was used for dimensionality reduction of LDPv features, Kabir et al. [56] found that 240 principal components were sufficient to produce the same results as presented in this table entry.

^dwLDP- weighted-LDP

^e σ - standard deviation

Table 8: A summary of some of the important works in computer vision-based facial expression recognition. These are all static, spatial approaches that use geometric, appearance-based or hybrid feature descriptors.

There has also been efforts to recognize expressions from image sequences [94, 33]. These methods use optical flow or face tracking to compute features such as direction and amount of movement of different parts of the face. These features are then used for training facial expression classifiers. Some of the earlier works that belong to the category of dynamic approaches for recognition of facial expressions are [59, 25, 99, 126, 90, 43, 33]. Some of these approaches recognize individual motion units from image sequences, and map them to facial expressions.

3.2 Facial Action Unit Recognition

In the previous section, some of the main approaches for the recognition of the six basic facial expressions and the neutral expression were discussed. These extract geometric and/or appearance-based features from the entire face region or its parts, and classify them into emotion categories. In this section, some of the key approaches for recognition of facial action units (AUs) are discussed. Facial AU recognition systems also employ geometric and appearance-based features, and exploit the spatial and temporal characteristics of facial muscle actions. Majority of these systems use dynamic approaches that are based on image sequences or videos. The facial action unit recognition systems that have been developed are capable of recognizing single AUs as well as combinations of AUs. However, systems that can recognize all the 44 AUs have not been built.

Some of the important works on automatic facial AU recognition are by Pantic et al. [93, 61, 55, 115], Tian et al. [110] and Bartlett et al. [19, 21]. A survey and comparison of the existing works on automatic facial AU recognition can be found in [61] and [115]. A few of the important works are summarized in table 9. Some of the works, such as [34, 20, 18, 61, 55], dealt with the recognition of AUs in images and image sequences of spontaneous expressions. The approaches that use the temporal dimension have been found to perform better than the approaches that use only the spatial information from static images [55]. The study [109] performed by Tian in 2004 showed that recognition of facial AUs is difficult at lower resolutions. However, at higher resolutions, similar performance could be achieved for recognition of facial expressions of emotion and for recognition of facial AUs, using geometric, appearance-based and hybrid face descriptors.

Reference	Method	Validation	Database	Results
Tian et al. [110] Year: 2001	Geometric + ANN	Single test set ^a	CK	10 lower facial AUs: 95.6%
			EH ^b	7 upper facial AUs: 95.4%
Pantic & Patras [93] Year: 2005	20 fiducial points + particle filter + temporal rules	-	CK	21 AUs: 93.3%
			MMI	9 AUs: 86.7%
Bartlett et al. [19] Year: 2005	Gabor + SVM	LOSO	CK + EH	17 AUs: 94.8%
Bartlett et al. [21] Year: 2006	Gabor + AdaSVM	LOSO	CK + EH	20 AUs: 90.9%
Littlewort et al. [65] Year: 2006	Gabor + AdaSVM	LOSO	CK	7 upper facial AUs: 92.9%
Tong et al. [111] Year: 2007	Gabor + AdaBoost + DBN	LOSO	CK	14 AUs: 93.3%
Lucey et al. [69] Year: 2007	S-PTS ^c + C-APP ^d + SVM	LOSO	CK	15 AUs: 95.5%
Koelstra et al. [61] Year: 2010	MHI ^e + GentleBoost + HMM	10-fold	MMI	27 AUs: 89.2%
	FFD ^f + GentleBoost + HMM	10-fold	MMI	27 AUs: 94.3%
			CK	18 AUs: 89.78%
Jiang et al. [55] Year: 2011	LBP _{8,1} ^{u2} + GentleBoost + SVM	LOSO	MMI	9 upper facial AUs: 85.8%
	LPQ + GentleBoost + SVM	LOSO	MMI	9 upper facial AUs: 90.2%
	LBP-TOP + GentleBoost + SVM	LOSO	MMI	8 upper facial AUs: 91.4%
	LPQ-TOP + GentleBoost + SVM	LOSO	MMI	8 upper facial AUs: 91.5%
Valstar & Pantic [115] Year: 2012	20 fiducial points + particle filter w/ FL ^g GentleBoost + SVM	LOSO	CK	16 AUs: 91.7%
			MMI	23 AUs: 95.3%

^aOnly one manually selected test set was used

^bEH- Ekman-Hager database [35]

^csimilarity normalized shape

^dcanonical appearance

^eMHI- Motion History Images

^fFFD- Free-Form Deformation

^gFL- Factorized Likelihoods

Table 9: A summary of some of the important works in facial action unit recognition.

3.3 Multimodal Emotion Recognition

Facial expressions are not the only way in which humans convey emotions. Emotions are also conveyed through other non-verbal means, such as, vocal intonation, touch, body posture and gestures [26]. Multimodal emotion recognition systems combine vocal and visual cues, and physiological signals to recognize the emotional state of a person. For the sake of completeness, some of the important works on multimodal emotion recognition are mentioned in this section.

Sebe et al. [104] proposed a Bayesian network topology to fuse visual and vocal cues for bimodal emotion recognition. Schuller et al. [102] integrated speech, touch and mouse gesture signals to recognize the prototypic and neutral expressions. Caridakis et al. [29] developed a multimodal emotion recognition system based on facial expressions, body and hand gestures, and speech. The emotions recognized included the six basic emotions and the emotions of irritation and pride. They found that the multimodal system performed better than the unimodal systems. Valstar et al. [119] investigated the use of a multimodal approach to automatically distinguish between posed smiles and spontaneous smiles. A survey of audio-based and audio-visual affect recognition methods is available in [128]. The article also provides a survey of the databases that are used for training and evaluation of these methods. Some of the other important works related to the domain of multimodal emotion recognition are [95, 27, 121].

3.4 Deficits

In this section, some of the deficits identified in the state of the art on automatic facial expression recognition and facial AU recognition are discussed.

From the study of the state-of-the-art, we observed that the same methods yielded different recognition rates, when trained and tested on different databases. The parameters of face descriptors¹⁰ and classifiers¹¹ that yielded the best results varied from database to database, and varied with the classification task¹² to be performed. In addition, the cross-database generalization performance of facial expression recognition methods is poor [65, 105, 125]. These show that the facial expression recognition and facial AU recognition research are not yet mature.

Most of the work on facial expression recognition still use manual input in the form of labelling of eye positions for face region estimation, labelling of facial fiducial points

¹⁰for example, scale and orientation of Gabor filters, the neighbourhood size of LBP, etc.

¹¹for example, kernel type and kernel parameters of SVMs, number of weak learners in AdaBoost, etc.

¹²number and type of expressions and facial AUs to be recognized

for feature tracking, etc. Appearance-based face descriptors require precise alignment of face region in images and image-sequences, and geometric face descriptor requires precise localization of facial features. Therefore, in order to automate the facial expression recognition process, it is necessary to first improve the reliability of facial feature detection and tracking methods.

Each of the face descriptors are suitable for different situations. Appearance-based face descriptors, in contrast to geometric face descriptors, are available even at low resolutions. LBP is robust to monotonic illumination variations, but not to non-monotonic illumination variations. LDP is robust to illumination variations and noise, and LPQ is insensitive to blurring. As can be seen, a single face descriptor does not perform well under different situations. Therefore, a combination of face descriptors should be used for facial expression recognition.

A number of different validation schemes have been used to evaluate the generalization performance of facial expression and facial AU recognition systems. While some researchers used 10-fold cross-validation scheme to compute the recognition rates, some others used the leave-one-subject-out (LOSO) scheme. A few of the authors used fixed training and test datasets. These differences in the performance evaluation protocol make the results of different works incomparable. In order to overcome this problem, a common evaluation protocol and performance metric(s)¹³ should be adopted for facial expression recognition, irrespective of the database used, and all researchers should report results accordingly.

Research has focused on frontal or nearly-frontal views of faces in images and image sequences. In real-life scenarios, however, faces appear commonly in profile, 3D-rotated or partially occluded views. In addition, most of the research has focused on posed expressions. Posed expressions are unnatural and differ from spontaneous expressions in spatial and temporal characteristics (see chapters 9 to 12 in [42]).

Compared to the datasets available for face detection, the databases available for facial expression recognition and facial AU recognition are very small in size. For face detection, the use of 5000 positive training samples is recommended [62]. The number of negative samples used for training frontal face detectors ranged from 10,000 [120] to millions [65]. However, such large datasets have not been used in facial expression recognition research. The need for such large datasets has been highlighted by Bartlett et al. in [65].

¹³Lucey et al. [68] have recommended the use of LOSO and confusion matrix for evaluating and reporting the results of facial expression recognition on the CK+ database.

The set of images used in experiments by different researchers is not uniform. Although CK database is commonly used, the researchers selected images and image sequences manually for their respective research works. The selected images differ from study to study. This makes comparison of results impossible.

The CK and JAFFE databases are the most widely used databases for research in static facial expression recognition. However, these databases have a few drawbacks. It is noted in [68] that the emotion labels of image sequences in the CK database were not validated. The image sequences in the CK database also do not contain the complete temporal pattern of facial expressions [96]. The sequences contain only the onset and apex phases; the offset phase is not available. In addition, some of the images in the CK database have date-time stamps appearing over the chin, thereby occluding the chin and its motion [96]. The JAFFE database, on the other hand, is a small database that contains only 213 images of 10 female subjects showing 7 expressions. Additionally, the expression of “fear” contained in JAFFE is less accurate [131].

There are many variables involved in automatic facial expression recognition and facial action unit recognition, which make automation challenging. For example, varying head sizes make it difficult to choose an optimal face model for face region estimation. Head poses, illumination conditions, image sharpness, noise, facial expression intensities, length of temporal phases (onset, peak, offset), facial characteristic features such as wrinkles, folds, scars, moles, etc.- all of these may vary from database to database, subject to subject and image to image. An automatic facial expression recognition system should be robust to all these variations. The systems developed so far consider only sub-sets of these variables, and assume the rest to be fixed or non-varying (for example, assumptions of frontal views, uniform lighting conditions, absence of occlusion, etc.).

However, recently, there have been efforts to use Bag of Words architecture with multi-scale dense scale-invariant feature transform (MSDF) and spatial pyramid matching (SPM) for facial expression recognition [108], with the goal of gaining invariance to scaling, rotation and translation, and thereby, achieving invariance to small amounts of in-plane head rotation, expression intensity, and wrinkling due to ageing. An overall recognition rate of 95.85% has been reported on CK+ database. However, the approach has not been tested systematically on young, middle-aged and older faces¹⁴, and age-invariance has not been verified and established.

¹⁴The CK+ database does not contain faces above the age of 50 years.

None of the works have studied the influence of age and ageing-related facial changes on the performance of facial expression and facial action unit recognition methods across a broad age spectrum.

4 Problem Formulation

Behaviour research has found that facial changes, such as wrinkles and folds, caused by ageing, resemble facial expressions of emotion [75], and reduce the clarity of the actual emotion conveyed [48]. ***There is no published work*** that investigates the influence of facial changes caused by ageing on the performance of automatic facial expression recognition systems¹⁵.

The review of the state-of-the art showed that similar face descriptors (see section 2.2.3), and similar approaches can be used for facial expression recognition as well as for age estimation. The appearance-based face descriptors, such as Gabor wavelets, LBP, LDP and their variants, describe the texture of the face. Gabor wavelets describe the face in terms of the wrinkles and folds of different frequencies and orientations appearing on the face. Feature extraction based on LBP, LDP and their variants count the different types of micro-patterns (wrinkles, folds, furrows, etc.) located at different parts of the face. Therefore, from a logical point of view, the facial wrinkles caused by ageing could be included in the features extracted for facial expression recognition.

The databases, namely, JAFFE, CK, POFA and MMI, which are commonly used for facial expression recognition and facial AU recognition, contain young faces (see table 10). The systems that have been developed so far are, therefore, trained and evaluated on young faces. Since the ageing-related facial changes could be included in the features extracted, these systems might not perform well on the faces of middle-aged and older age-groups. There is a possibility of increased confusion in the recognition of facial expressions of middle-aged and older people.

The main reason for the lack of research on the impact of facial ageing on facial expression recognition is that the databases that have been used in facial expression recognition research so far do not contain sufficient number of images of middle-aged faces, and also do not contain any images of older faces. Therefore, first, we need to identify or create a new facial expression database that contains an adequate number of images of facial expressions shown by young, middle-aged and older human subjects. After such a database

¹⁵Very few studies have been performed even in the behavioural and psychological fields on the impact of facial ageing on the perception of facial expressions.

is identified or created, experiments can be performed to examine whether ageing-related facial changes affect the performance of existing facial expression recognition approaches, and if so, how.

In our work, we do the following:

- Identify a suitable facial expression database that contains images of facial expressions posed by subjects belonging to young, middle-aged and older age-groups. In order to do this, we also examine the databases that have been used in human behavioural research.
- Evaluate and compare the performance of facial expression recognition systems that use appearance-based face descriptors, such as, Gabor filters and LBP, on young, middle-aged and older age-groups.
- Validate our hypotheses (see section 4.1) regarding the performance of facial expression recognition systems trained only on young faces.
- Identify the age-group that yields the best generalization performance across the entire age spectrum.
- Compare the performance of the facial expression recognition systems with the performance of humans¹⁶.
- Suggest improvements and generate ideas towards achieving better facial expression recognition performance for the middle-aged and older age-groups.

Key contributions:

- Evaluation and comparison of age-group-wise performance of facial expression recognition systems that use appearance-based features and that are trained only on young, middle-aged or older faces.
- Comparison of the performance of facial expression recognition systems with the performance of humans identified through psychological, neurological and behavioural studies.

In this work, only the problem of recognition of some of the basic expressions of emotion from 2D images is considered. Facial expression recognition from image sequences, and facial AU detection are out of scope of this work.

¹⁶In [68], the need to perform comparisons with the performance of humans on the same set of images that were used for evaluating the facial expression recognition system is highlighted.

4.1 Hypotheses

The hypotheses that are tested in our work are:

Hypothesis 1: If a facial expression recognition system that uses appearance-based face descriptors is trained only on facial expressions on young faces, then its generalization performance on older faces will be poor.

Hypothesis 2: If a facial expression recognition system that uses appearance-based face descriptors is trained only on facial expressions on young faces, then it will confuse the facial ageing-related wrinkles appearing on neutral faces of elderly with expressions of emotions.

5 Validation of Hypotheses

In this section, the process and results of validation of the hypotheses are elaborated. First, an appropriate facial expression database(s) for validating the hypotheses was selected. After selecting the database(s), the experiments to establish the validity of the hypotheses were designed and performed. The results of the experiments were then analysed to find evidence to support the hypotheses. For convenience, we also discuss in this section, how human performance compares with the system performance. The results of the comparison are, later, summarized in section 6.

5.1 Database Selection

To perform the age-group-wise analysis of the performance of facial expression recognition methods, we need a facial expression database that satisfies the following criteria:

- Contains 2D frontal images of faces showing the basic expressions of emotion as well as the neutral expression.
- Contains adequate¹⁷ number of subjects belonging to the young, middle-aged and older age-groups.
- Each subject displays the same set of facial expressions of emotion.
- No or minimal in-plane and out-of-plane head rotations.
- No occlusion of faces.

¹⁷The term “adequate” is used to indicate the preference for a good amount of variety in the faces belonging to each age-group in order to produce statistically significant results. For convenience, it was assumed to indicate the number 10 or more.

- Images were captured under identical illumination settings, and were post-processed identically.
- Ground truth was established by FACS experts or through experiments involving human subjects.
- All images are labelled based on the ground truth.
- Ethnic diversity not essential.

The different facial expression databases were evaluated against the criteria listed above. The results of the evaluation are listed in table 10. From the table, it is evident that FACES [36] is the only database in which the facial expressions as well as the age of faces are varied systematically, and which includes a good representation of the facial expressions of emotion in the older age-group category. In fact, each of the three age-categories, namely, young, middle-aged and older, are uniformly and adequately represented by the 171 human face models. Five of the six prototypic expressions, and the neutral expression are included in the database. Each image was captured under identical settings, and is labelled based on the ground truth established by trained human raters. In addition, the database has been validated by non-trained human subjects, and the validation results provide a summary of the human performance on the FACES database. The characteristics of the FACES database are discussed in detail in section 2.3.6. In our experiments, we use the images of resolution 335 pixels \times 419 pixels.

Database	JAFFE	POFA	MMI	FEED	CK+	FACES
2D frontal images	Yes	Yes	Yes ^a	Yes	Yes	Yes
Age-group coverage^b	Y	NA	NA	Y	NA	Y, M, O
Expressions^c	6B+N	6B+N	6B+N ^d	6B+N	6B+C+N ^e	5B+N
Facial occlusion^f	No	No	No ^g	No	No	No
Expert^h rating	No	Yes	Yes	Yes	Yes	Yes
Observerⁱ rating^j	Yes	Yes	No	No	No	Yes

^acontains profile images as well

^bY- Young, M- Middle-aged, O- Older, NA- Age-categorization and age-group labels not available

^cB- Basic, N- Neutral, C- Contempt

^dneutral expressions should be taken from the starting frames of videos or image sequences

^eneutral expressions should be taken from the starting frame of image sequences

^fglasses, scarf, moustache, beard, jewellery

^galso contains images with expresser wearing glasses, head scarf, or sporting beard

^htrained in FACS or in identifying facial expressions

ⁱa human subject who is not trained in FACS or in identifying basic facial expressions or AUs

^jor, human performance data

Table 10: Evaluation of facial expression databases against the selection criteria.

Guo and Wang introduced FACES database to the field of computer vision. We are the first to use FACES database to investigate facial expression recognition performance under ageing-related facial changes.

In addition to the FACES database, we also selected the JAFFE database to study the cross-database generalization performance of facial expression recognition methods. The images in the JAFFE database were also used for testing the software used in our study.

5.2 Human Performance on Selected Databases

In this section, a summary of the performance of humans on the databases selected for the validation of hypotheses is provided, so that the experimental results can be discussed in relation to human performance.

According to [70, 131], the semantic ratings given by humans on the JAFFE database agreed with the labels in 79.5% of the images. On excluding fear from the database ¹⁸, the performance of humans on JAFFE database increased to 85.6%. The reasons for this mismatch are attributed to a number of factors, which include, inaccurate or incorrect display of expressions, incorrect rating by human subjects, and possible mislabeling of images [131].

Tables 1, 2 and 3 provide insights into the performance of human subjects in recognizing the facial expressions displayed on images in the FACES database. The following can be inferred from the recognition rates provided in the tables:

- The overall performance of young human raters (84.51%) was higher than the overall performance of the middle-aged raters (81.32%) and the older raters (76.57%). The overall performance of older human raters was lower than the overall performance of the young and middle-aged raters.
- A pattern similar to the above can also be observed for the overall performance of the human raters in each expression category, with the exception of “joy”. [36] reports that:
 - The difference in performance between young and older raters was significant for the expressions “anger”, “disgust” and “sadness”.
 - The difference in performance between the young and middle-aged raters was significant for the expression “sadness”.

¹⁸The performance of the system was evaluated by excluding the images labelled with the expression “fear”, because “fear” is perceived and processed differently [11], is difficult for humans to perceive, and was expressed less accurately in the JAFFE database [131].

- No significant variation could be determined across the raters for the expression “joy”.
- The overall recognition rate, averaged over all the facial expressions and all human raters, was highest for the young faces (84.67%), followed by the middle-aged faces (81.61%). The recognition rate was lowest for the older faces (76.13%).
- A similar pattern was observed for all facial expressions, except “fear”. [36] reports that:
 - The performance of human raters on the older faces was lower than the performance of human raters on the young and middle-aged faces, for all expressions other than “fear”.
 - No significant variation in recognition rate could be determined across the age-group of face models for the expression “fear”.
 - The performance of human raters on middle-aged faces was significantly lower than that on young faces for the expressions “disgust”, “neutral” and “sadness”.
- Among all expressions, “joy” was the easiest for humans to identify (96%), and “disgust” was the most difficult to identify (68%). The expressions in the increasing order of difficulty for recognition are: joy, neutral, anger, fear, sadness, disgust.
- The overall recognition rate, averaged over all images in the FACES database and over all the human raters, was 80.80%. This is comparable to the human performance on the JAFFE database [131].

5.3 Experiments

In this section, the experiments that were performed to test the hypotheses stated in section 4.1 are described. Experiments were also performed to identify which single age-group enables facial expression recognition systems to generalize well across young, middle-aged and older faces. These experiments have also been included in this section.

5.3.1 Cross-Database Experiments

In this section, we examine how a facial expression recognition system trained on the images of young faces in the CK database, generalizes to the images of young faces in the JAFFE database, and to the images of young, middle-aged and older faces in the FACES database. The cross-database generalization performance of facial expression recognition systems on young faces is already known to be poor. Here, we investigate whether there is a further degradation in their cross-database performance, when tested on middle-aged and older faces.

5.3.1.1 Method

The facial expression recognition system implemented in [74], which uses Gabor wavelet coefficients¹⁹ as the face descriptor, and AdaBoost.MH²⁰ with multi-threshold stump learner²¹ as the classifier, was used to measure the cross-database performance on the images in the FACES and JAFFE databases. The model used in the experiments had been trained by Geovanny Macedo on the Gabor wavelet coefficients extracted from images of the six basic expressions and the neutral expression selected from the CK database. A bank of 40 Gabor filters with 5 different scales²² and 8 different orientations²³, was used. The implementation and the model are part of the ROS package named *brsu_facial_expression* in the RoboCupAtHome repository, available at [17]. An additional ROS node was written by us to communicate with the existing system, that is, to send the facial image data and to collect the recognition outputs, so as to perform evaluation on the test set. The code to compute performance metrics, such as, age-group-wise and overall expression recognition rate and confusion matrices, was also written.

The facial expression recognition system implemented in [74] expects the eye positions as input, based on which, the face model (see figure 1) is applied, and face region is extracted and normalized to 48x48 pixels. We performed two separate experiments- one in which the eye positions were detected automatically using the freely available open-source C library named *flandmark* [113, 112], and the other in which the eye positions were manually identified, and captured with the help of mouse event callback routines in OpenCV.

Note: *All processing was done internally as part of the feature extraction process. The original images in the FACES database were not modified in any way. The viewing and saving of any of the intermediate or final results of the pre-processing and feature extraction stages were disabled so as to abide by the “FACES Database Release Agreement”.*

¹⁹See section 2.2.3.1 for a detailed description of Gabor filters.

²⁰Number of iterations = 500.

²¹See [74] for a description of multi-threshold decision stumps.

²²3.5, 4.5, 7.5, 10.5, 15.5

²³0°, 22.5°, 45°, 67.5°, 90°, 112.5°, 135°, 157.5°

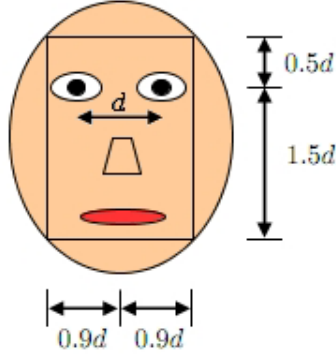


Figure 1: Face model used by Geovanny Macedo in [74], and adopted by us in our work. Image source: [74].

Eye detection method	FACES			JAFPE		
	Total	Correct	Accuracy	Total	Correct	Accuracy
Using landmark library	2052	614	29.92%	213	39	18.31%
Manual	2052	837	40.79%	213	84	39.44%

Table 11: Cross-database evaluation results: Overall recognition rates achieved on FACES and JAFPE databases using the facial expression recognition system based on Gabor features and AdaBoost.MH with MTSL, built in [74]. The facial expression recognition system was trained using images in the CK database [74].

5.3.1.2 Discussion

Table 11 shows that the overall facial expression recognition rate on FACES and JAFPE databases improved considerably, when the eye positions were identified manually. Table 12 reveals that the age-group-wise recognition rates obtained on the FACES database also increased when the eye positions were identified manually. Figures 2 and 3 also show a similar improvement in performance for the expressions in the FACES and JAFPE databases, respectively ²⁴. These results elucidate the importance of developing accurate and reliable facial feature detection methods, which form a pre-requisite for realizing reliable automatic facial expression recognition systems.

²⁴The only exception to this trend was the recognition rate for the expression of surprise in the JAFPE database (see figure 3). While 27 of the 30 images showing surprise could be recognized correctly when using the landmark library for eye position detection, 26 could be detected correctly when the eye positions were identified manually. The difference of one image is small, and therefore, does not affect the conclusion drawn.

FACES:Young			
Eye detection method	Total	Correct	Accuracy
Using flandmark library	696	300	43.10%
Manual	696	363	52.16%
FACES:Middle-aged			
Eye detection method	Total	Correct	Accuracy
Using flandmark library	672	193	28.72%
Manual	672	271	40.33%
FACES:Older			
Eye detection method	Total	Correct	Accuracy
Using flandmark library	684	121	17.69%
Manual	684	203	29.68%

Table 12: Cross-database evaluation results: Overall recognition rates achieved per age-group on FACES database using the facial expression recognition system based on Gabor features and AdaBoost.MH with MTSL, built in [74]. The facial expression recognition system was trained using images in the CK database [74].

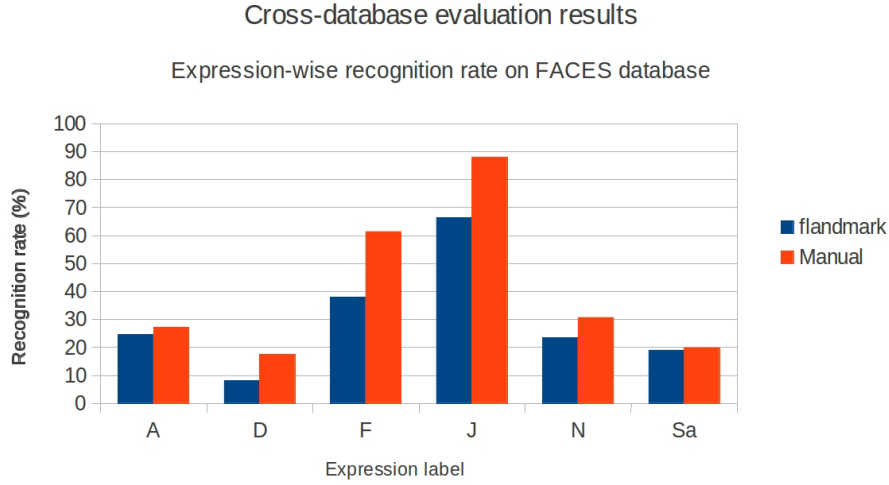


Figure 2: Cross-database evaluation results: Expression-wise recognition rates achieved on FACES database using the facial expression recognition system based on Gabor features and AdaBoost.MH with MTSL, built in [74]. The facial expression recognition system was trained using images in the CK database [74]. In the figure, A stands for anger, D for disgust, F for fear, J for joy, N for neutral and Sa for sadness. FACES database does not contain the facial expression of surprise.

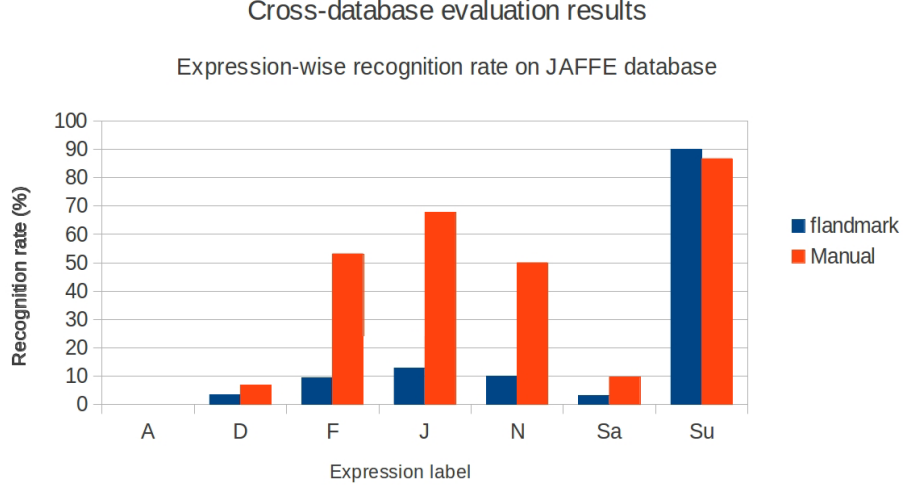


Figure 3: Cross-database evaluation results: Expression-wise recognition rates achieved on JAFFE database using the facial expression recognition system based on Gabor features and AdaBoost.MH with MTSL built in [74]. The facial expression recognition system was trained using images in the CK database [74]. In the figure, A stands for anger, D for disgust, F for fear, J for joy, N for neutral, Sa for sadness and Su for surprise. From the figure, it can be seen that none of the expressions of anger in the JAFFE database were recognized correctly.

The images from the CK database that were used for training the facial expression recognition system in [74] belonged to the age-group of 18 to 30 years, that is, to the young age-group. The age-group-wise recognition rates achieved on the FACES database (see table 12, rows corresponding to manual eye detection), show that the young age-group has the highest performance, followed by the middle-aged group and the older age-group, in that order. The accuracy of 52.16% achieved for the young age-group is comparable to the cross-database recognition performance achieved in other state-of-the-art studies (see table 13). The accuracy obtained for middle-aged and older age-groups are 40.33% and 29.68%, respectively, which are well below the accuracy obtained for the young age-group. These results show the inability of a facial expression recognition system, which is trained only on young faces using appearance-based face descriptors such as Gabor wavelet coefficients, to generalize well to middle-aged and older age-groups. **This is the first evidence in support of Hypothesis 1.**

Method	Trained on	Tested on	Performance (%)
Gabor + AdaSVM(linear) [65]	CK	POFA	56
Gabor + AdaSVM(RBF) [65]	CK	POFA	60
Boosted-LBP + SVM(linear) [105]	CK	MMI	50.8
Boosted-LBP + SVM(polynomial) [105]	CK	MMI	50.8
Boosted-LBP + SVM(RBF) [105]	CK	MMI	51.1
**Gabor + AdaBoost w/ MTSL	CK	FACES:Young	52.16
Boosted-LBP + SVM(linear) [105]	CK	JAFPE	40.4
Boosted-LBP + SVM(polynomial) [105]	CK	JAFPE	40.4
Boosted-LBP + SVM(RBF) [105]	CK	JAFPE	41.3
PCA + LBP + NN [125]	CK	JAFPE	38.57
PCA + LDP + NN [125]	CK	JAFPE	41.90
PCA + Weighted-LDP ^a + NN [125]	CK	JAFPE	43.38
** Gabor + AdaBoost w/ MTSL	CK	JAFPE	39.44

^aweighted with standard deviation

Table 13: Comparison of results of our cross-database experiments with the results reported in literature. (** denotes our results.)

Now, let us look at the recognition rates achieved for each expression and each age-group in the FACES database. These are shown in figure 4. It can be seen that the recognition rates for all expressions except sadness were highest for the young age-group, lower for the middle-aged group, and lowest for the older age-group²⁵. The greatest drop in recognition rate with ascending age-group was shown for the neutral expression. The neutral expression was confused greatly with fear, sadness and surprise, especially for the middle-aged and older age-groups (see figure 5). This indicates that the wrinkles and folds on the faces due to ageing are mistaken for facial expressions. **This is the first evidence in support of Hypothesis 2.**

The confusion of ageing-related wrinkles with the expression of sadness explains why the recognition rates for sadness across the different age-groups showed a reverse trend, with higher recognition rates for sadness expressed by the middle-aged and older age-groups than for sadness expressed by the young age-group.

²⁵Recognition rates for the expression of joy, for young and middle-aged groups, were almost identical.

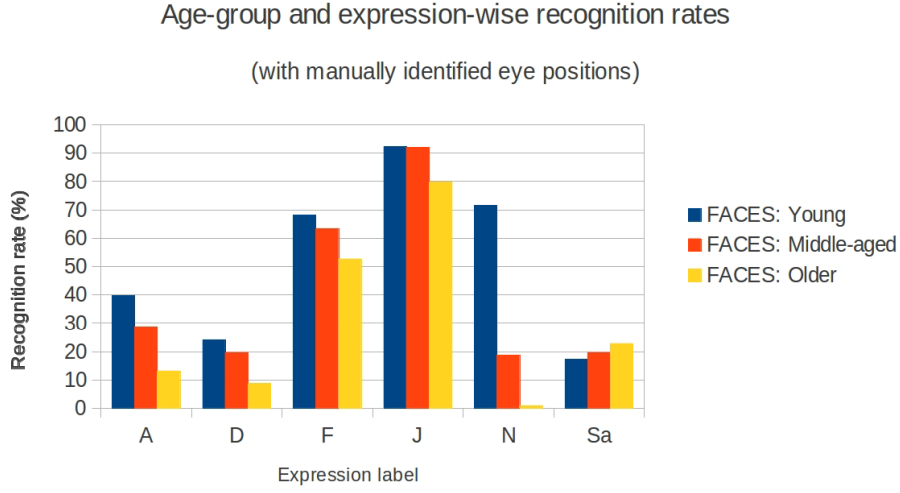


Figure 4: Cross-database evaluation results: Age-group and expression-wise recognition rates achieved on FACES database using the facial expression recognition system based on Gabor features and AdaBoost.MH with MTSL, built in [74], with manually identified eye positions. The facial expression recognition system was trained using images in the CK database [74]. In the figure, A stands for anger, D for disgust, F for fear, J for joy, N for neutral and Sa for sadness.

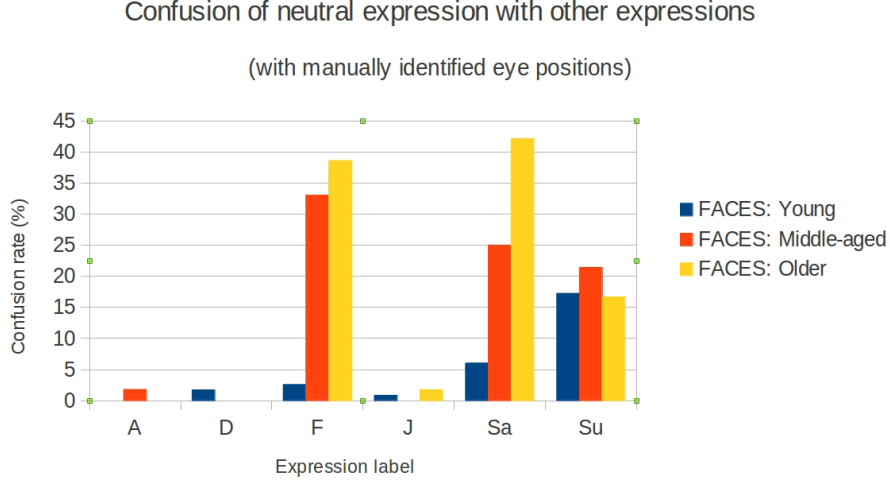


Figure 5: Cross-database results: The confusion of neutral expression with other expressions, for each age-group in the FACES database, for the case of manually identified eye positions. In the figure, A stands for anger, D for disgust, F for fear, J for joy, Sa for sadness and Su for surprise.

As can be seen from the figures 2, 3 and 4, the recognition rates²⁶ were higher for the expressions of joy, fear and surprise. The recognition rates for the neutral expression was also high for the images of young faces (that is, the images in FACES:Young and JAFFE). Anger, disgust and sadness were the hardest to recognize.

The confusion matrices for the JAFFE database and the young age-group of FACES database are given in table 14. It can be seen that, for the set of images in FACES:Young, anger was mainly confused with disgust, sadness and surprise, disgust was mainly confused with fear, sadness and surprise, and sadness was heavily confused with fear (64.66%). For the images in the JAFFE database, anger was heavily confused with surprise (60%), disgust was mainly confused with fear, sadness and surprise, and sadness was confused with fear and surprise.

Even though the recognition rates obtained for FACES and JAFFE databases improved when the eye positions were manually located, and the performance on the FACES: Young and JAFFE were comparable to the state of the art (see table 13), the cross-database generalization performance is still far from satisfactory. The low recognition rates show that variations in display and intensity of expressions, and variations in image characteristics affect the generalization accuracy. This issue needs to be handled before facial expression recognition systems can be deployed in real-life situations, where the environmental conditions and the characteristics of facial expressions are extremely variable.

5.3.1.3 Conclusion

The following are the key conclusions that can be drawn from the cross-database experiments described in this section:

- Reliable facial feature recognition methods are crucial for the success of image-based automatic facial expression recognition systems.
- Systems trained only on young faces cannot provide reliable performance on the middle-aged and older age-groups.
- Systems trained only on young faces confuse the ageing-related facial features on neutral faces of middle-aged and older human subjects with other expressions of emotions.
- Joy and surprise were amongst the easiest expressions to recognize, and sadness was amongst the hardest to recognize for all age-groups.

²⁶for the case of manually identified eye positions

FACES: Young							
	A	D	F	J	N	Sa	Su
A	39.66	15.52	4.31	4.31	4.31	18.10	13.79
D	1.72	24.14	38.79	1.72	0	21.55	12.07
F	0	0	68.1	1.72	6.03	2.59	21.55
J	0	0	6.9	92.24	0.86	0	0
N	0	1.72	2.59	0.86	71.55	6.03	17.24
Sa	0.86	0	64.66	4.31	3.45	17.24	9.48

JAFFE							
	A	D	F	J	N	Sa	Su
A	0	0	13.33	10	3.33	13.33	60
D	0	6.9	37.93	3.45	6.9	10.34	34.48
F	0	0	53.13	0	3.13	0	43.75
J	0	0	12.90	67.74	6.45	0	12.90
N	0	0	6.67	0	50	0	43.33
Sa	0	0	35.48	6.45	0	9.68	48.39
Su	0	0	0	3.33	10	0	86.67

Table 14: Cross-database evaluation results: Confusion matrices for FACES:Young and JAFFE. In the table, A stands for anger, D for disgust, F for fear, J for joy, N for neutral, Sa for sadness and Su for surprise. All the values in the table are in percentage (%), and were computed based on the recognition results obtained for the case of manually identified eye positions.

5.3.2 Intra-age-group Experiments

In this section, we describe the experiments that were performed on images belonging to the same age-group in the FACES database. The objective was to study the generalization performance of a facial expression recognition system on images belonging to the same age-group as that used to train the system. To be precise, we examine the following:

- How does a facial expression recognition system perform, when it is trained and tested on images of middle-aged faces?
- How does the system perform, when it is trained and tested on older faces?
- How do the performances in the above cases compare with the performance of the system when it is trained and tested on images of young faces?

5.3.2.1 Method

The images in the FACES database were categorized into three groups, according to the age-group to which the face models belonged- FACES:Young, FACES:Middle-aged and FACES:Older. The facial expression recognition systems used for the experiments were based on $LBP_{8,R}^{u2}$ ²⁷ and AdaBoost.MH. We used four different systems based on four different LBP-based feature descriptors, namely, $LBP_{8,1}^{u2}$, $LBP_{8,2}^{u2}$, block-wise $LBP_{8,2}^{u2}$ and block-wise $LBP_{8,2}^{riu2}$. The performance of all the four systems on FACES:Young, FACES:Middle-aged and FACES:Older image sets was studied. The particulars of the facial expression recognition systems used in the intra-age-group experiments, are detailed in the paragraphs below.

In the pre-processing stage, the images were converted to grayscale and the face regions were estimated based on manually identified eye positions and a pre-defined face model (see figure 1). The estimated face region in each image was normalized to a resolution of 48x48 pixels. The pre-processing steps are identical to those performed by Geovanny Macedo in [74], and we have re-used the OpenCV-based routines for face region estimation and normalization implemented by him as part of [74].

The LBP-based operators were then applied to the normalized face region to extract facial expression information. We used the open-source C++ implementation of LBP-based face descriptors written by Navid Nourani-Vatani, which is downloadable at [84]. Nourani-Vatani created the C++ implementation based on the Matlab code (available at [16]) written by the original creators of LBP face descriptors. We conducted experiments

²⁷See section 2.2.3.2 for a detailed description of LBP operators.

using four different LBP-based feature descriptors. Of these, $LBP_{8,1}^{u2}$ and $LBP_{8,2}^{u2}$ use 8 samples from a circular neighborhood of radius 1 pixel and 2 pixels, respectively²⁸, to compute the binary pattern and decimal code for the center pixel. After applying the LBP operators to all pixels in the image, the histogram of patterns in the whole face region is computed. The histograms for $LBP_{8,1}^{u2}$ and $LBP_{8,2}^{u2}$ contain a total of 59 bins²⁹. We use the 59-bin histograms in unnormalized form as the feature vector representing facial expression information.

Block-wise $LBP_{8,2}^{u2}$ and block-wise $LBP_{8,2}^{riu2}$ are variants of $LBP_{8,2}^{u2}$. These also use 8 samples from a circular neighbourhood of radius 2 pixels to compute the LBP codes for each pixel in the image. However, these differ from $LBP_{8,2}^{u2}$ in the manner in which the feature vector is computed. The normalized face region is divided into blocks and separate histograms are computed for each block. The histograms for all blocks are then concatenated to form a sequence of histograms. The histogram sequence thus obtained is used as the feature vector. When block-wise $LBP_{8,2}^{u2}$ is used, 59-bin histograms based on uniform patterns are extracted from each block. When block-wise $LBP_{8,2}^{riu2}$ is used, 10-bin histograms based on rotation-invariant uniform patterns are extracted from each block. In our experiments, we divided the face region into 6x6 non-overlapping blocks³⁰. Therefore, the length of a feature vector computed for block-wise $LBP_{8,2}^{u2}$ is $59 \times 36 = 2124$, and the length of a feature vector computed for block-wise $LBP_{8,2}^{riu2}$ is $10 \times 36 = 360$.

The feature vectors extracted from the images were used to train an AdaBoost.MH learner with multi-threshold stump learner (MTSL) as the weak learner. We used the MultiBoost package [22] that provides a C++ implementation of AdaBoost.MH. The MultiBoost package is distributed under GPL licence³¹. The simple command-line interface provided by the package was used to train an AdaBoost.MH learner, to test the generated strong hypothesis model, and to produce the confusion matrix. The number of iterations was set to 100³².

²⁸This is denoted by the subscripts 8,1 and 8,2 respectively.

²⁹In the implementation [84] used in our experiments, the uniform patterns are assigned to bins with index ranging from 0 to 57, and all non-uniform patterns are assigned to the bin with index 58.

³⁰Shan et al. [105] have divided the face region into 6x7 blocks. The resolution of the normalized face region in our experiments is 48x48 pixels. Therefore, to create blocks of identical size, we divided the face region into $6 \times 6 = 36$ non-overlapping blocks.

³¹The MultiBoost package can be downloaded from the MultiBoost webpage at [5]

³²Preliminary experiments using local binary pattern rotation-invariant histogram Fourier features and ‘n’ AdaBoost.MH iterations, on selected images in FACES:Young, showed no variations in overall recognition rates. The values set for ‘n’ were 100, 150, 200, 250, 300, 350, 400, 500, 600 and 1000. The highest recognition rate was obtained using 100 iterations, and hence 100 was chosen for ‘n’ in the intra- and inter-age-group experiments.

“Leave-One-Subject-Out” cross-validation scheme was used to evaluate the generalization performance³³. Under LOSO, the classifier is trained using all images except those of one subject. The images of the left-out subject are used for testing. The process is then repeated by leaving out another subject and using the images of all remaining subjects for training. The process of training and testing is repeated until all subjects have been used once for testing. The average of the recognition rates obtained in each run is computed and reported as the overall generalization performance of the facial expression recognition method. We have computed the average overall recognition rate, as well as the average recognition rate for each expression.

The experimental set up is summarized in table 15.

	FACES:Young	FACES:Middle-aged	FACES:Older
No. of subjects	58	56	57
Types of expression	6	6	6
No. of images per subject per expression	2	2	2
No. of images per subject	12	12	12
Total no. of images	696	672	684
Feature descriptor	LBP-based	LBP-based	LBP-based
Strong learner	AdaBoost.MH	AdaBoost.MH	AdaBoost.MH
Weak learner	MTSL	MTSL	MTSL
No. of weak learners used	100	100	100
Cross-validation scheme	LOSO	LOSO	LOSO
Training set size in each run of LOSO	684	660	672
Test set size in each run of LOSO	12	12	12
Total train-test runs for LOSO	58	56	57

Table 15: Experimental set-up for intra-age-group experiments performed on the FACES database. LOSO stands for “Leave-One-Subject-Out”. MTSL stands for “Multi-Threshold Stump Learner”.

³³The use of LOSO is recommended in [68] for reporting generalization performance on CK+ database. We adopt the same strategy for our experiments on FACES database.

Note: All processing was done internally as part of the feature extraction process. The original images in the *FACES* database were not modified in any way. The viewing and saving of any of the intermediate or final results of the pre-processing and feature extraction stages were disabled so as to abide by the “*FACES Database Release Agreement*”.

5.3.2.2 Discussion

The overall recognition rates obtained for the intra-age-group experiments on *FACES:Young*, *FACES:Middle-aged* and *FACES:Older* are listed in table 16. From the table, the following observations can be made:

Face descriptor	<i>FACES:Young</i>	<i>FACES:Middle-aged</i>	<i>FACES:Older</i>
$LBP_{8,1}^{u2}$	55.028%	45.833%	38.011%
$LBP_{8,2}^{u2}$	60.775%	52.975%	41.373%
Block-wise $LBP_{8,2}^{u2}$	87.499%	85.416%	77.631%
Block-wise $LBP_{8,2}^{riu2}$	82.614%	73.065%	64.911%

Table 16: Overall recognition rates obtained for intra-age-group experiments. The experimental set-up is given in table 15. Border copy was enabled while extracting LBP-based features in order to facilitate the computation of LBP codes for the pixels along the image boundary. For block-wise LBP operators, the face region was divided into $6 \times 6 = 36$ non-overlapping blocks.

- $LBP_{8,2}^{u2}$ produced better overall performance than $LBP_{8,1}^{u2}$ for all three age categories. This shows that $LBP_{8,2}^{u2}$ has more discriminative power than $LBP_{8,1}^{u2}$ for facial expression recognition. The increase in performance achieved for middle-aged group was higher than that for young age-group. The increase in performance achieved for older age-group was lower than that for young age-group. This could be due to the difference in the nature of the ageing-related facial wrinkles and folds on middle-aged and older faces. The ageing-related facial changes are more prominent on older faces than on middle-aged faces. This would result in less performance gain from increased neighbourhood radius for older faces than for middle-aged faces.
- Block-wise $LBP_{8,2}^{u2}$ yielded significant improvements in performance compared to $LBP_{8,2}^{u2}$ for all three age-groups. An increase of 26.724% in recognition rate was obtained for *FACES:Young*, 32.441% for *FACES:Middle-aged* and 36.258% for *FACES:Older*. The increase in recognition rate was the highest for the older age-group, followed by the middle-aged group. This could be due to the fact that there are fewer ageing-related wrinkles and folds in each block, as opposed to the face taken as a

whole. Consequently, the influence of ageing-related wrinkles is considerably minimized through the use of block-wise $\text{LBP}_{8,2}^{u2}$ operator. This, in turn, increases the recognition rates for middle-aged and older age-groups by a greater margin compared to the young age-group.

- Feature extraction based on $\text{LBP}_{8,2}^{u2}$ computes a single 59-bin histogram for the entire face. In this process, regional information about occurrence of patterns is lost. Block-wise $\text{LBP}_{8,2}^{u2}$, on the other hand, computes a separate 59-bin histogram for each sub-region (block) of the face, and concatenates them to form the feature vector, thereby, capturing regional pattern information over the entire face. The significant improvements in performance produced by block-wise $\text{LBP}_{8,2}^{u2}$ features over $\text{LBP}_{8,2}^{u2}$ features show that regional pattern information increases the discriminative power of the system to recognize facial expressions of emotions.
- Block-wise $\text{LBP}_{8,2}^{u2}$ also performed better than block-wise $\text{LBP}_{8,2}^{riu2}$ for all three age-groups. Block-wise $\text{LBP}_{8,2}^{riu2}$ counts rotation-invariant uniform patterns to construct the histogram for each block. Each uniform pattern is represented by the rotationally equivalent uniform pattern with the smallest decimal value [86]. This causes information about the orientation of the micro-patterns to be lost. The lower performance obtained using block-wise $\text{LBP}_{8,2}^{riu2}$ shows that orientation information is crucial for facial expression recognition. The degradation in performance when using rotation-invariant uniform patterns was greater for middle-aged and older age-groups. This could be due to the increased presence of ageing-related facial wrinkles and folds in the counts of rotation-invariant uniform patterns.
- Amongst the four LBP-based face descriptors evaluated, block-wise $\text{LBP}_{8,2}^{u2}$ yields the best performance. The recognition rates for all three age-groups obtained using block-wise $\text{LBP}_{8,2}^{u2}$ are comparable to the results reported in the state of the art (see tables 7 and 8).
- The recognition rates for all three age-groups obtained using block-wise $\text{LBP}_{8,2}^{u2}$ are comparable to the intra-age-group performance of humans on FACES database (see figure 6). The performance of young human raters on FACES:Young, and the performance of the facial expression recognition (FER) system trained and tested on FACES:Young were almost identical. However, the recognition rates achieved by middle-aged human raters on FACES:Middle-aged, and the recognition rates achieved by older human-raters on FACES:Older, were lower than the recognition rates obtained by the facial expression recognition system trained and tested on FACES:Middle-aged, and trained and tested on FACES:Older, respectively. The lower performance of middle-aged and older human raters could be due to the impact

Comparison of intra-age-group performance of humans and FER system

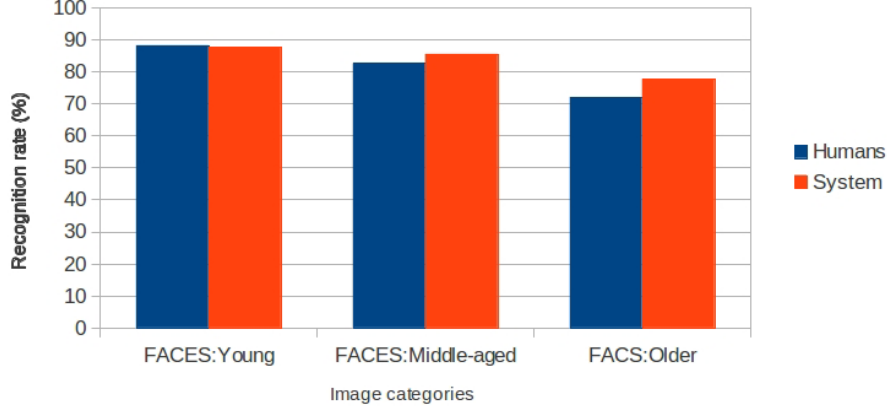


Figure 6: Comparison of intra-age-group performance of humans and the facial expression recognition (FER) system for each age-group. The intra-age-group performance of humans was determined from table 3. The recognition rates shown for the FER system are those corresponding to block-wise $LBP_{8,2}^{u2}$ (see table 16). The human performance on FACES:Young should be read as “Performance of young human raters on FACES:Young”. A similar remark holds for the human performance on FACES:Middle-aged and FACES:Older. The system performance on FACES:Young should be read as “Performance of the FER system trained and tested on FACES:Young”. A similar remark holds for the system performance on FACES:Middle-aged and FACES:Older.

of neuropsychological changes caused by adult ageing on facial expression recognition, as suggested by Ruffman et al. in [100].

- The system performance on the older age-group was poor compared to its performance on middle-aged and young age-groups, for all the four face descriptors used. The recognition performance trend over the three age-groups is similar to the human performance on FACES database (see figure 6).

Block-wise $LBP_{8,2}^{u2}$ produced the best performance for all age-groups. In table 17, we report the expression-wise recognition rates obtained using block-wise $LBP_{8,2}^{u2}$ for each of the three age-groups. Comparison of these recognition rates with the human performance on FACES database is illustrated in figures 7, 8 and 9.

Image category	A	D	F	J	N	Sa
FACES:Young	79.31	89.655	92.241	100	92.241	71.551
FACES:Middle-aged	78.571	91.071	92.857	96.428	86.607	66.964
FACES:Older	78.07	72.807	84.21	94.736	72.807	63.157

Table 17: Intra-age-group experiment results: Comparison of expression-wise performance of the facial expression recognition system based on block-wise $LBP_{8,2}^{u2}$ for each of the three age-groups. In the table, A stands for anger, D for disgust, F for fear, J for joy, N for neutral and Sa for sadness.

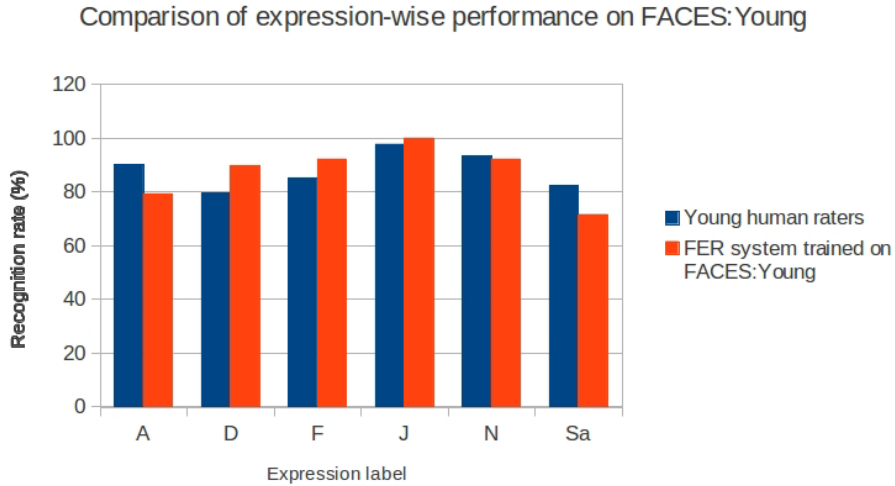


Figure 7: Intra-age-group experiment results: Comparison of expression-wise performance of humans and FER system on FACES:Young. The recognition rates for humans are taken from table 3. The recognition rates for the FER system are taken from table 17. In the figure, A stands for anger, D for disgust, F for fear, J for joy, N for neutral and Sa for sadness.

Comparison of expression-wise performance on FACES:Middle-aged

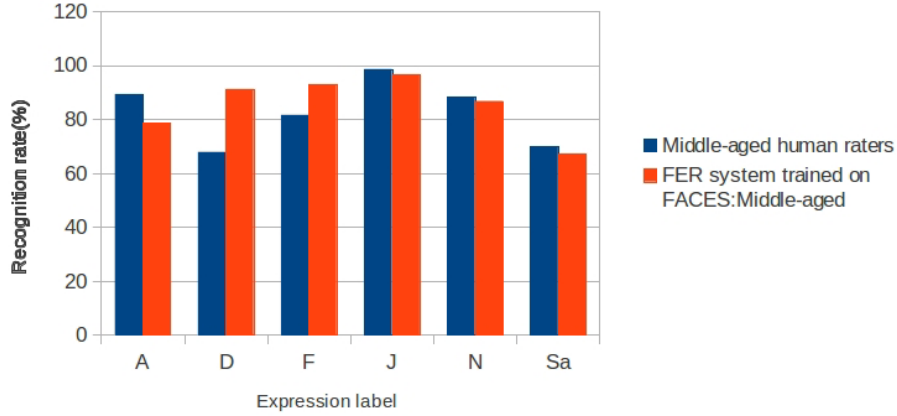


Figure 8: Intra-age-group experiment results: Comparison of expression-wise performance of humans and FER system on FACES:Middle-aged. The recognition rates for humans are taken from table 3. The recognition rates for the FER system are taken from table 17. In the figure, A stands for anger, D for disgust, F for fear, J for joy, N for neutral and Sa for sadness.

Comparison of expression-wise performance on FACES:Older

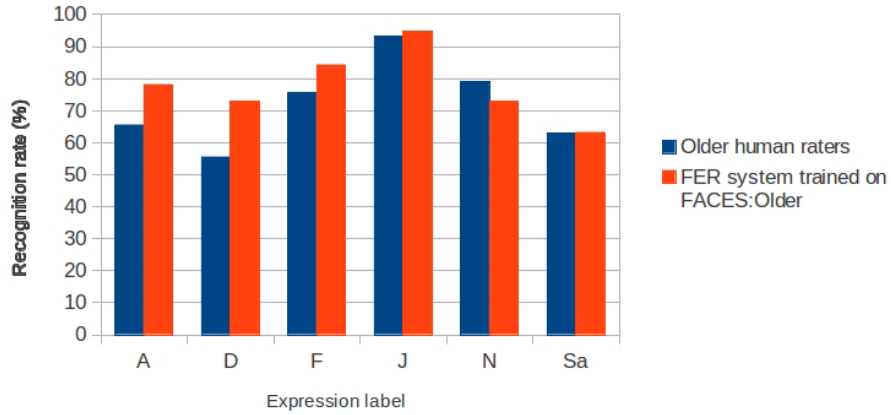


Figure 9: Intra-age-group experiment results: Comparison of expression-wise performance of humans and FER system on FACES:Older. The recognition rates for humans are taken from table 3. The recognition rates for the FER system are taken from table 17. In the figure, A stands for anger, D for disgust, F for fear, J for joy, N for neutral and Sa for sadness.

From the table 17 and the figures 7, 8 and 9, the following can be observed:

- For all three age-groups, the highest recognition accuracy was obtained for the expression of joy. Joy was also the easiest for humans to recognize. The recognition rates achieved by the system and the human raters for the expression of joy were almost identical. However, humans slightly outperformed the system in recognizing joy.
- For all three age-groups, the expression of sadness was the hardest to recognize for the system. For humans, sadness was the second most difficult expression to recognize.
- For all three age-groups, the performance of the system in recognizing the expression of anger was identical. Anger was also the second hardest expression for the system to recognize, with the exception of disgust on older faces.
- Humans performed better than the system in recognizing the expressions of anger and sadness for the young and middle-aged groups. For the older age-group, the system performed better than humans in recognizing anger. The performance of humans and the system were identical for the expression of sadness on older faces.
- The system outperformed humans in recognizing the expressions of disgust and fear for all three age-groups.
- The performance of the system on expressions of disgust and fear were nearly the same for young and middle-aged groups. However, there was a considerable drop in the performance of the system on disgust and fear expressed on older faces.
- The neutral expression could be recognized equally well by humans and the system for the young and middle-aged groups. However, for the older age-group, humans performed better than the system in recognizing the neutral expression.
- The recognition rates obtained by the system on neutral and sad faces decreased as the age-group changed from young to middle-aged, and to older.
- The system confused the neutral expression on older faces with sadness (14.035%), anger (6.14%), fear (4.385%) and disgust (2.631%).
- The system confused sadness with the neutral expression (13.157%) for the older age-group, and with the expression of anger (12.5%) for the middle-aged group.

5.3.2.3 Conclusion

The key conclusions that can be drawn from the results of the intra-age-group experiments are summarized below:

- An (8,2) neighbourhood for LBP operator is more resistant to ageing-related facial changes than an (8,1) neighbourhood.
- Information about regional distribution of micropatterns on the face is essential for achieving good facial expression recognition performance for all three age-groups.
- Information about orientation of micropatterns is essential for good facial expression recognition performance for all three age-groups.
- Regional pattern information and information about pattern orientation minimize the effect of ageing-related facial changes on the performance of the facial expression recognition system.
- The overall intra-age-group performance achieved by the system for the three age-groups is comparable to the state of the art, as well as to the performance of humans. Having said that, the system performs better than humans on middle-aged and older faces.
 - The performance of older human raters are reported to be affected by ageing-related neuropsychological changes [100]. A facial expression recognition system does not suffer from such cognitive capacity degradation, and therefore, produces better overall intra-age-group performance on older age-groups, compared to humans.
- **The performance of the system decreases as the age-group to which it is applied is changed from young to middle-aged and to older.** This trend is similar to that shown by humans.
- However, there are differences in the expression-wise performance of humans and the system for the different age-groups. While the system outperforms humans in recognizing disgust and fear for all age-groups, humans are better at recognizing anger and sadness on young and middle-aged faces. Humans also perform better than the system in recognizing the neutral expression on older faces.
- Irrespective of the age-group, joy is the easiest for the system to recognize, and sadness is the hardest to recognize.

- The intra-age-group performance of the system in recognizing anger is almost identical for the three age-groups.
- The performance of the system trained and tested on FACES:Young, and that of the system trained and tested on FACES:Middle-aged are almost identical for the expressions of anger, disgust and fear. There is a slight drop in the performance of the latter for the expressions of joy, neutral and sadness.
- The performance of the system trained and tested on FACES:Young and that of the system trained and tested on FACES:Older are almost identical for the expression of anger. For the remaining five expressions, there is a significant drop in the performance of the latter.
- The performance of the system trained and tested on FACES:Middle-aged and that of the system trained and tested on FACES:Older are almost identical for the expressions of anger and joy. For the remaining four expressions, there is considerable drop in the performance of the latter. The highest drop in performance between the former and the latter was for the expression of disgust.

5.3.3 Inter-age-group Experiments

In this section, we describe the experiments performed to study the ability of a facial expression recognition system that is trained on faces belonging to one of the three age-groups, to generalize to the other two age-groups. The objective was to identify which of the three age-groups provides the best generalization performance over the other age-groups. To be precise, we examine the following:

- How does a facial expression recognition system that is trained on young faces perform on middle-aged and older faces?
- How does a facial expression recognition system that is trained on middle-aged faces perform on young and older faces?
- How does a facial expression recognition system that is trained on older faces perform on young and middle-aged faces?
- Which of the above three systems performs best across all three age-groups?

5.3.3.1 Method

The architecture of the facial expression recognition system used for inter-age-group experiments is similar to that used for the intra-age-group experiments described in section 5.3.2. We used three different facial expression recognition systems based on block-wise $LBP_{8,2}^{u2}$ ³⁴. One of these systems was trained on FACES:Young, the other was trained on FACES:Middle-aged and the third was trained on FACES:Older. The performance of each of these system on the other two image sets was then studied.

The stages involved in facial expression recognition are pre-processing, feature extraction and expression classification. Images were first loaded in grayscale format, and then, based on manually identified eye positions, a face model (figure 1) was applied to each grayscale image to estimate the face region. The estimated face region was scaled to a resolution of 48x48 pixels. Block-wise $LBP_{8,2}^{u2}$ -based feature extraction method was applied to the normalized face region to compute the feature vector. The feature extraction method applied $LBP_{8,2}^{u2}$ operator on the entire face region, divided the face region into $6 \times 6 = 36$ non-overlapping blocks, constructed a 59-bin histogram of patterns for each block, and concatenated the histograms to form the feature vector. Feature vectors extracted from images in the training sets were used to train the facial expression recognition systems, and those extracted from images in the test sets were used for evaluating the generalization performance of the trained systems. AdaBoost.MH with multi-threshold stump learner was used as the classifier, and the number of iterations was set to 100.

The pre-processing stage is similar to that used by Geovanny Macedo in [74]. We have re-used the OpenCV-based pre-processing module implemented by him. For extracting LBP-based features, we used the open-source C++ implementation [84] from Navid Nourani-Vatani³⁵. The MultiBoost package [22] was used to train and test AdaBoost.MH learner for facial expression classification.

Note: *All processing was done internally as part of the feature extraction process. The original images in the FACES database were not modified in any way. The viewing and saving of any of the intermediate or final results of the pre-processing and feature extraction stages were disabled so as to abide by the “FACES Database Release Agreement”.*

³⁴Block-wise $LBP_{8,2}^{u2}$ had produced the best performance in the intra-age-group experiments.

³⁵Nourani-Vatani’s implementation of LBP and its variants was based on the Matlab code [16] from the original creators.

5.3.3.2 Discussion

The generalization performance of the three facial expression recognition systems is summarized in table 18. Figure 10 compares the inter- and intra-age-group recognition rates achieved.

Trained on	Tested on	Overall Recognition Rate
FACES:Young	FACES:Middle-aged	79.612
	FACES:Older	66.958
FACES:Middle-aged	FACES:Young	85.631
	FACES:Older	77.777
FACES:Older	FACES:Young	80.603
	FACES:Middle-aged	82.886

Table 18: Inter-age-group experiment results. Blockwise $LBP_{8,2}^{u2}$ operator was used to generate the feature vectors. AdaBoost.MH with MTSL was chosen as the classifier, and was configured to run for 100 iterations.

From the table 18 and the figure 10, it can be seen that:

- The facial expression recognition system trained on FACES:Young generalized poorly to middle-aged and older age-groups. There is a degradation of performance as the age-group changes from young to middle-aged, and to older. This is similar to the findings of cross-database experiments in section 5.3.1. **This is yet another evidence in support of Hypothesis 1**, and shows that ageing-related wrinkles and folds affect the performance of a facial expression recognition system that is trained solely on young faces.
- The facial expression recognition system trained on FACES:Middle-aged generalized well to young and middle-aged faces. The performance on these two age-groups were identical. However, there was a significant drop in the performance of the system on older age-group.
- The facial expression recognition system trained solely on FACES:Older generalized reasonably well to young and middle-aged faces. The performance on these two age-categories were comparable³⁶ However, the performance on the older age-group was much lower.

³⁶Even when the classifier was trained for 150 iterations using all images in FACES:Older, almost identical performance was obtained on the young (82.183%) and middle-aged (81.398%) faces.

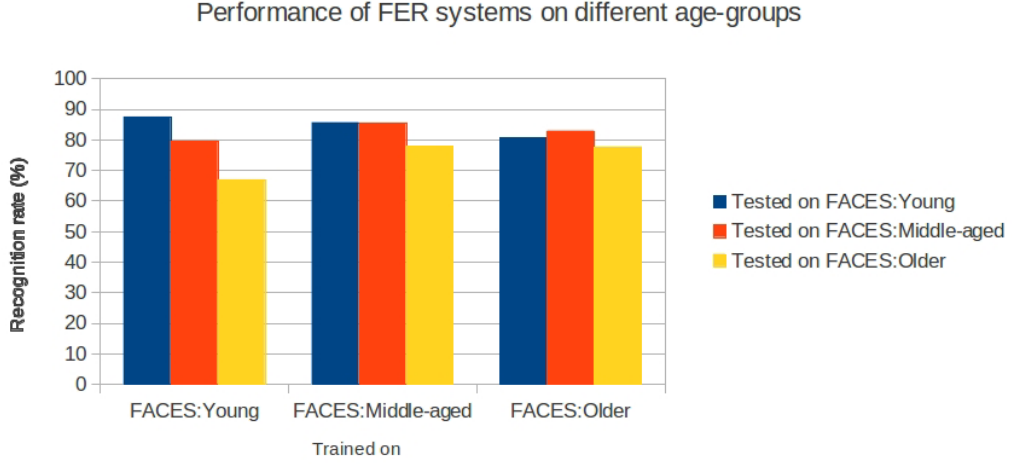


Figure 10: Performance of facial expression recognition systems based on block-wise $LBP_{8,2}^{u2}$ and AdaBoost.MH that have been trained and tested on young, middle-aged and older age-groups. The own-age-group recognition rates were taken from the results of intra-age-group experiments (row 4, table 16). The inter-age-group recognition rates were taken from table 18

- Among the three facial expression recognition systems, the one that was trained on FACES:Young produced the lowest recognition rate on middle-aged and older age-groups, and the highest on the young age-group.
- The performance of the facial expression recognition system trained on FACES:Middle-aged, and the performance of the system trained on FACES:Older were identical for the older age-group. In addition, the former performed better than the latter on young and middle-aged faces.
- Moreover, the performance of the system trained on FACES:Young was only slightly better than the performance of the system trained on FACES:Middle-Aged for the young age-group.
- From the above three bullet points, we can conclude that the system trained on FACES:Middle-aged provides the best generalization performance across all three age-groups.
- An “own-age-group bias” was observed for the young age-group. The system trained on FACES:Young performed best on young faces.

- The system trained on FACES:Middle-aged produced the best performance on middle-aged faces, among all systems.
- The best recognition rate that could be obtained on FACES:Older was only 77.777%.

In table 19, we put together, the expression-wise inter-age-group and intra-age-group recognition rates achieved by facial expression recognition systems trained separately on the three age-group-wise image categories in FACES database. The figures 11, 12 and 13 illustrate graphically, how the expression-wise performances of the systems for the three age-groups compare with each other.

Trained on FACES:Young							
Tested on FACES:	Anger	Disgust	Fear	Joy	Neutral	Sadness	Overall
Young	79.31%	89.66%	92.24%	100%	92.24%	71.55%	87.5%
Middle-aged	75.89%	89.29%	98.21%	94.64%	57.14%	62.5%	79.61%
Older	57.02%	85.96%	92.11%	92.98%	19.3%	54.39%	66.96%
Overall	70.74%	88.30%	94.19%	95.87%	56.23%	62.81%	78.02%

Trained on FACES:Middle-aged							
Tested on FACES:	Anger	Disgust	Fear	Joy	Neutral	Sadness	Overall
Young	81.03%	87.93%	89.66%	97.41%	93.97%	63.79%	85.63%
Middle-aged	78.57%	91.07%	92.86%	96.43%	86.61%	66.96%	85.42
Older	79.82%	84.21%	82.46%	94.74%	73.68%	51.75%	77.78%
Overall	79.81%	87.74%	88.32%	96.19%	84.75%	60.84%	82.94%

Trained on FACES:Older							
Tested on FACES:	Anger	Disgust	Fear	Joy	Neutral	Sadness	Overall
Young	80.17%	84.48%	92.24%	94.83%	80.17%	51.72%	80.60%
Middle-aged	83.93%	81.25%	92.86%	97.32%	84.82%	57.14%	82.89%
Older	78.07%	72.81%	84.21%	94.74%	72.81%	63.16%	77.63%
Overall	80.72%	79.51%	89.77%	95.63%	79.27%	57.34%	80.37%

Table 19: Performance of facial expression recognition systems trained and tested on young, middle-aged and older faces in FACES database. The own-age-group recognition rates were taken from the expression-wise results of intra-age-group experiments (see table 17).

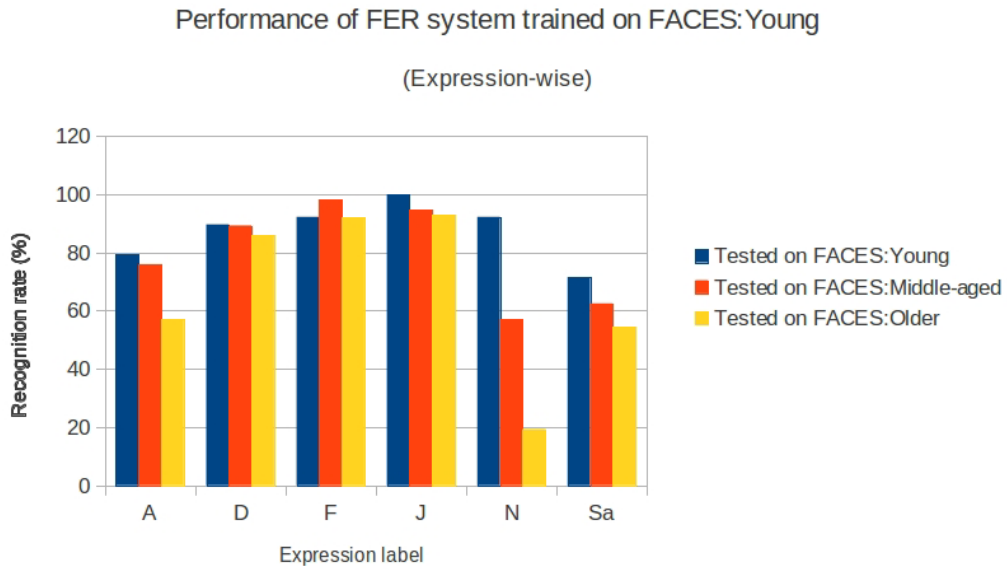


Figure 11: Expression-wise performance of facial expression recognition systems trained on FACES:Young and tested on young, middle-aged and older age-groups. The recognition rates were taken from table 19.

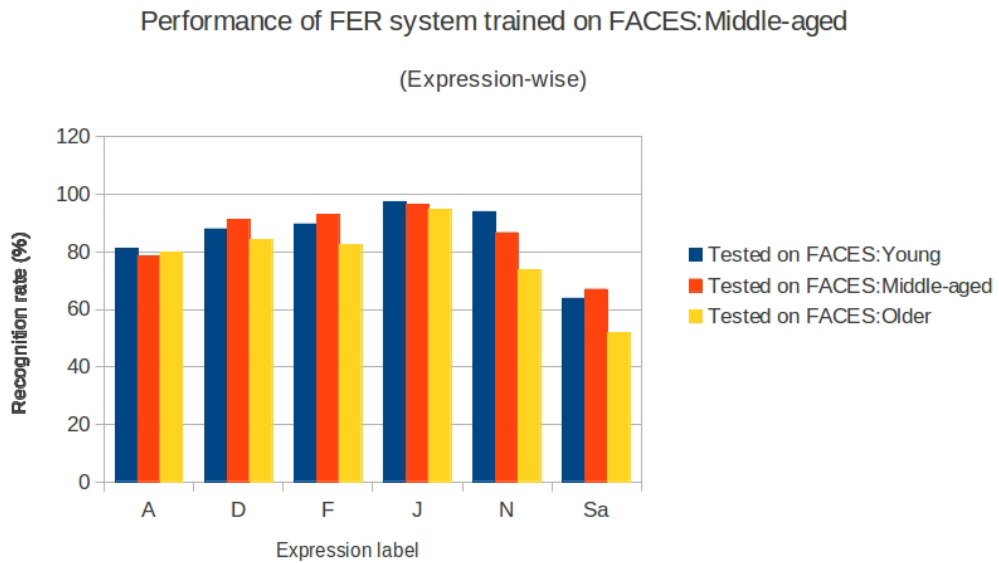


Figure 12: Expression-wise performance of facial expression recognition systems trained on FACES:Middle-aged and tested on young, middle-aged and older age-groups. The recognition rates were taken from table 19.

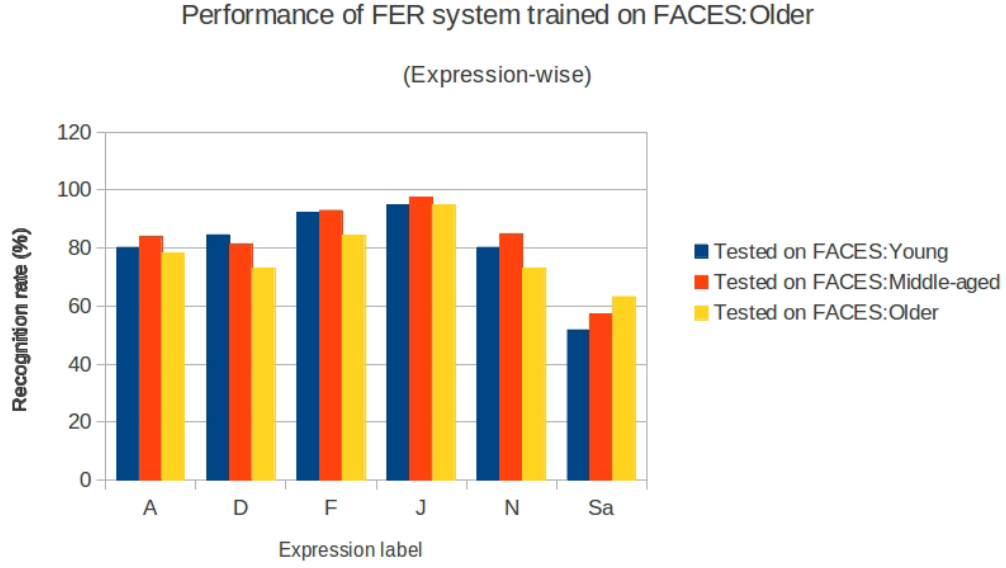


Figure 13: Expression-wise performance of facial expression recognition systems trained on FACES:Older and tested on young, middle-aged and older age-groups. The recognition rates were taken from table 19.

From the figures, we observed the following:

- From figure 11, it can be seen that the performance of the facial expression recognition system trained on FACES:Young showed drastic decline for the neutral expression on middle-aged and older faces. The neutral expression on middle-aged faces was confused greatly with sadness (15.178%), fear (10.714%) and anger (8.928%). The neutral expression on older faces was confused greatly with fear (25.438%), joy (21.052%) and sadness (19.298%). **This is yet another evidence in support of Hypothesis 2.** The system also performed very poorly in recognizing sadness on middle-aged and older faces, and in recognizing anger on older faces. Sadness on middle-aged and older faces were confused with disgust (17.857%, 14.912%) and fear (8.928%, 14.912%). Anger on older faces was confused greatly with disgust (16.67%) and sadness (12.28%). Joy, fear and disgust were easier for the system to recognize across all age-groups.
- From figure 12, it can be seen that the facial expression recognition system trained on FACES:Middle-aged performed very poorly in recognizing sadness (63.793% on young faces, 66.964% on middle-aged faces, 51.754% on older faces). Sadness on young faces was greatly confused with anger (12.068%) and neutral expression (16.379%). Sadness on older faces was confused greatly with fear (13.157%), neutral (11.403%) and disgust (10.526%). The recognition rate for neutral expression on older faces

was also low (73.684%). Joy, disgust and fear were easier for the system to recognize across all age-groups. The recognition rates for anger showed no significant variation across age-groups.

- From figure 13, it can be seen that the facial expression recognition system trained on FACES:Older also performed very poorly in recognizing sadness (51.724% on young faces, 57.142% on middle-aged faces, 63.157% on older faces). As can be seen, the recognition rates obtained for sadness increased with ascending age-group. Sadness on young faces was greatly confused with anger (21.551%) and neutral expression (18.965%). Sadness on middle-aged faces was greatly confused with anger (18.75%), disgust (12.5%) and neutral expression (8.928%). The performance of the system was more or less identical for anger, disgust, fear, joy and neutral expression displayed on young and middle-aged faces.

We also compared the inter-age-group performance (see table 18) of the facial expression recognition systems to that of humans (see table 3). The following observations were made:

- Young human raters performed significantly better than the system that was trained on FACES:Young in recognizing expressions on middle-aged and older faces.
- The performance of middle-aged human raters and the system that was trained on FACES:Middle-aged were almost identical for the young and older age-groups.
- Older human raters and the system trained on FACES:Older perform identically in recognizing expressions on young faces. However, the system outperforms older human raters in recognizing expressions on middle-aged faces.
- Young humans outperformed the system trained on FACES:Young in recognizing anger, sadness and neutral expression on middle-aged and older faces. The system outperformed young human raters in recognizing disgust and fear on middle-aged and older faces. Humans were slightly better than the system in recognizing joy on middle-aged and older faces.
- Middle-aged human raters outperformed the system in recognizing anger on young faces, joy and neutral expression on older faces, and sadness on young and older faces. The system trained on FACES:Middle-aged outperformed middle-aged human raters in recognizing anger on older faces, and disgust and fear on young and older faces. Humans were also slightly better than the system in recognizing joy on young faces.
- Older human-raters outperformed the system trained on FACES:Older in recognizing neutral expression on young faces, and sadness on young and middle-aged faces. The

system outperformed older human raters in recognizing anger, disgust and fear on young and middle-aged faces, and joy on middle-aged faces. Older human raters and the system performed almost identically in recognizing joy on young faces and neutral expression on middle-aged faces.

5.3.3.3 Conclusion

The key conclusions that can be drawn based on the inter-age-group experiments are summarized below:

- A facial expression recognition system that is trained only on young faces performs poorly on middle-aged and older age-groups.
- A facial expression recognition system that is trained only on middle-aged faces provides identical performance on young and middle-aged faces, but performs poorly on older faces.
- A facial expression recognition system that is trained only on older faces generalizes reasonably well to young and middle-aged groups.
- Ageing-related wrinkles and folds appearing on the faces of middle-aged and older humans, greatly affect the performance of a facial expression recognition system that is trained only on young faces. Such a system misinterprets the ageing-related wrinkles as expressions of emotions such as sadness and fear. The interference was greater in the case of older faces than middle-aged faces³⁷.
- Among the facial expression recognition systems examined, the one that is trained only on middle-aged faces produces the best generalization performance across the three age-groups.
- The system trained only on young faces showed an “own-age bias” in facial expression recognition.
- Recognition of expressions on older faces is the hardest among the three age-groups.
- Irrespective of the age-group, the expression of sadness is amongst the hardest for facial expression recognition systems to recognize. Recognition of sadness was hard for humans as well, but humans were better at it than the systems.
- The systems significantly outperformed humans in recognizing disgust and fear for all age-groups.

³⁷This could be due to the greater prominence and extent of ageing-related facial changes on older faces.

5.4 Scientific Results

In this work, we evaluated and compared the performance of facial expression recognition systems based on appearance-based face descriptors such as Gabor filters and LBP, across three different age-groups, namely, young, middle-aged and older. We generated evidence to establish the validity of the hypotheses stated in section 4.1, and also identified the age-group that can provide good generalization rates for all three age-groups. The main results are as follows:

- The generalization performance of a facial expression recognition system trained only on young faces is very poor on middle-aged and older age-groups [*Hypothesis 1 validated*]. The decline in performance of the system was greater for the older age-group than for the middle-aged group.
- A facial expression recognition system trained only on young faces greatly confuses the ageing-related wrinkles on neutral middle-aged and older faces with facial expressions of emotions [*Hypothesis 2 validated*].
- A facial expression recognition system trained only on young faces produced the best recognition rates for young age-group, among all systems. A facial expression recognition system trained only on middle-aged faces produced the best recognition rates for middle-aged and older age-groups, among all systems.
- The middle-aged group produces the best generalization performance across the young, middle-aged and older age-groups. A facial expression recognition system trained only on middle-aged faces produced the best overall performance, averaged over all expressions and age-groups.
- Among the three age-groups, recognizing the expressions on older faces is the hardest.
- Among the six expressions studied³⁸, recognizing the expression of joy was the easiest, whereas, recognizing sadness was amongst the hardest.

The validity of the hypotheses 1 and 2 have been established. The need for developing/identifying methods to improve facial expression recognition performance in the presence of ageing-related facial changes is revealed.

³⁸the six expressions that are included in the FACES database, namely, anger, disgust, fear, joy, neutral and sadness.

5.5 Enhancements: Suggestions, Ideas

5.5.1 Improvements to Evaluated Systems

Based on the insights gained from the state of the art and the results of the experiments performed, we suggest a few enhancements that can be made to the facial expression recognition systems evaluated in our work, that might improve their performance on all age-groups.

- The size of faces in the images in the FACES database are not uniform [36]. Since we used a fixed face model based on the eye positions to estimate the face region, some of the estimated face regions included parts of the neck and ears, as well as hair covering the upper forehead, either side of the face and/or falling behind or on the shoulders. These interfere with the feature vector extracted from the estimated face region after normalization. In order to avoid the impact of variations in face sizes, it might be better to identify numerous fiducial points on the face and then estimate the face region based on these fiducial points. This might possibly provide more uniform estimates, and in turn, improve facial expression recognition performance. Such an approach has been used in the Computer Expression Recognition Toolbox (CERT) [66] for face region estimation.
- The use of head pose estimation and face rotation techniques to correct in-plane head rotations, might improve the performance of the system not only in nearly frontal facial views, but also in tilted views. A survey and comparison of head pose estimation methods is available in [81].
- The experiments done by Shan et al. [105] showed that the recognition rates were higher when higher resolution face regions were used. In our experiments, a resolution of 48×48 pixels was used, which is quite low. Since the extraction of LBP features is not memory- and time-intensive [105], we expect that increasing the resolution of the face region would improve the overall recognition rate without heavily impacting the time and space requirements. However, higher resolutions might further reduce the generalization performance of classifiers on images of middle-aged and older age-groups, because the wrinkles and folds due to ageing would be more prominent at higher resolutions.
- In our experiments, the features were extracted from the entire face region. Since it is the region around the eyes, eyebrows and mouth that carry discriminative information about the facial expression [71, 105], extracting features only from these locations might enhance the performance of facial expression recognition systems on all age-groups, as shown by Nagi et al. in [82] on young faces.

- The expressions are almost always not pure. They are a blend of other emotions, as revealed by the semantic ratings of the images in the JAFFE database [3]. Every image in the JAFFE database was rated as containing the 6 basic expressions at varying intensities. Humans do not recognize the intensity of emotions uniformly [48, 3]. Performance varies from person-to-person [3], and depends on the age of the perceiver as well as the expresser [48]. Hence, judging the success rate of automatic expression recognition systems based on a single winner category, appears to be unrealistic. Establishing ground truth in the form of confusion matrices and/or in the form of semantic ratings such as that given to images in the JAFFE database [3], would be a more realistic alternative. The performance of a facial expression recognition system, in such cases, could be determined using statistical procedures that relate the estimates with the ground truth. One example of a system that uses the semantic ratings provided by humans is given in [132].
- The cross-database performance of facial expression recognition methods is poor. This is due to the variations in the conditions under which the images in different databases have been captured, the variations in the facial features of expressers, and variations in the way expressions are posed. Training facial expression recognition systems on a combination of multiple databases might yield more realistic estimates of performance of the algorithms, and help to identify improvements.
- The use of difference images, obtained by subtracting the neutral face from the expressive face might reduce the effect of wrinkles and folds due to ageing and enable the estimation of the true emotional state based on the facial expression. Precise face alignment would be crucial for such an approach.
- LBP feature selection using AdaBoost and classification using SVM has been found to yield better performance than using AdaBoost alone on LBP features [105]. We could use the 1-vs-rest AdaBoost learners to select discriminative features, and SVM for classification, as mentioned by Littlewort et al. in [65].
- In our experiments, we used block-wise $LBP_{8,2}^{u2}$ for feature extraction. The face region was divided into $6 \times 6 = 36$ non-overlapping blocks. The use of such blocks introduces a dependency on block-size and block location. To minimize the dependence of the LBP features on block size and block location, we could use overlapping blocks of varying sizes, such as those used by Shan et al. in [105].
- Jabid et al. [52] proposed an appearance-based face descriptor, LDP, that is robust to non-monotonic illumination changes, unlike LBP. LDP has been shown to perform better than LBP in recognizing facial expressions [53]. The use of LDP instead of

LBP might yield better facial expression recognition performance for the three age-groups.

- Lucey et al. [69] had found that concatenating two different facial representations derived from AAM improved performance for facial AU recognition. A two-layer architecture based on Gabor filters and LBP was proposed for facial AU recognition in [124], and was found to perform better than Gabor filters and LBP used individually. In the light of the findings of these studies, we expect that using a combination of multiple appearance-based face descriptors might improve the facial expression recognition performance across different age-groups.

5.5.2 Proposed Approach for Facial Expression Recognition across Age-groups

Based on the insights gained from this work, we propose an approach to handle facial expression recognition across the three age-groups. The approach is illustrated in figure 14 and is described in the paragraphs below.

After extracting the feature vector from the normalized face region in the input image, estimate the age-group to which the subject posing the expression belongs. If the age-group is estimated as “young”, use a classifier for facial expression recognition that has been trained only on young faces³⁹. For middle-aged and older age-groups, process the feature vector to reduce the contamination from ageing-related wrinkles, and use a classifier trained only on middle-aged faces for facial expression recognition⁴⁰. Identifying methods to reduce the contamination from ageing-related wrinkles is part of future research.

As can be seen, the proposed approach uses the same feature vector for age-group estimation and facial expression recognition. It may be a better alternative to use separate feature vectors for age-group estimation and expression recognition. For example, rotation invariant uniform LBP patterns for age-group estimation and block-wise uniform LBP patterns for expression recognition. An approach to estimate the age of a person even under facial expression changes was developed by Guo and Wang in [47] and could be adopted for the age-group estimation step. In figure 14, we have also shown a single control path for both middle-aged and older age-groups. It may be a better alternative to process the feature vector for the two age-groups separately, and use separate classifiers for expression recognition.

³⁹We have seen that a facial expression recognition system that is trained only on young faces yields the best performance for the young age-group, among all systems.

⁴⁰We have seen that a facial expression recognition system trained only on middle-aged faces performs best on middle-aged group, and generalizes well to older age-groups.

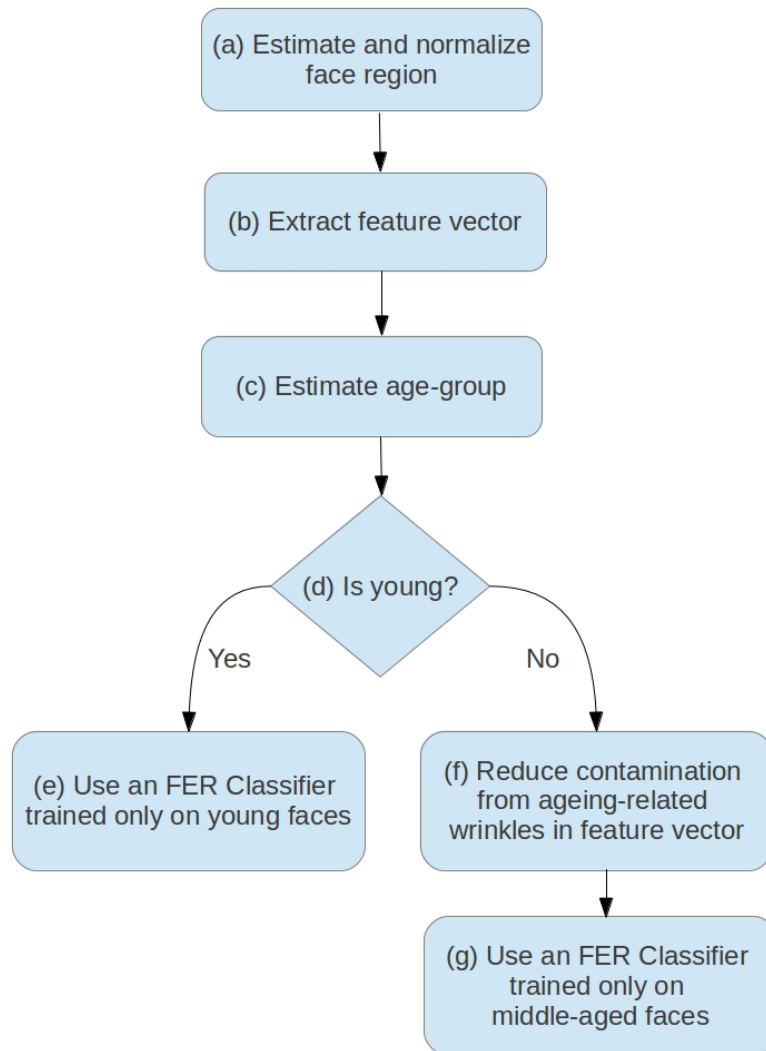


Figure 14: Proposed approach for facial expression recognition across young, middle-aged and older age-groups.

The age-group estimation step can be excluded and a single classifier trained on images of subjects belonging to all age-groups can be used, if multi-layer architectures such as Bag of Words with scale-invariant features [108] are found to have age-group-invariance property.

The approach that has been proposed is based on the insights gained from this project, and has not been experimentally validated.

6 Comparison of Human and System Performance

In this section, we summarize the results of comparison of performance of humans and that of the facial expression recognition systems evaluated in our work.

The results of the comparison of performance of human raters on FACES database, and the performance of facial expression recognition systems based on block-wise $LBP_{8,2}^{u2}$ and trained separately on young, middle-aged and older faces in FACES database are given in symbolic form in table 20. The table is self-explanatory.

In section 2.4, we have looked at some of the findings of neurological, psychological and behavioural research on performance of humans in recognizing emotions conveyed through facial expressions. In table 21, we identify which of these findings hold true also for facial expression recognition systems that use appearance-based face descriptors such as Gabor wavelets and LBP.

Young human raters v/s System trained on FACES:Young							
Performance on FACES:	Anger	Disgust	Fear	Joy	Neutral	Sadness	Overall
Young	H	S	S	s	h	H	≈
Middle-aged	H	S	S	H	H	H	H
Older	H	S	S	h	H	H	H
Overall	H	S	S	≈	H	H	H

Middle-aged human raters v/s System trained on FACES:Middle-aged							
Performance on FACES:	Anger	Disgust	Fear	Joy	Neutral	Sadness	Overall
Young	H	S	S	≈	s	H	≈
Middle-aged	H	S	S	h	h	h	s
Older	S	S	S	h	H	H	s
Overall	H	S	S	h	h	H	s

Older human raters v/s System trained on FACES:Older							
Performance on FACES:	Anger	Disgust	Fear	Joy	Neutral	Sadness	Overall
Young	s	S	S	≈	H	H	≈
Middle-aged	S	S	S	s	≈	H	S
Older	S	S	S	s	H	≈	S
Overall	S	S	S	s	H	H	S

Table 20: Performance comparison between humans and facial expression recognition systems based on block-wise $LBP_{8,2}^{u2}$ that were trained and tested on young, middle-aged and older faces in FACES database. Symbolic coding scheme: If the absolute value of difference in recognition rates between humans and systems $> 3\%$, then ‘H’ or ‘S’ is used to indicate whether humans or system performs better. If the absolute value of the difference is between 1% and 3%, then ‘h’ or ‘s’ is used to indicate whether humans or system performs better. If the absolute value of the difference is less than 1%, then the performance of humans and systems are considered to be almost identical, and is indicated by ‘≈’

Findings of reasearch on human performance	Observed in FER systems?
Young human raters performed better than middle-aged human raters on FACES database [36]	N
Middle-aged human raters performed better than older human raters on FACES database	N
Older human raters performed poorly in decoding emotions [76] and facial expressions [100]	N
Young humans confused neutral expression on older faces with non-neutral emotions [48]. The confusion was much higher compared to that for neutral young faces [48].	Y
Young humans perceived expressions on older faces less clearly [48]	Y
Humans showed “own-age-bias” in facial expression recognition [76]	P ^a
Humans have difficulty in recognizing the expression of fear in JAFFE database [131]	Y
Expression of joy is the easiest for humans to recognize [92, 28, 36]	Y
Expression of disgust is the hardest for humans to recognize in FACES database [36]	N
Sadness is hard for humans to recognize [92, 100, 36]	Y
Recognizing expressions on young faces is the easiest and on older faces is the hardest in FACES database [36]	Y

^aonly for young age-group.

Table 21: Some of the findings of research on human performance in recognizing emotions from facial expressions are listed in the first column. In the second column, we state whether a similar performance trend was observed for facial expression recognition systems studied in this work. ‘Y’ stands for “yes, observed”; ‘N’ stands for “no, not observed”; ‘P’ stands for “partial- applies to some, but not to all three age-groups”.

7 Summary

Existing approaches for facial expression and facial action unit recognition have been validated on facial images of young people. How these systems perform on facial images of middle-aged and older people has so far not been studied. In this work, the impact of ageing-related facial changes on the performance of facial expression recognition systems based on Gabor filters and uniform LBP was examined. The systems were trained and tested on the FACES database, which contains images of young, middle-aged and older face models showing the neutral expression and the expressions of anger, disgust, fear, joy and sadness. A degradation in performance of the systems trained only on young faces was observed when they were tested on middle-aged and older age-groups. The system greatly confused neutral expression on middle-aged and older faces with other expressions. These results show that ageing-related facial changes interfere with facial expression recognition.

Among the three age-groups, overall facial expression recognition performance on the older age-group was the lowest. Among the six expressions, the recognition rates achieved for joy was the highest, and that for sadness was amongst the lowest. Of all systems, the one trained only on middle-aged faces generalized well across the entire age spectrum. The performance of the systems was also compared to that of humans and certain similarities and dissimilarities were observed. While humans were better at recognizing sadness, the systems were better at recognizing disgust and fear. Improvements to the evaluated facial expression recognition systems were suggested, and a general approach was proposed for improving facial expression recognition performance on middle-aged and older age-groups. Reduction of the presence of ageing-related facial wrinkles in the feature vectors extracted for facial expression recognition is the central idea of the proposed approach, and techniques to realize it need to be identified.

The results of our work can motivate future research in the direction of developing age-independent facial expression recognition systems.

8 Future Work

In this section, we identify the open items, which need to be addressed in future research, and tasks that can be performed as a continuation of this work.

- In this work, we have considered only 5 of the 6 basic expressions. The expression of surprise was not available in the FACES database, and hence was not included. The performance of facial expression recognition systems in recognizing surprise on faces of subjects belonging to different age-groups need to be studied. However, in

order to do this, images of subjects belonging to the three different age-groups and displaying the expression of surprise should be compiled.

- An empirical evaluation of geometric and appearance-based descriptors, and their combinations, for facial expression recognition on the three age-groups should be performed, so as to identify which face descriptor(s) are best for each age-group.
- Wrinkles and folds that appear on the face due to ageing are included in the appearance-based feature vector representing the facial expression, as seen from the experiments described in section 5.3. The micro-patterns that characterize facial ageing should be examined and ways to minimize its impact on facial expression recognition should be identified, so that the recognition rates for the older and middle-aged groups can be further improved. In order to do this, the feature vectors extracted from neutral and expressive faces of subjects should be analysed, for all age-groups and all expressions.
- Experimental validation of the enhancements suggested in section 5.5 need to be performed.
- Image sequences contain more information about expressive facial actions [55]. Therefore, dynamic and spatio-temporal approaches based on hybrid face descriptors might improve performance across age-groups. The lack of a benchmark database of image sequences in which both the facial expression as well as the age of the expresser are systematically varied, presents a hurdle to investigating this option in the near future.

References

- [1] Cohn-Kanade AU-Coded Facial Expression Database. <http://www.pitt.edu/~jeffcohn/CKandCK+.htm>. [Online; accessed 26-Aug-2013].
- [2] FACES. <http://faces.mpg.de/faces/>. [Online; accessed 19-Aug-2013].
- [3] JAFFE Information. http://www.kasrl.org/jaffe_info.html. [Online; accessed 18-Aug-2013].
- [4] MMI Facial Expression Database. <http://www.mmifacedb.com/>. [Online; accessed 27-Aug-2013].
- [5] MultiBoost. <http://www.multiboost.org/>. [Online; accessed 17-Sep-2013].
- [6] Pictures of Facial Affect (POFA). <http://www.paulekman.com/product/pictures-of-facial-affect-pofa/>. [Online; accessed 19-Aug-2013].

- [7] The Japanese Female Facial Expression (JAFPE) Database. <http://www.kasrl.org/jaffe.html>. [Online; accessed 18-Aug-2013].
- [8] FaceGenModeller (version 3.2). <http://www.facegen.com/>, 2007. [computer software].
- [9] Guaita A., Malnati M., Vaccaro R., Pezzati R., Marcionetti J., Vitali S.F., and Colombo M. Impaired facial emotion recognition and preserved reactivity to facial expressions in people with severe dementia. *Archive of Gerontology and Geriatrics*, 49:135–146, 2009.
- [10] Ralph Adolphs. Recognizing emotion from facial expressions: Psychological and neurological mechanisms. *Behavioral and cognitive neuroscience reviews*, 1(1):21–62, 2002.
- [11] Ralph Adolphs, Daniel Tranel, Hanna Damasio, and Antonio Damasio. Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. *Nature*, 372(6507):669–672, 1994.
- [12] T. Ahonen, A. Hadid, and M. Pietikainen. Face description with local binary patterns: Application to face recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(12):2037–2041, 2006.
- [13] Zara Ambadar, Jeffrey F. Cohn, and Lawrence Ian Reed. All smiles are not created equal: Morphology and timing of smiles perceived as amused, polite, and embarrassed/nervous. *Journal of Nonverbal Behavior*, 33(1):17–34, 2009.
- [14] Jeffrey S. Anastasi and Matthew G. Rhodes. Evidence for an own-age bias in face recognition. *North American Journal of Psychology*, 2006.
- [15] C. Anitha, M.K. Venkatesha, and B. Suryanarayana Adiga. A survey on facial expression databases. *International Journal of Engineering Science and Technology*, 2(10):5158–5174, 2010.
- [16] Department of Computer Science at University of Oulu and Engineering. LBP Matlab code. <http://www.cse.oulu.fi/CMV/Downloads/LBPMatlab>. [Online; accessed 16-Sep-2013].
- [17] b-it-bots Team. ROS Stacks for RoboCup@Home. <https://github.com/b-it-bots/robocup-at-home>. [Online; accessed 4-Sep-2013].
- [18] Marian Stewart Bartlett, Gwen C. Littlewort, Mark G. Frank, Claudia Lainscsek, Ian R. Fasel, and Javier R. Movellan. Automatic recognition of facial actions in spontaneous expressions. *Journal of Multimedia*, 1(6):22–35, 2006.

- [19] M.S. Bartlett, Gwen Littlewort, M. Frank, C. Lainscsek, Ian Fasel, and J. Movellan. Recognizing facial expression: machine learning and application to spontaneous behavior. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 2, pages 568–573 vol. 2, 2005.
- [20] M.S. Bartlett, Gwen Littlewort, M. Frank, C. Lainscsek, Ian Fasel, and J. Movellan. Recognizing facial expression: machine learning and application to spontaneous behavior. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 2, pages 568–573 vol. 2, 2005.
- [21] M.S. Bartlett, Gwen Littlewort, M. Frank, C. Lainscsek, Ian Fasel, and J. Movellan. Fully automatic facial action recognition in spontaneous behavior. In *Automatic Face and Gesture Recognition, 2006. FGR 2006. 7th International Conference on*, pages 223–230, 2006.
- [22] Djalel Benbouzid, Róbert Busa-Fekete, Norman Casagrande, François-David Collin, and Balázs Kégl. MultiBoost: a multi-purpose boosting package. *Journal of Machine Learning Research*, 13(1):549–553, March 2012.
- [23] Vinay Bettadapura. Face expression recognition and analysis: the state of the art. *arXiv preprint arXiv:1203.6722*, 2012.
- [24] Deborah Bier. Improving Alzheimers and Dementia Care: Emotions Rule. <http://psychcentral.com/lib/2012/improving-alzheimers-and-dementia-care-emotions-rule/all/1/>, 2012. [Online]. Accessed: Apr. 28, 2013.
- [25] M.J. Black and Y. Yacoob. Tracking and recognizing rigid and non-rigid facial motions using local parametric models of image motion. In *Computer Vision, 1995. Proceedings., Fifth International Conference on*, pages 374–381, 1995.
- [26] Cynthia Breazeal. Emotion and sociable humanoid robots. *International Journal of Human-Computer Studies*, 59(1):119–155, 2003.
- [27] Carlos Busso, Zhigang Deng, Serdar Yildirim, Murtaza Bulut, Chul Min Lee, Abe Kazemzadeh, Sungbok Lee, Ulrich Neumann, and Shrikanth Narayanan. Analysis of emotion recognition using facial expressions, speech and multimodal information. In *Proceedings of the 6th international conference on Multimodal interfaces, ICMI '04*, pages 205–211, New York, NY, USA, 2004. ACM.
- [28] Manuel G. Calvo and Daniel Lundqvist. Facial expressions of emotion (KDEF): Identification under different display-duration conditions. *Behavior Research Methods*, 40(1):109–115, 2008.

- [29] George Caridakis, Ginevra Castellano, Loic Kessous, Amaryllis Raouzaïou, Lori Malatesta, Stelios Asteriadis, and Kostas Karpouzis. Multimodal emotion recognition from expressive faces, body gestures and speech. In Christos Boukis, Aristodemios Pnevmatikakis, and Lazaros Polymenakos, editors, *Artificial Intelligence and Innovations 2007: from Theory to Applications*, volume 247 of *IFIP The International Federation for Information Processing*, pages 375–388. Springer US, 2007.
- [30] James M. Carroll and James A. Russell. Facial expressions in hollywood’s portrayal of emotion. *Journal of Personality and Social Psychology*, 72(1):164–176, 1997.
- [31] Modesto Castrillón, Oscar Déniz, Daniel Hernández, and Javier Lorenzo. A comparison of face and facial feature detectors based on the viola-jones general object detection framework. *Machine Vision and Applications*, 22(3):481–494, 2011.
- [32] Susan Turk Charles, Mara Mather, and Laura L Carstensen. Aging and emotional memory: the forgettable nature of negative images for older adults. *Journal of Experimental Psychology: General*, 132(2):310, 2003.
- [33] Ira Cohen, Nicu Sebe, Ashutosh Garg, Lawrence S. Chen, and Thomas S. Huang. Facial expression recognition from video sequences: temporal and static modeling. *Computer Vision and Image Understanding*, 91(1-2):160 – 187, 2003.
- [34] J.F. Cohn, L.I. Reed, Z. Ambadar, Jing Xiao, and T. Moriyama. Automatic analysis and recognition of brow actions and head motion in spontaneous facial behavior. In *Systems, Man and Cybernetics, 2004 IEEE International Conference on*, volume 1, pages 610–616 vol.1, 2004.
- [35] G. Donato, M.S. Bartlett, J.C. Hager, P. Ekman, and T.J. Sejnowski. Classifying facial actions. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 21(10):974–989, 1999.
- [36] Natalie C. Ebner, Michaela Riediger, and Ulman Lindenberger. FACES—A database of facial expressions in young, middle-aged, and older women and men: Development and validation. *Behavior Research Methods*, 42(1):351–362, 2010.
- [37] Micah Eckhardt, Ian Fasel, and Javier Movellan. Towards practical facial feature detection. *International Journal of Pattern Recognition and Artificial Intelligence*, 23(03):379–400, 2009.
- [38] Irenäus Eibl-Eibesfeldt. *Human Ethology*. Aldine de Gruyter, New York, 1989.
- [39] P. Ekman and W. V Friesen. Pictures of facial affect. Human Interaction Laboratory, University of California Medical Center, San Francisco, CA, 1976.

- [40] Paul Ekman and Wallace V Friesen. Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, 17(2), 1971.
- [41] Paul Ekman and Wallace V Friesen. *Unmasking the face: A guide to recognizing emotions from facial clues*. Ishk, 2003.
- [42] Paul Ekman and Erika L. Rosenberg, editors. *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*. Oxford University Press, 2005.
- [43] I.A. Essa and A.P. Pentland. Coding, analysis, interpretation, and recognition of facial expressions. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 19(7):757–763, 1997.
- [44] B. Fasel and Juergen Luetttin. Automatic facial expression analysis: a survey. *Pattern Recognition*, 36(1):259 – 275, 2003.
- [45] D. Gabor. Theory of communication. *Journal of the Institution of Electrical Engineers*, 93(26):429–457, November 1946.
- [46] Ralph Gross. Face databases. In *Handbook of Face Recognition*, pages 301–327. Springer, 2005.
- [47] Guodong Guo and Xiaolong Wang. A study on human age estimation under facial expression changes. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 2547–2553, 2012.
- [48] Ursula Hess, Reginald B. Adams Jr., Annie Simard, Michael T. Stevenson, and Robert E. Kleck. Smiling and sad wrinkles: Age-related changes in the face and the perception of emotions and intentions. *Journal of Experimental Social Psychology*, 48(6):1377 – 1380, 2012.
- [49] A. Higashi, T. Yasui, Y. Fukumizu, and H. Yamauchi. Local gabor directional pattern histogram sequence (lgdphs) for age and gender classification. In *Statistical Signal Processing Workshop (SSP), 2011 IEEE*, pages 505–508, 2011.
- [50] Di Huang, Caifeng Shan, M. Ardabilian, Yunhong Wang, and Liming Chen. Local binary patterns and its application to facial image analysis: A survey. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 41(6):765–781, 2011.
- [51] Derek M Isaacowitz, Heather A Wadlinger, Deborah Goren, and Hugh R Wilson. Is there an age-related positivity effect in visual attention? A comparison of two methodologies. *Emotion*, 6(3):511, 2006.

- [52] T. Jabid, M.H. Kabir, and O. Chae. Local directional pattern (ldp) for face recognition. In *Consumer Electronics (ICCE), 2010 Digest of Technical Papers International Conference on*, pages 329–330, 2010.
- [53] T. Jabid, M.H. Kabir, and Oksam Chae. Facial expression recognition using local directional pattern (LDP). In *Image Processing (ICIP), 2010 17th IEEE International Conference on*, pages 1605–1608, 2010.
- [54] T. Jabid, M.H. Kabir, and Oksam Chae. Gender classification using local directional pattern (ldp). In *Pattern Recognition (ICPR), 2010 20th International Conference on*, pages 2162–2165, 2010.
- [55] Bihan Jiang, M.F. Valstar, and M. Pantic. Action unit detection using sparse appearance descriptors in space-time video volumes. In *Automatic Face Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*, pages 314–321, 2011.
- [56] Md Hasanul Kabir, Taskeed Jabid, and Oksam Chae. Local directional pattern variance (LDPv): A robust feature descriptor for facial expression recognition. *Int. Arab J. Inf. Technol.*, 9(4):382–391, 2012.
- [57] M.H. Kabir, T. Jabid, and Oksam Chae. A local directional pattern variance (LDPv) based face descriptor for human facial expression recognition. In *Advanced Video and Signal Based Surveillance (AVSS), 2010 Seventh IEEE International Conference on*, pages 526–532, 2010.
- [58] T. Kanade, J.F. Cohn, and YingLi Tian. Comprehensive database for facial expression analysis. In *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on*, pages 46–53, 2000.
- [59] MASE Kenji. Recognition of facial expression from optical flow. *IEICE TRANSACTIONS on Information and Systems*, 74(10):3474–3483, 1991.
- [60] Martin Knapp, M. Prince, E. Albanese, S. Banerjee, S. Dhanasiri, J.L. Fernandez, C. Ferri, P. McCrone, T. Snell, and R. Stewart. Dementia UK—A report into the prevalence and cost of dementia prepared by the Personal Social Services Research Unit (PSSRU) at the London School of Economics and the Institute of Psychiatry at King’s College London, for the Alzheimer’s Society. *THE FULL REPORT*, 2007.
- [61] S. Koelstra, M. Pantic, and I. Patras. A dynamic texture-based approach to recognition of facial actions and their temporal models. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(11):1940–1954, 2010.

- [62] Szidónia Lefkovits. Assessments of building classifiers for face detection. In *International Conference On Recent Achievements in Mechatronics, Automation, Computer Sciences and Robotics MACRO*, pages 175–186, 2009.
- [63] R. Levenson. Human emotions: A functional view. In P. Ekman and R. Davidson, editors, *The Nature of Emotion*, pages 123–126. Oxford University Press, New York, NY, 1994.
- [64] Michael Lewis. The emergence of human emotions. *Handbook of emotions*, 2:265–280, 2000.
- [65] Gwen Littlewort, Marian Stewart Bartlett, Ian Fasel, Joshua Susskind, and Javier Movellan. Dynamics of facial expression extracted automatically from video. *Image and Vision Computing*, 24(6):615 – 625, 2006.
- [66] Gwen Littlewort, J. Whitehill, Tingfan Wu, Ian Fasel, M. Frank, J. Movellan, and M. Bartlett. The computer expression recognition toolbox (cert). In *Automatic Face Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*, pages 298–305, 2011.
- [67] A Lobo, L.J. Launer, L. Fratiglioni, K. Andersen, A. Di Carlo, M.M.B. Breteler, J.R.M. Copeland, J-F Dartigues, C. Jagger, J. Martinez-Lage, et al. Prevalence of dementia and major subtypes in Europe: a collaborative study of population-based cohorts. *Neurology*, 54(11):S4–S9, 2000.
- [68] P. Lucey, J.F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews. The extended Cohn-Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, pages 94–101, 2010.
- [69] Simon Lucey, Ahmed Bilal Ashraf, and Jeffrey Cohn. Investigating spontaneous facial action recognition through AAM representations of the face. *Face recognition*, pages 275–286, 2007.
- [70] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba. Coding facial expressions with Gabor wavelets. In *Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on*, pages 200–205, 1998.
- [71] Michael J. Lyons, J. Budynek, and S. Akamatsu. Automatic classification of single facial images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 21(12):1357–1362, 1999.

- [72] Michael J. Lyons, Miyuki Kamachi, and Jiro Gyoba. Japanese Female Facial Expressions (JAFFE), Database of digital images. 1997.
- [73] Bingpeng Ma, Wenchao Zhang, Shiguang Shan, Xilin Chen, and Wen Gao. Robust head pose estimation using lgbp. In *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, volume 2, pages 512–515, 2006.
- [74] Geovanny Rafael Giorgana Macedo. Facial expression recognition from video sequences using spatial and spatiotemporal face descriptors. Master’s thesis, University of Applied Sciences Bonn-Rhein-Sieg, Sankt Augustin, Germany, February 2011.
- [75] C. Z. Malatesta, M. J. Fiore, and J. J. Messina. Affect, personality, and facial expressive characteristics of older people. *Psychology and Aging*, 2(1):64–69, 1987.
- [76] Carol Z. Malatesta, Carol E. Izard, Clayton Culver, and Mark Nicolich. Emotion communication skills in young, middle-aged, and older women. *Psychology and Aging*, 2(2):193–203, 1987.
- [77] A. M. Martinez and R. Benavente. The AR Face Database, CVC Technical Report#24. Technical report, June 1998.
- [78] Mara Mather and Laura L Carstensen. Aging and attentional biases for emotional faces. *Psychological Science*, 14(5):409–415, 2003.
- [79] Christoph Mayer and Frank Wallhoff. Metadata for FEED database. <http://www.mmk.ei.tum.de/~waf/fgnet/metadata-feedtum.csv>. [Online; accessed 28-Sep-2013].
- [80] A. Mehrabian. Communication without words. *Psychology Today*, 2(4):53–56, 1968.
- [81] E. Murphy-Chutorian and M.M. Trivedi. Head pose estimation in computer vision: A survey. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(4):607–626, 2009.
- [82] Gawed M Nagi, Rahmita Wirza OK Rahmat, Fatimah Khalid, and Muhamad Taufik Abdullah. Region-based facial expression recognition in still images. *JIPS*, 9(1):173, 2013.
- [83] United Nations. Wall Chart on Population Ageing and Development 2012. http://www.un.org/esa/population/publications/2012PopAgeingDev_Chart/2012AgeingWallchart.html, November 2012. [Online; accessed 25-Sep-2013].

- [84] Navid Nourani-Vatani. LBP. <https://github.com/nourani/LBP>. [Online; accessed 16-Sep-2013].
- [85] T. Ojala. Multichannel approach to texture description with feature distributions, Technical Report CAR-TR-846. Technical report, Center for Automation Research, University of Maryland, 1996.
- [86] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(7):971–987, 2002.
- [87] Timo Ojala, Matti Pietikäinen, and David Harwood. A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition*, 29(1):51 – 59, 1996.
- [88] Ville Ojansivu and Janne Heikkilä. Blur insensitive texture classification using local phase quantization. In Abderrahim Elmoataz, Olivier Lezoray, Fathallah Nouboud, and Driss Mammass, editors, *Image and Signal Processing*, volume 5099 of *Lecture Notes in Computer Science*, pages 236–243. Springer Berlin Heidelberg, 2008.
- [89] Andrew Ortony and Terence J. Turner. What’s basic about basic emotions? *Psychological Review*, 97(3):315–331, July 1990.
- [90] T. Otsuka and Jun Ohya. Recognizing multiple persons’ facial expressions using hmm based on automatic extraction of significant frames from image sequences. In *Image Processing, 1997. Proceedings., International Conference on*, volume 2, pages 546–549 vol.2, 1997.
- [91] Curtis Padgett and Garrison Cottrell. Identifying emotion in static face images. In *Proceedings of the 2nd Joint Symposium on Neural Computation*, volume 5, pages 91–101. Citeseer, 1995.
- [92] Romina Palermo and Max Coltheart. Photographs of facial expression: Accuracy, response times, and ratings of intensity. *Behavior Research Methods, Instruments, & Computers*, 36(4):634–638, 2004.
- [93] M. Pantic and I. Patras. Detecting facial actions and their temporal segments in nearly frontal-view face image sequences. In *Systems, Man and Cybernetics, 2005 IEEE International Conference on*, volume 4, pages 3358–3363 Vol. 4, 2005.
- [94] M. Pantic and L. J M Rothkrantz. Automatic analysis of facial expressions: the state of the art. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(12):1424–1445, 2000.

- [95] M. Pantic and L. J M Rothkrantz. Toward an affect-sensitive multimodal human-computer interaction. *Proceedings of the IEEE*, 91(9):1370–1390, 2003.
- [96] M. Pantic, M. Valstar, R. Rademaker, and L. Maat. Web-based database for facial expression analysis. In *Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on*, pages 5 pp.–, 2005.
- [97] M Pietikäinen, T Ojala, and Z Xu. Rotation-invariant texture classification using feature distributions. *Pattern Recognition*, 33(1):43 – 52, 2000.
- [98] M. Prince, R. Bryce, E. Albanese, A. Wimo, W. Ribeiro, and C.P. Ferri. The global prevalence of dementia: a systematic review and metaanalysis. *Alzheimer’s & Dementia*, 9(1):63–75, January 2013.
- [99] M. Rosenblum, Y. Yacoob, and L.S. Davis. Human expression recognition from motion using a radial basis function network architecture. *Neural Networks, IEEE Transactions on*, 7(5):1121–1138, 1996.
- [100] Ted Ruffman, Julie D. Henry, Vicki Livingstone, and Louise H. Phillips. A meta-analytic review of emotion recognition and aging: Implications for neuropsychological models of aging. *Neuroscience & Biobehavioral Reviews*, 32(4):863 – 881, 2008.
- [101] Modesto Castrillón Santana, Oscar Déniz-Suárez, Luis Antón-Canalís, and Javier Lorenzo-Navarro. Face and facial feature detection evaluation - performance evaluation of public domain haar detectors for face and facial feature detection. In Alpesh Ranchordas and Helder Araújo, editors, *VISAPP (2)*, pages 167–172. INSTICC - Institute for Systems and Technologies of Information, Control and Communication, 2008.
- [102] B. Schuller, M. Lang, and G. Rigoll. Multimodal emotion recognition in audiovisual communication. In *Multimedia and Expo, 2002. ICME ’02. Proceedings. 2002 IEEE International Conference on*, volume 1, pages 745–748 vol.1, 2002.
- [103] N. Sebe, M.S. Lew, Y. Sun, I. Cohen, T. Gevers, and T.S. Huang. Authentic facial expression analysis. *Image and Vision Computing*, 25(12):1856 – 1863, 2007.
- [104] Nicu Sebe, Ira Cohen, and Thomas S. Huang. Multimodal emotion recognition. *Handbook of Pattern Recognition and Computer Vision*, 4:387–419, 2005.
- [105] Caifeng Shan, Shaogang Gong, and Peter W. McOwan. Facial expression recognition based on local binary patterns: A comprehensive study. *Image and Vision Computing*, 27(6):803 – 816, 2009.

- [106] Phillip Shaver, Judith Schwartz, Donald Kirson, and Cary O’connor. Emotion knowledge: further exploration of a prototype approach. *Journal of Personality and Social Psychology*, 52(6):1061–1086, June 1987.
- [107] Frank Y. Shih, Chao-Fa Chuang, and Patrick S. P. Wang. Performance comparisons of facial expression recognition in JAFFE database. *International Journal of Pattern Recognition and Artificial Intelligence*, 22(03):445–459, 2008.
- [108] Karan Sikka, Tingfan Wu, Josh Susskind, and Marian Bartlett. Exploring bag of words architectures in the facial expression domain. In Andrea Fusiello, Vittorio Murino, and Rita Cucchiara, editors, *Computer Vision - ECCV 2012. Workshops and Demonstrations*, volume 7584 of *Lecture Notes in Computer Science*, pages 250–259. Springer Berlin Heidelberg, 2012.
- [109] Ying-Li Tian. Evaluation of face resolution for expression analysis. In *Computer Vision and Pattern Recognition Workshop, 2004. CVPRW ’04. Conference on*, pages 82–82, 2004.
- [110] Ying-Li Tian, T. Kanade, and J.F. Cohn. Recognizing action units for facial expression analysis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(2):97–115, 2001.
- [111] Yan Tong, Wenhui Liao, and Qiang Ji. Facial action unit recognition by exploiting their dynamic and semantic relationships. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(10):1683–1699, 2007.
- [112] Michal Uříčář and Vojtěch Franc. flandmark - open-source implementation of facial landmark detector. <http://cmp.felk.cvut.cz/~uricamic/flandmark/>. [Online; accessed 02-Sep-2013].
- [113] Michal Uříčář, Vojtěch Franc, and Václav Hlaváč. Detector of facial landmarks learned by the structured output SVM. In Gabriela Csurka and José Braz, editors, *VISAPP ’12: Proceedings of the 7th International Conference on Computer Vision Theory and Applications*, volume 1, pages 547–556, Portugal, February 2012. SciTePress — Science and Technology Publications.
- [114] M. Valstar and M. Pantic. Fully automatic facial action unit detection and temporal analysis. In *Computer Vision and Pattern Recognition Workshop, 2006. CVPRW ’06. Conference on*, pages 149–149, 2006.
- [115] M.F. Valstar and M. Pantic. Fully automatic recognition of the temporal phases of facial actions. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 42(1):28–43, 2012.

- [116] M.F. Valstar, I. Patras, and M. Pantic. Facial action unit detection using probabilistic actively learned support vector machines on tracked facial point data. In *Computer Vision and Pattern Recognition - Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on*, pages 76–76, 2005.
- [117] M.F. Valstar, I. Patras, and M. Pantic. Facial action unit detection using probabilistic actively learned support vector machines on tracked facial point data. In *Computer Vision and Pattern Recognition - Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on*, pages 76–76, 2005.
- [118] Michel Valstar and Maja Pantic. Induced disgust, happiness and surprise: an addition to the MMI facial expression database. In *Proc. Int’l Conf. Language Resources and Evaluation, W’shop on EMOTION*, pages 65–70, 2010.
- [119] Michel F. Valstar, Hatice Gunes, and Maja Pantic. How to distinguish posed from spontaneous smiles using geometric features. In *Proceedings of the 9th international conference on Multimodal interfaces, ICMI ’07*, pages 38–45, New York, NY, USA, 2007. ACM.
- [120] Paul Viola and MichaelJ. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154, 2004.
- [121] J. Wagner, E. Andre, and F. Jung. Smart sensor integration: A framework for multimodal emotion recognition in real-time. In *Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on*, pages 1–8, 2009.
- [122] Frank Wallhoff. Facial Expressions and Emotion Database. <http://www.mmk.ei.tum.de/~waf/fgnet/feedtum.html>. Technische Universität München, 2006. [Online; accessed 28-Aug-2013].
- [123] Li Wang and Dong-Chen He. Texture classification using texture spectrum. *Pattern Recognition*, 23(8):905 – 910, 1990.
- [124] Tingfan Wu, N.J. Butko, P. Ruvolo, J. Whitehill, M.S. Bartlett, and Javier R. Movellan. Multilayer architectures for facial action unit recognition. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 42(4):1027–1038, 2012.
- [125] Tianwei Xu, Juxiang Zhou, and Yunqiong Wang. A variation of local directional pattern and its application for facial expression recognition. In Tai-hoon Kim, Hjjat Adeli, Carlos Ramos, and Byeong-Ho Kang, editors, *Signal Processing, Image Processing and Pattern Recognition*, volume 260 of *Communications in Computer and Information Science*, pages 36–47. Springer Berlin Heidelberg, 2011.

- [126] Y. Yacoob and L.S. Davis. Recognizing human facial expressions from long image sequences using optical flow. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 18(6):636–642, 1996.
- [127] Lijun Yin, Xiaozhou Wei, Yi Sun, Jun Wang, and M.J. Rosato. A 3d facial expression database for facial behavior research. In *Automatic Face and Gesture Recognition, 2006. FGR 2006. 7th International Conference on*, pages 211–216, 2006.
- [128] Z. Zeng, M. Pantic, G.I. Roisman, and T.S. Huang. A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(1):39–58, 2009.
- [129] Guangcheng Zhang, Xiangsheng Huang, StanZ. Li, Yangsheng Wang, and Xihong Wu. Boosting local binary pattern (lbp)-based face recognition. In StanZ. Li, Jianhuang Lai, Tieniu Tan, Guocan Feng, and Yunhong Wang, editors, *Advances in Biometric Person Authentication*, volume 3338 of *Lecture Notes in Computer Science*, pages 179–186. Springer Berlin Heidelberg, 2005.
- [130] Wenchao Zhang, Shiguang Shan, Wen Gao, Xilin Chen, and Hongming Zhang. Local gabor binary pattern histogram sequence (lgbphs): a novel non-statistical model for face representation and recognition. In *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, volume 1, pages 786–791 Vol. 1, 2005.
- [131] Zhengyou Zhang, M. Lyons, M. Schuster, and S. Akamatsu. Comparison between geometry-based and gabor-wavelets-based facial expression recognition using multi-layer perceptron. In *Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on*, pages 454–459, 1998.
- [132] Wenming Zheng, Xiaoyan Zhou, Cairong Zou, and Li Zhao. Facial expression recognition using kernel canonical correlation analysis (KCCA). *Neural Networks, IEEE Transactions on*, 17(1):233–238, 2006.

Appendices

A Facial Action Coding System (FACS)

FACS is a comprehensive coding system for describing facial muscle activity, which enables human observers to code facial expressions. It was created in its current form by Ekman, Friesen and Hager [42]. Each atomic facial muscle action is called an action unit (AU). Each AU is given a unique numeric code. There are 44 single AUs in all. AUs may be displayed either alone or in combination with other AUs. FACS also allows the intensity of an AU to be rated on a five-point scale. In addition to coding facial muscle activities, FACS also defines codes for head and eye movements [42]. Ekman and Friesen also developed Emotion FACS (EMFACS), which includes only those AUs defined in FACS that are relevant for facial expressions of emotions. EMFACS is used by behavioural scientists to interpret emotions from facial actions. A detailed description of how to recognize emotions by analysing the face is provided in [41].

The single facial AUs and the facial activity described by them are listed below:

1 : Inner Brow Raiser	24: Lip Pressor
2 : Outer Brow Raiser	25: Lips Part
4 : Brow Lowerer	26: Jaw Drop
5 : Upper Lid Raiser	27: Mouth Stretch
6 : Cheek Raiser	28: Lip Suck
7 : Lid Tightener	29: Jaw Thrust
8 : Lip Toward Each Other	30: Jaw Sideways
9 : Nose Wrinkler	31: Jaw Clencher
10: Upper Lip Raiser	32: Lip Bite
11: Nasolabial Fold Deepener	33: Blow
12: Lip Corner Puller	34: Puff
13: Cheek Puffer	35: Cheek Suck
14: Dimpler	36: Tongue Bulge
15: Lip Corner Depressor	37: Lip Wipe
16: Lower Lip Depressor	38: Nostril Dilator
17: Chin Raiser	39: Nostril Compressor
18: Lip Puckerer	41: Glabella Lowerer
19: Tongue Out	42: Inner Eyebrow Lowerer
20: Lip Stretcher	43: Eyes Closure
21: Neck Tightener	44: Eyebrow Gatherer
22: Lip Funneler	45: Blink
23: Lip Tightener	46: Wink

B General Resources

Conferences and Workshops

International Conference on Pattern Recognition (ICPR)
IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)
IEEE International Conference on Computer Vision (ICCV)
IEEE International Conference on Systems, Man and Cybernetics
International Conference on Automatic Face and Gesture Recognition (FGR)
IEEE International Conference on Multimedia and Expo
IEEE Statistical Signal Processing Workshop (SSP)
Conference on Computer Vision and Pattern Recognition Workshop (CVPRW)
International Conference on Computer Vision Theory and Applications
International Conference on Multimodal Interfaces

Psychological Journals

Behavioral and Cognitive Neuroscience Reviews
Behavior Research Methods
Behavior Research Methods, Instruments and Computers
Emotion
Journal of Nonverbal Behavior
Journal of Experimental Psychology
Journal of Experimental Social Psychology
Journal of Personality and Social Psychology
Nature
Neuroscience and Biobehavioral Reviews
North American Journal of Psychology
Psychological Science
Psychology and Aging
Psychology Today
The Nature of Emotion

Technical Journals

IEEE Transactions on Pattern Analysis and Machine Intelligence
IEEE Transactions on Systems, Man and Cybernetics
IEEE Transactions on Image Processing
IEEE Transactions on Neural Networks
Journal of Multimedia

Journal of Information Processing Systems
International Journal of Computer Vision
International Journal of Pattern Recognition and Artificial Intelligence
International Journal of Human-Computer Studies
International Journal of Engineering Science and Technology
Pattern Recognition
Face Recognition
Lecture Notes in Computer Science: Image and Signal Processing
Signal Processing: Image Communication
Image and Vision Computing
International Journal of Pattern Recognition and Artificial Intelligence
Communications in Computer and Information Science: Signal Processing, Image Processing and Pattern Recognition
Computer Vision and Image Understanding

Books

Handbook of Pattern Recognition and Computer Vision
Unmasking the Face: A Guide to Recognizing Emotions from Facial Clues

Research Groups and Labs

Intelligent Behaviour Understanding Group (iBUG) at Imperial College, London
Computational Face Group at the University of California, San Diego
Affect Analysis Group at the University of Pittsburgh, Pennsylvania
Affective Computing Group at MIT Media Lab
Image Processing Lab, Kyung Hee University

Key Researchers

Paul Ekman	Marian Stewart Bartlett
Wallace V. Friesen	Gwen Littlewort
Joseph C. Hager	Takeo Kanade
Maja Pantic	Jeffrey F. Cohn
Michel Valstar	Cynthia Breazeal
Christoph Mayer	