Granular fuzzy model with hyperboxes for facial expression recognition

by

Eduardo Morales Vargas

Thesis submitted for partial requirement to obtain the degree of

Master of science, in computer science

by the

Instituto Nacional de Astrofísica, Óptica y Electrónica

July, 2017

Tonantzintla, Puebla

Advisors:

Dr. Carlos Alberto Reyes García

Dra. Hayde Peregrina Barreto

Computational sciences Division

INAOE

Contents

List of Figures

List of Tables

Acknowledgments

In a special way, I would like to express gratitude to my advisors Ph.D. Carlos Alberto Reyes and Ph.D. Hayde Peregrina Barreto because of all their sincere help, the continuous support, for their motivation, for all the knowledge that they shared with me and for the quality of people who are.

To overcome the stronger bonds that must be broken are those of the family. It is difficult to be far away, thanks for understanding that Irene, Jorge, Sheyla, Violeta, Samuel, Olga, and Ivan. Without forgetting other important members of my family.

Nothing can be done without any help. To all my friends at INAOE, in a special way to Jesus, Daniel and Elizabeth. Sitting at the table is not just for eating. Is to share, grow, regain courage and be happy;

I express my gratitude to all the researchers of INAOE which supported me, in a special way to Ph.D. Felipe Orihuela Espina and Ph.D. Luis Villaseñor Pineda.

And CONACyT, which helped me with a scholarship.

Acronyms

AAM Active appearance model

FACS Facial Action Coding System

FER Facial Expression Recognition

AU Action Unit

ANFIS Adaptive Neuro-Fuzzy Inference System

FCM Fuzzy C-Means

SC Subtractive Clustering

ANN Artificial Neural Networks

SVM Support Vector Machine

PA Procrustes Analysis

AT Affine Transformations

GA Genetic Algorithm

CK+ Cohn-Kanade Plus

AC Action Unit

Nomenclature

st	Set of facial landmarks
μ_s	Membership value given to a fuzzy set s
nls	Normalized landmarks
$BasicCar$	Raw Feature vector
is	Sequence of images
ar	Aspect ratio of a landmarks set
$AMAR$	Alignment Maintaining the facial Aspect Ratio
$AITNS$	Alignment In Terms of The Neutral State
$FERFAU$	Facial Expression Recognition with Fuzzy rules over Action Units
ns	Normalized landmarks
sz	size of the neutral state
r	Set of reference points
as	Aligned facial landmarks
ac	Area changes in the triangulated shape
ts	Triangulated shape
mo	Magnitude and Orientation of facial landmarks displacement
$FRGAS$	Fuzzy Rule Generation algorithm with antecedent selection

Summary

Facial expression recognition is related to the automatic identification of the overt manifestation of affective states of a subject by computational means and has applications in security, human computer interaction among others.

This work focuses on the design of a model for the recognition of the seven basic facial expressions: anger, contempt, disgust, fear, happiness, sadness and surprise. Facial expressions description in terms of action units is used as depart point. Fuzzy models are used in order to maintain a relation between the facial muscle appearance and the fuzzily associated facial expressions.

The proposed method, model facial expressions using granular fuzzy models, finding automatically fuzzy rules that can describe the output class with a low number of rules. The reason of use a model with a low number of rules lies in the necessity of a simple model that do not loose the ability to explain why is making a decision.

First, heuristic guided affine transformations align facial landmarks of the neutral and the target expression. Second, features are extracted describing face movements in terms of changes in orientation (angle and magnitude) of distinctive facial areas. Third, the full featured representation is embedded into a compact one by means of pooling. Finally, a Sugeno-type adaptive Neuro Fuzzy Inference System is used for each action unit to generate a description of the movements in the face that

identifies the facial expression present in an image sequence.

For evaluating the method the **CK+** database is used, it contains 327 labeled frontal image sequences from 123 healthy subjects in which one of the seven basic facial expressions is represented. Each sequence begins with a subject in a neutrally affective state and ends with a facial expressions.

The proposed model discriminates facial expressions with mean accuracy of 89.04±0.91% with a maximum accuracy of 91.41±28% . Further, distinctly to current solutions the model can also describe why is reaching such decision.

Chapter 1

Introduction

Facial expressions communicate emotions. They are practically the first thing that we look at when interacting with someone. Facial Expression recognition Facial Expression Recognition (FER) can be used for many applications such as security, human-computer interaction, driver safety, and health care [24]. The study of FER is important in areas such as psychology, neuroscience, education, or sociology [29]. In this thesis, we focus on FER through the description provided by Facial Action Coding System FACS which beyond emotion science and because of its descriptive power [9] it is possible to detect facial neuromuscular disorders [11].

Many studies have been made in FER, actual systems often proceed by extracting features from the input image set to feed a subsequent classifier that outputs the inferred facial expression [36], [27]. State-of-the-art algorithms in FER report maximum accuracies in the range of 90.51±0.64%. But current solutions have favored discriminative over explicative power e.g. [24], [17] .Consequently, a general limitation of current developments is their explicative capacities. Explicative models go beyond predictive and discriminative models affording not only an output label but also accompanying it with procedural mechanics. In FER, there are initial steps in this direction [15], [18], but admittedly, there is room for improvement.

Usually, there are two main approaches for FER [17]. The static approach is based on local descriptors of the whole image in which a facial expression is represented. Dynamic approaches take as reference the difference between the neutral state and the representation of a facial expression (Fig. 5.11). Another kind of approach for facial expressions recognition problems is to use facial distinctive areas or use the face as a whole:

- Based on facial distinctive areas: face distinctive areas(nose, mouth, eyes, eyebrows) are detected and then each one is segmented; features for each area are extracted to discriminate between the facial expressions. [2]
- Holistic approach: Statistical methods are used to extract features of the face as a whole. [2]

Figure 1.1: Basic facial expressions, from left to right: angry, disgust, contempt, happy, fear, sadness and surprise. Images taken from the **CK+** dataset.[24].

[2] say that the facial expressions recognition problem is divided into four major steps:

1. Face detection detects the face in the image.
2. Normalization, lightning, and other effects are reduced to enhance the image quality.

3. Feature extraction, the extraction of relevant features to describe facial expressions is performed in this step.

4. Classification, in this step facial expressions, are classified by a trained model.

In this thesis, we focus on the normalization, feature extraction, and classification steps of FER problem, by using a dynamic facial expression recognition approach of facial distinctive areas (nose, mouth, eyes, etc.) because a comparison gave evidence that facial expressions recognition systems with local features are better than with global ones [15]. We take as reference the movements of facial distinctive areas which are encoded as Action Units Action Unit (AC) [13] and then this representation is used to describe each facial expression in terms of AU, taking advantage of the easiness provided by fuzzy inference systems to use linguistic variables and fuzzy rules. Each facial movement is called an AU and describes the smallest visually discriminable facial deformation.

1.1 Problem

Given an image sequence $is = \{I, t\}$ where I represents a facial frontal image in which a facial expression is shown and t represents the temporal location. Active Appearance Models (AAM) [12] retrieve 68 coordinates pairs from the face images describing the face shape, each one corresponding to a vertex of a face descriptors set $st = \{\{x_i, y_i, x_2, y_2, ..., x_n, y_n\}, t\}$ where x and y are coordinates of points located over the image I_t. Is intended to use Facial landmarks st to recognize the following facial expressions: anger, contempt, disgust, fear, happiness, sadness, and surprise.

The necessity of a simple and interpretable model which also can provide procedural mechanics to facial expression recognition guided us to the use of fuzzy models. Fuzzy models with a higher number of rules usually exhibit higher accuracies than one with a less number (a trivial consequence of increasing the model parameters), but in losing simplicity, they lose the ability to explain why the model is making a decision and are more prone to over-fitting. We thus strive to reduce the dimensionality of the representation to generate a simpler model affording fewer rules.

The computational problem attended in this proposal are:

1. The design of a modeling strategy to the classification of facial expressions in terms of AU. Being an AU the smallest movement is present in the face.
2. The generation of fuzzy models with a small number of fuzzy rules.
3. The automatic generation of fuzzy rules for fuzzy modeling.

1.2 Research questions

The research questions what we are responding to are the following:

- Fuzzy models can be tapped to afford both discriminative and explicative outcomes of facial expressions from frontal facial image sequences?
- How is it possible to recognize facial expressions using fuzzy rules over the description of the expressions in terms of action units?
- How does Pooling helps to reduce the complexity of fuzzy models while they keep their effectiveness?
- How is it possible to model facial expressions using an automated algorithm for fuzzy rules generation?
- Fuzzy rule antecedent selection helps to afford discriminative power in facial expression recognition?

- Fuzzy rule antecedent selection before fuzzy rule generation by algorithms allows generating low complexity fuzzy models?

1.3 Hypothesis

Under controlled conditions (negligible camera rotations or illumination changes, absence of zooming operations and occlusions) automatically generated fuzzy rules defined over Action Units Action Unit (AU) can exhibit a high discriminate power between facial expressions whilst concomitantly explaining the actions of the model.

1.4 Main objective

The main goal of this research is to model facial expressions through automatically generated fuzzy rules defined over action units based on the Facial Action Coding System FACS.

1.5 Specific Objectives

The specific objectives are listed below:

- Design a feature representation which describes facial movements.
- Design and implement a granular fuzzy model to generate a description of the movements in the face in terms of action units.
- Design, implement and test a granular fuzzy model for the classification of facial expressions based on FACS.
- Optimize the parameters of the model generated.

- Validate the model using data labeled by experts.

1.6 Contributions

As a final result of this thesis work, we obtained a granular fuzzy model for facial expression classification. The model is the result of a proposed methodology for the automatic generation of fuzzy rules. In addition, the main contributions of this work are:

- A feature representation that describes the movement in the face in a compact way.
- A feature pooling scheme of facial movement characteristics.
- Identification of action units that better describes facial expressions through granular fuzzy models.
- A model which generates a description of the movements in the face in terms of action units.
- A model which infers which facial expression is presented in the frontal image sequence of the face.
- An algorithm for automatic generation of fuzzy rules with rule antecedents selection.

1.7 Publications

The following publications have been a direct result of this research:

- E. Morales-Vargas, C.A. Reyes-Garcia, Hayde Peregrina-Barreto, Facial expression recognition based on the dynamic of facial landmarks, 9o Congreso Mexicano de Inteligencia Artificial COMIA 2017. Accepted 08/05/2017

- E. Morales-Vargas et al, Facial expression recognition with fuzzy explainable models, 10 th International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications MAVEBA 20017. Accepted 10/07/2017

1.8 Scope and limitations

For this work, the following facial expressions will be recognized: anger, disgust, contempt, happiness, fear, sadness, and surprise. [31, 8, 29]

We focus on facial expressions recognition. We do not attempt to validate or evaluate any model for action unit recognition.

We do not intend to give conclusive explanations in the areas of anatomy, psychology, or neuroscience on the interpretable results of the models.

Our model is limited to a dataset of images taken into a controlled environment without the interference of illumination and facial occlusions. We part from facial landmarks contained in the dataset, we do not intend to obtain any other facial landmarks.

1.9 Document structure

This work is composed of six chapters which are briefly described below. In chapter two, the theoretical framework, and the main concepts that serve to understand and give support to this thesis work are presented. In chapter three, the state of the art, and the review of works related to the research project are shown. In chapter four, the proposed solution is presented. In chapter five, the experiments and results obtained are shown, and finally, in chapter six, conclusions and future work, we talk about some conclusions that we obtain after the completion of the research.

Chapter 2

Theoretical framework

In this chapter are presented some basic concepts which are used in the document such as, among others. The following sections briefly describe each concept.

2.1 Facial Action Coding System FACS

The FACS [9], [14], [13] specifies 9 action units in the upper face and 18 in the lower face. In addition, there are more movements not related to facial expressions. Each facial movement is called Action Unit and describes the smallest visually discriminable facial deformation. FACS is used to name facial movements and is a standard to classify the facial expression of emotion through rules. By using FACS, human experts can detect and encode basic emotions by observing Action Units presented on the face.

In the facial expression description, in terms of action units, only are considered the movements of the upper and lower face. In Table 2.1 AU related to facial expressions are shown.

Table 2.1: Action Units and his corresponding movement

AU	Movement	AU	Movement
0	Neutral state	14	Dimpler
1	Inner brow raiser	15	Lip corner depressor
2	Outer brow raiser	16	Lower lip depressor
4	Brow lowerer	17	Chin raiser
5	Upper lid raiser	18	Lip puckerer
6	Cheek raiser	20	Lip stretcher
7	Lid tightener	22	Lip funneler
9	Nose wrinkler	23	Lip tightener
10	Upper lip raiser	24	Lips pressor
11	Nasolabial deepener	25	Lips parted
12	Lip corner puller	26	Jaw drop
13	Cheek puffer	27	Mouth stretch

2.2 Active appearance Models

Active Appearance Models [12] have been used in areas such as anthropology or computer sciences [24] to explain images by generating a model conformed by coordinates fitted to a shape of the picture which is going to be explained. Fitting an AAM to an image consists of minimizing the error between the input image and the closest model instance solving a nonlinear optimization problem (2.1). The usual approach is to use an iterative algorithm to solve the parameters for incremental additive updates in order to match the generated model and the input image (See Fig. 2.1 [25]).

A facial model in a 2D space is a triangulated shape $s = [x_1, y_1, x_2, y_2 ..., x_n, y_n]$ where x and y are coordinates in an image [24]. Suppose that given an input image

(a) Initialization (b) Matching

Figure 2.1: Matching between a face model and an image

$I(z)$ that is wanted to be fitted by an AAM and that is known the optimal shape p and appearance δ parameters for the fit. This means that the image $I(z)$ and the model instance $M(W(z;p)) = A(z)$ must be similar. The fitting process is related to minimize the error between $I(z)$ and $M(W(z;p)) = A(z)$. There are two coordinate frames in which the error can be computed, the coordinate frame of I and the coordinate frame of the AAM.

2.3 Affine transformations

An affine transformation is a linear mapping method of the euclidean plane. F, is a mapping that maps each point X to a point $F(x)$ defined by $F(x) = AX$ [5]. Affine transformations are widely used to correct geometric distortions in images or landmarks. Affine transformations preserve the shape, for example, a set of parallel lines remain parallel after an affine transformation.

$$F(X) = \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ 0 & 0 & 1 \end{bmatrix} X \tag{2.1}$$

Affine transformations can be used in an n-dimensional space. The following equations show the most widely used affine transformations: Eq. 2.2 being the rotation, Eq. is the 2.3 translation and Eq. 2.4 the scale on a 2-dimensional space.

$$[x, y] = \begin{bmatrix} \cos(an) & \cos(an) & 0 \\ -\sin(an) & \cos(an) & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x & y \end{bmatrix} \tag{2.2}$$

$$[x, y] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \Delta x & \Delta y & 1 \end{bmatrix} \begin{bmatrix} x & y \end{bmatrix} \tag{2.3}$$

$$[x, y] = \begin{bmatrix} sx & 0 & 0 \\ 0 & sy & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x & y \end{bmatrix} \tag{2.4}$$

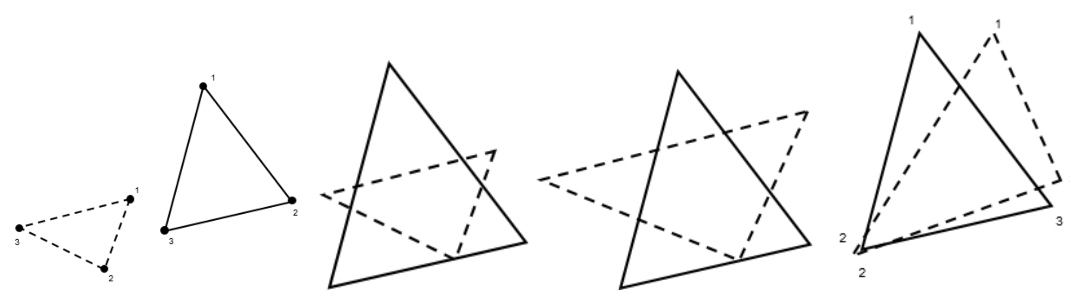
Where: x and y are coordinates in a euclidean space. In our application x and y are coordinates defining the facial shape. an defines the rotation angle. Δx and ΔY specifies the displacement in the x and y axes. Finally, sx and sy are the scale factor for x and y axes respectively.

2.4 Procrustes analysis

Procrustes analysis (PA), also known as Procrustes superimposition, has been used for shape analysis with various applications [12]. Procrustes Analysis Procrustes Analysis (PA) superimposes shapes by optimally translating, rotating and uniformly scaling objects [25], [26]. Procrustes Analysis mitigate geometric distortions in images or landmarks using affine transformations by minimizing the distance between two shapes using an error function. Procrustes analysis is divided into three steps:

1. A translation puts the centroid of all the analyzed shapes at a converging point
2. A rescaling gives all shapes a centroid size of 1
3. A rotation is performed iteratively so that the distances between all shapes is minimized

In Fig. 2.2 the graphical representation of the three superimposition steps is depicted. First, in 4.3a the original coordinates before the process is shown. In Fig. 2.2b a representation of a translation that put the centroid of both triangles at a converging point. Then a scale set the centroid size to 1 in Fig. 2.2c. Finally, rotations are performed so that the minimum distance between the two triangles is found (Fig. 2.2d).

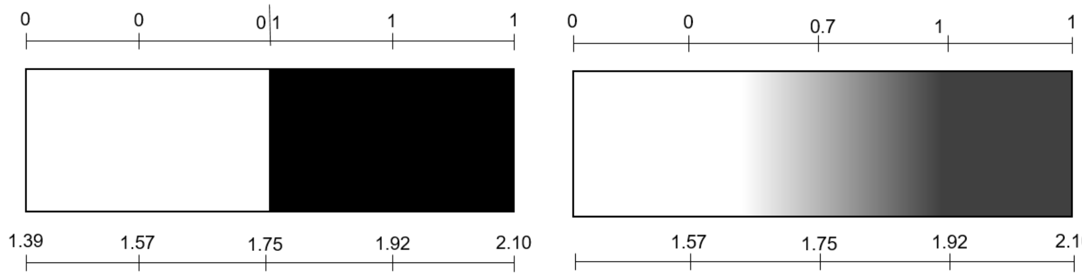


(a) Landmarks (b) Translation (c) Scale (d) Rotation

Figure 2.2: Graphical representation of Procrustes superimposition process, image taken from [37]

2.5 Fuzzy logic

Fuzzy logic is the formalization proposed to handle semantic and subjective ambiguity developed by Lofti Zadeh in 1965. In classical logic, an element of a set either belongs or does not belongs to a set. In fuzzy theory the same element belongs with a certain grade to a set in the interval [0,1]. The principal difference between boolean and fuzzy sets consists of the addition of a characteristic function that map all the elements of the fuzzy set into values between 0 and 1; this function is called membership function μ_f. In the particular case in which the function μ_f contains the restriction to either values 0 or 1, then this set is reduced to a boolean one [1]



(a) Crisp values

(b) Fuzzy values

Figure 2.3: Upper metric correspond to the range of logical values for the example and lower metric illustrate the limits. [35]

A classic example consists on to label subjects into the set "high" taking into account the height of each subject. In boolean logic, it is necessary to establish a lower limit (ex. 1.75 m.), when this limit is passed the subject is labeled as "high". Besides, in the fuzzy set, the membership is gradual.

As can be seen in table 2.2 the difference between boolean logic and fuzzy logic lies in the membership of an element of a set. In the case of boolean logic, always a value of 0 or 1 describes this membership. In the case of fuzzy logic, values between [0,1] are used to describe this membership.

2.5.1 Fuzzy sets

Let $u = X$ be the universe or set of all the possible elements for a model. A fuzzy subset F or either for simplicity called fuzzy set defined in the universe U is characterized by a membership function $\mu_F(x)$ which maps all the elements of a domain or universe with values in the interval $[0, 1], F : X \Rightarrow [0, 1]$. Membership

Table 2.2: Fuzzy logic and Boolean logic comparison

Subject	Height	Boolean logic	Fuzzy logic
Andrea	2.05 m	1	1.00
Pedro	1.96 m.	1	1.00
Luis	1.80 m.	1	.85
Jesus	1.75 m,	0	.75
Carlos	1.70 m	0	.40

values change between 0 and 1, where 0 denotes that the element x absolutely does not belong to F and the closer to 1, the more is the membership of the object x to F set

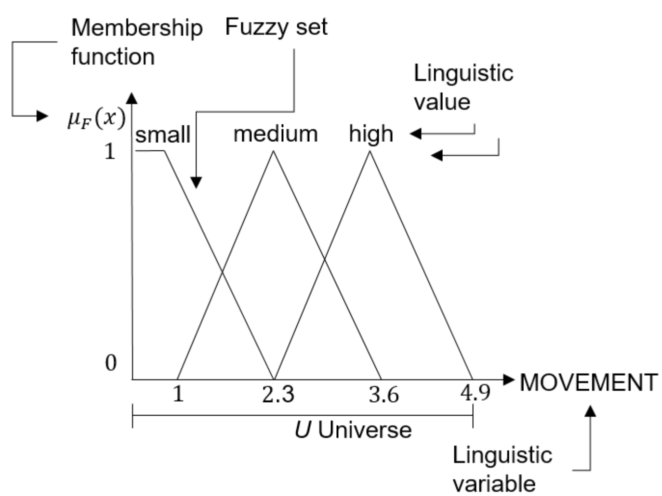


Figure 2.4: Graphic representation of fuzzy set components

Let be X a set of elements x. A fuzzy set F is a collection of ordered pairs $x, \mu_F(x)$, for x∈X. where X is the universe and $\mu_A : x \rightarrow [0, 1]$, this is a common notation when X is countable and is represented as $A = \mu_A(x_1)/x_1 + \mu_A(x_2)/x_2 + ... + \mu_A(x_n)/x_n$.

2.5.2 Linguistic variable

Linguistic variables are used every day to express the importance of an entity and its context, for example, it represents an opinion independent of the measuring system and it has information that most listeners will understand. A linguistic variable means a variable whose values are words or sentences in a natural or artificial language. For example, "movement" is a linguistic variable if its values are linguistic rather than numerical, slow movement, medium movement, quick movement, etc., rather than 0.1, .5 or 1. [39]. Basically, a linguistic variable, is "whose values are words or sentences in a natural or artificial language". Formally, a linguist variable is a 4-tuple (T, X, G, M) where T is a set of natural language terms called linguistic values, X is the universe, G is a context used to generate elements of T, and M is a mapping from T to the fuzzy subsets of X [1].

For example, x being an element of the universe X:

$$T = \{small, medium, large, ...\}, X = [0, 250]$$

M is a mapping, in this example we will examine M for medium.

$$\mu_A(x) = \begin{cases} 0 & if, x \leq 170, \\ \frac{x-170}{15} & if, 170 < x \leq 185, \\ 1 & if, 185 < x. \end{cases}$$

2.5.3 Fuzzy rules

Rules are used commonly in knowledge representation. They can be defined as an IF-THEN structure that relates given information or facts. A rule proves some description of how to solve a problem or to model some knowledge.

Rules consist of two parts, the IF also known as the antecedent, and the THEN, known as the consequent.

IF <antecedent>
THEN <consequent>

A rule can have multiple antecedents joined by the keywords AND, OR and its use implies conjunction or disjunction respectively

IF <antecedent 1>
AND <antecedent 2>
OR < antecedent 3>
...
THEN <consequent>

In fuzzy systems there are fuzzy rules, the general form of fuzzy rules is:

antecedent	consequent
IF u_i is A_1 AND u_2 is A_2	THEN y is B

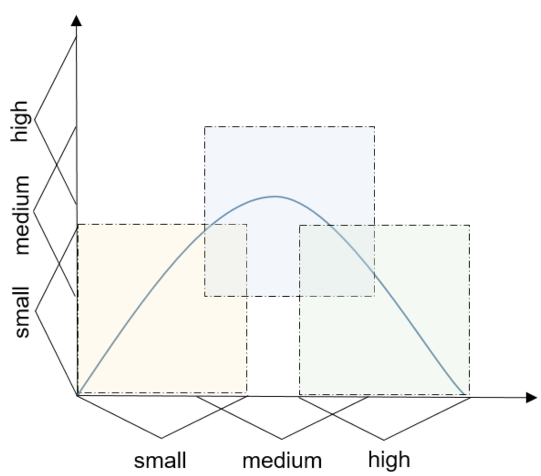
Where u is a crisp input from the universe, y is a linguistic variable, A and B are linguistic values determined by fuzzy sets. Fuzzy rules are a key tool for expressing pieces of knowledge in fuzzy logic. Let's consider the following rule IF $x \in A$ THEN $y\ inB$ where A and B are ordinary subsets and x and y are variables in the universe U. The rule IF $\mu_a(x) = 1$ THEN $\mu(y) = 1$ provides a description of a relationship between x and y only in a particular case where $\mu_a(x) = 1$, is seen that the consequent if crisp. Fuzzy rules are rules whose antecedents, consequences or both are fuzzy rather than crisp.

2.6 Granularity

In fuzzy logic, everything is allowed to be granulated. An information granule is a clump of attribute-values drawn together by similarity, proximity, or functionality

[40]. Graduated granulation is inspired by the way in which humans deal with complexity and imprecision and may be viewed as a form of information compression of variables and input/output relations. Informally, a human face can be clustered in granules or distinctive areas, such as eyes, mouth or eyebrows, and the eyebrows can be divided into inner and outer eyebrows, this is an example of granularity levels where each granule contains its own information and characteristics.

More specifically, consider a variable x which takes values in U. Let u be a value of x. Informally, if u is known precisely, then u is referred to as a singular point value of x, if x is not known precisely but the is some information that constrains possible values of u, then the constraint of u defines a granular value of x. A granular variable is a variable that takes granular values, in this sense, a linguistic variable is a granular variable that carries linguistic labels. For instance, fuzzy clustering allows generating of granular variables to model the possible values of u. A granule can be defined as a cluster or a membership function, moreover, a granule is a formalization of a hyperbox. A hyperbox is a region in the decision space.



If X is small then Y is small

if X is medium then Y is medium

if X is large then Y is small

(a) Function

(b) Fuzzy rules

Figure 2.5: Granulation of a function using three granules, the number of granules can impact in the generalization capabilities of the model [34].

2.6.1 Fuzzy models

Fuzzy models are widely used because they have a lot of applications [39]. Mathematically, a standard fuzzy system is a static nonlinear mapping between its inputs and outputs. It is assumed that the fuzzy systems have inputs $u_i \in U_i$ where $i = 1, 2, ..., m$, the inputs and outputs are crisp, that is, they are real numbers not fuzzy sets. Fuzzy models are practically composed of three blocks, the fuzzification block, the inference mechanism, and the defuzzification block. The fuzzification block converts the crisp inputs to fuzzy sets, the inference mechanism uses the fuzzy rules in the rule-base to produce fuzzy conclusions, and the defuzzification block converts these fuzzy conclusions into the crisp outputs [30].

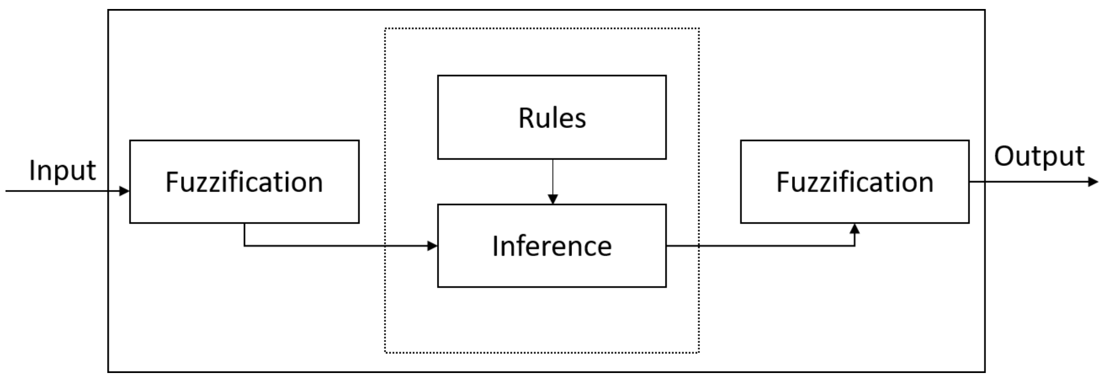


Figure 2.6: Composition of fuzzy models. Fuzzy models are composed by three blocks: Fuzzyfication block, inference mechanism, and defuzzification block.

The principal difference between Mamdani type FIS and Sugeno type FIS is the way the crisp output is generated from the fuzzy inputs [16]. Sugeno-type FIS uses the weighted average result of fuzzy rules to compute the crisp output. On the other hand, Mamdani-type FIS uses a deffuzification method. The most used fuzzy models are Mamdani and Takagi-Sugeno Fuzzy Inference Systems. Mamdani type is the most used to describe expert knowledge. Moreover, the Sugeno type is computationally more effective and works well fitting functions.

2.7 Fuzzy clustering

Clustering algorithms are grouping methods for a dataset using some similarity or distance criterion defined through one operation or function. Fuzzy clustering is a kind of clustering in which each element of the set has a membership grade to the groups. In this section, some fuzzy clustering algorithms are briefly described.

2.7.1 Fuzzy c-means

Fuzzy c-means algorithm is a fuzzy clustering technique which part a data set $X = \{x_1, x_2, ..., x_n\}$ into groups also called clusters $C = \{c_1, c_2, ...c_{cn}\}$, in which is obtained for each data point of the data set a partition matrix $W =_{i,j} \in [0, 1], i = 1, ..., n, j- = 1, ..., cn$ where each element $w_{i,j}$ denotes the membership degree to which element x_i belongs to each cluster c_j . There are a large number of modifications to this algorithm but the base algorithm is the same: use the data points of the dataset to iteratively optimize an objective function (Eq. 2.5), [6].

$$J_m(U, v) = \sum_{i=1}^{c}\sum_{k=1}^{n}(u_{ik})^m d_{ik}^2 + a \ \sum_{i=1}^{c}\sum_{k=1}^{n}(u_{ik} - f_{ik})^m d_{ik}^2 \tag{2.5}$$

Where:

U	fuzzy partition matrix
v	group centers matrix
u_{ik}	fuzzy membership value for i-th data point with a value between 0 and 1
d_{ik}	distance for the k-th data point with the i-th data center
f_{ik}	membership value of the data point member of the i-th group
c	Number of clusters
n	Number of data points

The a coefficient denotes the scale factor and m the diffusion factor. a role consists into keep a balance between supervised and no supervised components in the optimization mechanism, m controls diffusions level in the classification. For this parameters the most used values are $m = 2$ and $a = L/n$ where L equals to the size of the labeled samples. Function J_m can take a bigger number of values and it is associated with the best grouping.

2.7.2 Subtractive Clustering

Subtractive clustering was proposed in [21]. For this algorithm, the training data set Z is normalized into values in the range [0,1] for each dimension. The algorithm is based on selecting the data point with the best cluster potential P_j^* which is calculated taking into account the distance with all the other data points.

$$P_i^* = \sum_{j=1}^{n} e^{-\alpha \|x^i - x^j\|^2} \tag{2.6}$$

Where:

$$\alpha = \frac{\gamma}{\gamma_a} \tag{2.7}$$

Where:

P_i^*	cluster potential value for the i-th value of the training data set
γ_a	weight between i-th and j-th data point
x	data point
γ	is variable, commonly set 4
γ_a	positive constant called cluster radius

The potential of a data point to be a cluster center is higher when more data points are closer, consequently, small groups of points can generate a certain number of cluster to represent them. The data point with the highest potential, denoted by

P_i^* is considered as the first cluster center $c_1 = (d_1, e_1)$. The potential is then recalculated for all other points excluding the influence of the first cluster center according to:

$$P_i^* = P_i^* - P_k^* \zeta \tag{2.8}$$

Where:

$$\zeta = e^{-\beta \|x^i - c^k\|} \tag{2.9}$$

$$\beta = \frac{4}{\gamma_b^2} \tag{2.10}$$

$$\gamma_b = \gamma_a x \eta \tag{2.11}$$

P_i^*	new potential value for the i-th data point
P_k^*	potential value as a cluster center
c	cluster center of data
β	weight of i-th data point to the cluster center
γ_i	distance between the cluster center
η	diffusion factor

Again, the data point with the highest potential P_k^* is considered to be the next cluster center c_k if:

$$\frac{d_{min}}{\gamma_a} + \frac{P_k^*}{P_1^*} \geq 1 \tag{2.12}$$

Finally, the clustering end if the following condition is fulfilled.

$$P_k^* < \epsilon P_i^* \tag{2.13}$$

Where: ϵ is the reject ratio.

Equation 2.14 is the common form of subtractive clustering

$$\mu_j^{ik} = e^{-\alpha \|x_j^i - c_j^k\|^2} \tag{2.14}$$

2.8 Takagi-Sugeno Fuzzy Inference System

2.9 Pooling

In general terms, the objective of pooling is to transform a feature representation into a compact representation that preserves important information [7]. Pooling features of a neighborhood creates invariance to changes and reduce the dimensionality of the feature representation. The pooling operation is typically a sum, an average, max, or some other user-defined operation. Some well-known works that use pooling are Scale Invariant Feature Transform (SIFT), and Histogram of Oriented Gradients (HOG), among others, [23]. Equation 2.15 show the average pooling operation and in equation 2.16 the max operation can be seen. The most difficult decision to pool features consists on identifying which features belong to a certain group and selecting the operation to be used in the input space [19].

$$f_p(v) = \frac{1}{p}\sum_{i=1}^{p} v_i \tag{2.15}$$

$$f_m = max_i v_i \tag{2.16}$$

Chapter 3

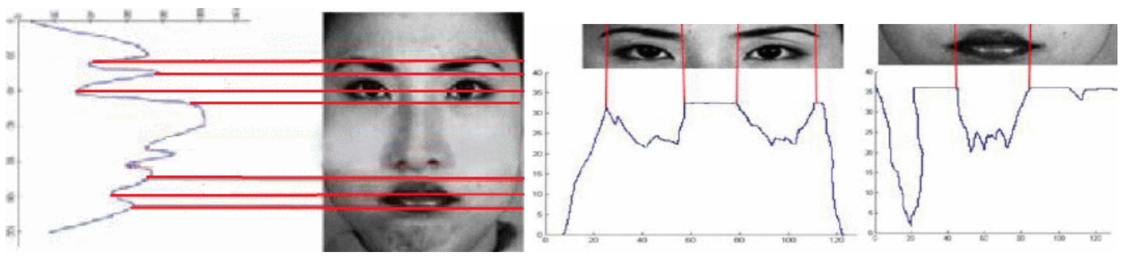
Related work

In this chapter, some works related to this thesis are presented. Basically, the works described have two interesting features: the models have the capability to be explainable or the feature extraction describes the movement presented in facial distinctive areas.

[15] proposed a facial expression recognition system using facial distinctive areas and fuzzy logic on the JAFFE dataset. The authors proposed a method consisting of two major steps: facial feature extraction and classification based on a fuzzy rule-based system. The facial features extraction step consists of applying horizontal integral projection on the original binary image using Eq. 3.1, and later vertical integral projection is applied using the horizontal projection. The abscissa axis of the eyebrow, eyes, and mouth is identified, and the vertical integral projection indicates the ordinate axis of the eyes and the mouth areas. In this way, the facial distinctive areas are identified by using Bezier curves to model smooth curves because the original integral projection curves are irregular.

$$H(y) = \frac{1}{X_2 - X_1} \sum_{x=X1}^{Y2} I(x, y) \tag{3.1}$$

$$V(X) = \frac{1}{y_2 - y_1} \sum_{y=Y1}^{X2} I(x, y) \tag{3.2}$$



(a) Horizontal integral projection (b) Vertical integral projection

Figure 3.1: Horizontal and vertical integral projections.

Then, considering three facial distinctive areas (eyes, eyebrows, and mouth) the appearance of these elements is used as input for a fuzzy model. The encoded measurements are the following: opening for eyes and mouth,constriction of eyebrow and mouth. Three fuzzy sets (low, moderate and high) were used with Gaussian membership functions. The number of rules generated were of 565 for this task. Below two sample rules presented in the work are shown.

- IF eye opening is very high AND eyebrow constriction is very low AND mouth opening IS very high AND mouth constriction IS low THEN Surprise
- IF eye opening is very LOW AND eyebrow constriction is very hight AND mouth opening IS very low AND mouth constriction IS hight THEN Disgust

Besides than obtaining a fuzzy model (Fig. 3.2 with an accuracy of 96.42% for the JAFFE dataset, the "thoughts" of the model can be interpreted. Interpretability is a desired characteristic for today's systems, even so, a limited number of facial appearances can be inferred from his rules.

Another facial expression recognition system, in where local features are extracted from the whole image, is the one presented in [18]. This system consists

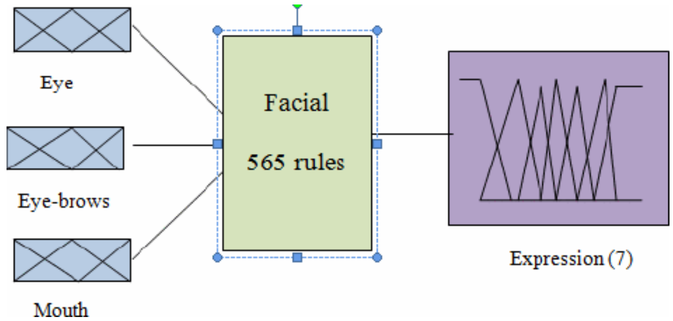


Figure 3.2: Model overview proposed in [15]

of 4 modules: preprocessing, region extraction, feature extraction, and expression recognition based on a Mamdani fuzzy inference system.

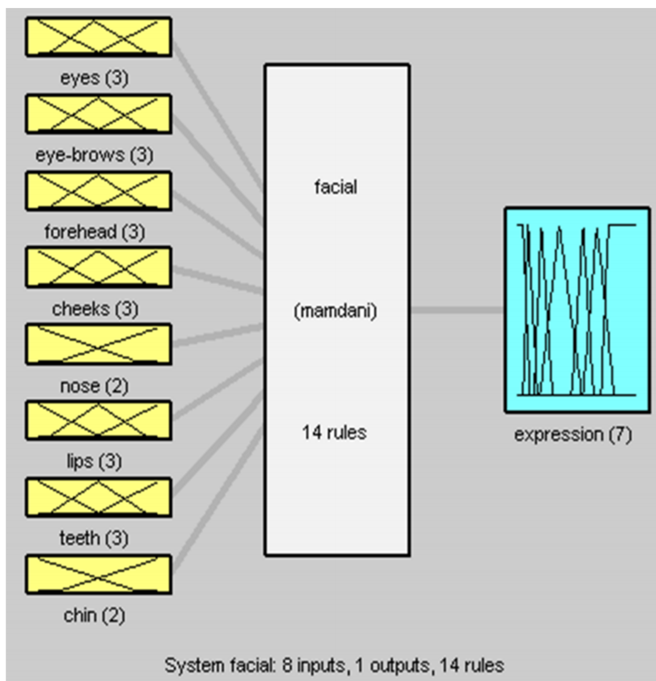
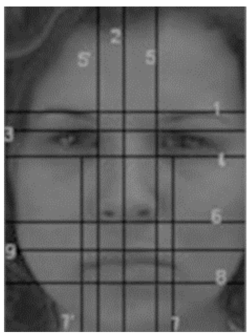


Figure 3.3: Model overview proposed in [18]

First, in the preprocessing module, the face is extracted from the background and scaled to a predetermined size. Later, 9 basic lines for the region extraction which carry a semantic significance were defined (Fig. 3.4b). Next, these lines are used to extract the region associated with each one with the aim of extracting distinctive facial areas. Then, algorithms were designed for finding facial actions in each distinctive area; it is not mentioned which algorithms were used. The facial action elements considered for expression output were eyebrows, forehead, nose, chin, teeth, cheeks, and lips. States of these facial elements act as input to the fuzzy system shown in Fig. 3.3



(a) Image Lines for region extraction

No.	Semantic significance
1	Eyebrows top
2	Face middle
3	Eyes top
4	Eyes bottom
5	Eyes inner corner
6	Face middle
7	Lips outer corner
8	Lips bottom
9	Lips top

(b) Lines for region extraction

Figure 3.4: Image and lines for region extraction, figure and table taken from [18]

This fuzzy model [18] can also explain some actions presented in the face, such as the openness of the eyes or the stretch of the eyebrows, obtaining an average accuracy of 87.5% for six facial expressions.

[24] presented a dataset containing 327 image sequences for action units and facial expressions recognition, each sequence contains the representation of a facial expression and is labeled with the present action units and their respective intensity. Through AAM, a set of 68 coordinates describing the face are obtained, each one representing a vertex of a shape called hereafter facial landmark. After facial landmarks were got, a normalization step where noise caused by location, size, and orientation is reduced using Procrustes superimposition, is followed. Similarity Normalized Shape (SPTS) is obtained and it refers to the abscissa and ordinate

coordinates corresponding to the 68 vertex points, i. e., a 136 dimensional feature vector. The Action Unit 0 (AU0) normalization was used in these coordinates and it consists of subtracting the features of the first frame (neutral state) from the image sequence. Also, Canonical Normalized Appearance (CAPP) features were extracted by applying a piece-wise affine warp on each triangle path appearance in the source image so that it aligns with the base face shape resulting in an 87 x 93 synthesized grayscale image. The best-reported accuracy obtained for the combination of both features was 88%.

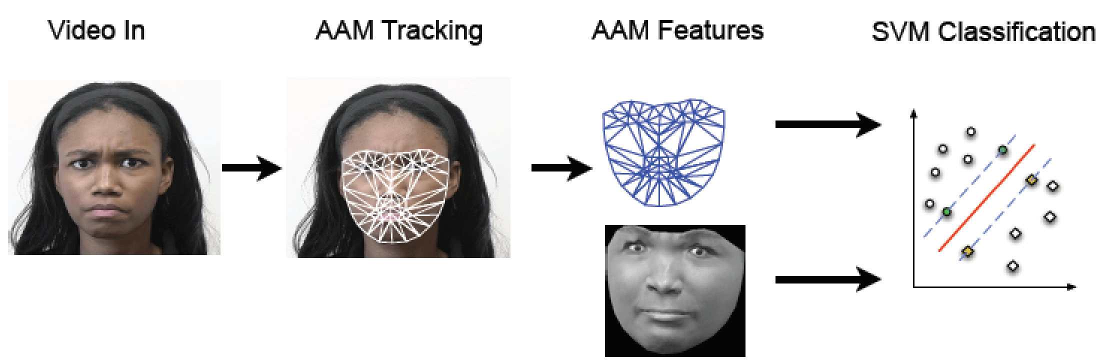


Figure 3.5: Block diagram of the system proposed in [24]. The face is tracked using an AAM and from this they get features for classification (SPTS and CAPP) using a linear Support Vector Machine. SVM

Moreover, [17] proposed a simple descriptor based on capture the angles obtained trough the selection of three facial landmarks in a location. This is a size invariant approach given that only it is measured the variation of facial movement. The used descriptor can only have three discrete values given by the difference in magnitude of the movement between the facial expression state and the neutral state (Eq. 3.3). The best-reported accuracy was of 86% using 560 angles of the 5,000 possible combinations employing 68 facial landmarks.

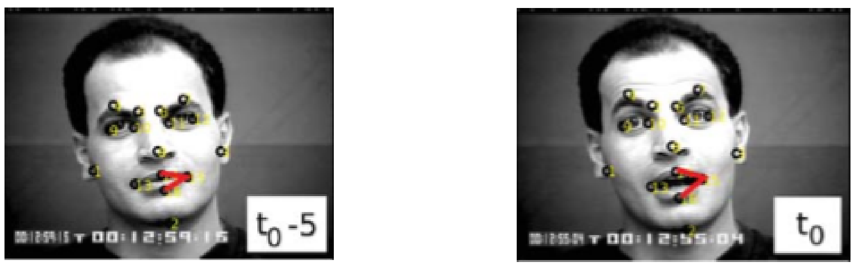
$$f_{l_i,l_2,l_3}(t) = \begin{cases} -1 & \text{if } d_{l_i,l_2,l_3}(t) < -\theta \\ 0 & \text{if } \|d_{l_i,l_2,l_3}(t)\| \leq \theta \\ +1 & \text{if } d_{l_i,l_2,l_3}(t) > \theta \end{cases} \tag{3.3}$$

Where:

$d_{l_i,l_2,l_3}(t) = a_{l_i,l_2,l_3}(t) - a_{l_i,l_2,l_3}(t-T)$

a_{l_i,l_2,l_3} Angle, where l_1 landmark as the central point l_2, l_3 extreme points

θ threshold that validates an angle change as significant



(a) Neutral state (b) Angle change

Figure 3.6: Angle change in the mouth triangle. In this sequence surprise is represented. Images taken from [17]

Some works were presented with a global approach for facial feature extraction, [22] used local binary patterns as a pattern to classify the facial expressions, [27] proposed a fuzzy local binary pattern descriptor for facial expression recognition using the CK+ dataset. Also works with cross dataset testing were presented, [41] evaluated the generalization of a model training with 6 or 7 databases and testing with the missing one. On the other hand, [3] evaluated their model training with one database and testing with another one with a method based on HOG features. On the other hand, [43] used facial landmarks with a manifold approach for the recognition of facial expressions, and [38] used features of regions of interest from the face with Convolutional Neural Networks.

Some work related to facial landmarks extraction was presented, Open face is an open source analysis toolkit presented in [4]. it contains an algorithm to facial landmarks extraction based on HOG features. Also [33] proposed an algorithm based on HOG features. the difference lies in a windowed analysis.

3.1 Discussion

The works related to the proposed methodology are summarized in Table 5.8. As can be seen, different approaches have been proposed for facial expression recognition. In this case, there are two main general approaches: static and dynamic. But also different kinds of works can be found, the ones which are explainable.

Table 3.1: State of the art for facial expression recognition

Work	Task	Approach	Landmarks	Interpretable
[24]	FER	holistic	yes	no
[17]	FER	holistic	no	no
[18]	FER	ROI	yes	yes
[15]	FER	ROI	yes	yes
[27]	FER	holistic	no	-
[22]	FER	ROI	no	no
[4]	FER	-	-	-
[33]	FER	-	-	-
[43]	FER	yes	no	
[3]	FER	holistic	no	no
[41]	FER	holistic	no	no
[38]	FER	ROI	no	no
Proposed	FER	ROI	yes	yes

[24] presented three methodologies based on a dynamic approach. The first two use facial landmarks to obtain the movement or deformations in the following

an SVM classification obtaining an accuracy of 0.64 for the SPTS features and an accuracy of 0.8 for the CAPP features. Later, an assembly of these two classifiers was made resulting in the improvement of the accuracy, obtaining an 0.88 for S+C. On the other hand, [17] proposed two approaches. The first one is based on the dynamic of 560 angles formed by 68 landmark points and uses Conditional Random Fields as classifiers. The second approach proposes a method that uses Oriented Fast and Rotated BRIEF (ORB) with SVM obtaining an accuracy of 0.86 and 0.92, respectively. Although most of the dynamic approaches capture the movement of the face, they do not focus on explaining what is happening in the sequence. For example, if it is found any movement in the eyes or the mouth. The exception are the works presented in [18] and [15]. In these works, a number of states for each distinctive facial area are used to identify the represented facial expression. [18] uses a Mamdani Fuzzy System to model facial states obtaining an accuracy of 0.96. Finally, [15] extract regions in the face in order to explain the states of each one; the studies reported accuracy of 0.87. Explicative models go beyond predictive and discriminative models in trying to afford not only an output label but also accompany it with procedural mechanics. In facial expressions recognition, there are initial steps in this direction. The solutions in [15] or [18] using the advantages of fuzzy models explained the resolved decisions with a limited number of actions. Admittedly there is room for improvement.

Chapter 4

Proposed solution

We proposed a methodology for facial expression recognition divided into three steps: facial landmarks alignment, feature extraction, and classification.

Is often convenient that descriptors of facial expressions are invariant to scale, orientation and translation, for this reason, is needed to remove the size, orientation, and location of the shape. Procrustes analysis is one method used in the literature for this task [24],[12]. We proposed a heuristic method based on affine transformations to perform this step.

The next step consists on extract features that describe the movements in the face. The basic feature representation consists of the concatenation of the intensity, angle orientation of facial landmarks, and size changes in facial distinctive areas (such as eyes or mouth). Later, the basic representation is pooled to generate a simple and compact one, this allows to reduce the complexity of the Fuzzy Inference Systems used to predict which facial expression is present in a sequence. Finally, one Sugeno Adaptive Neuro-Fuzzy Inference System (ANFIS) is trained for each AU. Then, the set of systems is used to describe the facial movements of each sequence in terms of AU's to generate the final descriptor set.

The generation of fuzzy rules for facial expression recognition is through a modification of an algorithm proposed in [32] and ANFIS training. The enhancement consists of the selection of antecedent rule before the rule generation, this allows the generation of a simple model with few rules with a direct impact on the effectiveness of the model. Also, the algorithm was modified to assign a semantic to each rule, facilitating interpretation. Such semantics assignment was made by partitioning the membership range [0,1].

The proposed methodology is shown in the algorithm 1.

Algorithm 1 Facial expression recognition modeling with fuzzy rules over action units

Require: A set of facial landmarks $st = \{\{x_i, y_i, x_2, y_2, ..., x_n, y_n\}, t\}$

Ensure: A membership value μ_{au} for each considered action unit, a membership value μ_{fe} for each facial expression.

procedure FERFAU(st)
    $nls \leftarrow LandmarksAlignment(st)$ ▷ Facial landmarks alignment
    $BasicCar \leftarrow FeatureExtraction(nls)$ ▷ Feature extraction starts from here
    $PooledCar \leftarrow Pooling(BasicCar)$
    for each $action\_unit$ **do**
        $au\_models \leftarrow FuzzyModeling(PooledCar)$
    end for
    $\mu_{au} \leftarrow FuzzyModelsEvaluation(au\_models)$
    for each $facial\_expression$ **do** ▷ Classification starts from here
        $fe\_models \leftarrow FuzzyModeling(m_{au})$
    end for
    $\mu_{fe} \leftarrow FuzzyModelsEvaluation(fe\_models)$
end procedure

4.1 Facial landmarks alignment

AAM and others algorithms to identify facial landmarks, retrieve 68 coordinates pairs $st = \{\{x_i, y_1, y_1, y_2, ..., x_{68}, y_{68}\}, t\}$ from the face images describing the face shape, each one corresponding to a vertex of a face descriptor set for each image I of the image sequence $is = \{I, t\}$. When is required to compare two sets of facial landmarks, one method to align the face shape $st_{t=0}$ with $st_{t\neq 0}$ is Procrustes Analysis PA [12], [10]. PA superimposes shapes by optimally translating, rotating, and uniformly scaling objects. PA mitigates geometric distortions in images or landmarks in terms of affine transformations. Although, PA is highly sensitive to noise, geometric distortions, and outliers values. When a facial expression is represented usually a geometric distortion happens, thus, PA could not be the best choice to align facial landmarks.

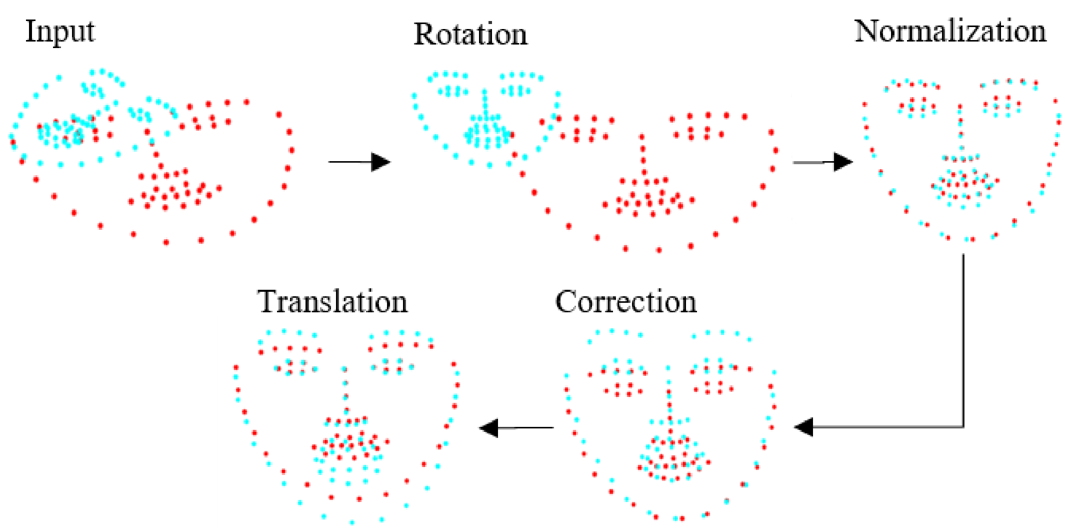


Figure 4.1: Graphical representation of the two facial landmarks alignment Processes

Facial landmarks are aligned to reduce the noise in the system using a heuristic method based on affine transformations, ensuring invariance to scale, orientation and translation. The full process is depicted in Fig. 4.1. To align facial landmarks, first, shape values are normalized between [0,1]. Later, to maintain the deformations caused by facial movement a correction is performed. There are two variants for performing the proposed method; The first variation of the facial landmarks alignment method consists of a size variation correction in terms of the neutral state. In the case of the first variation, the max size of the neutral state is used in all landmarks to perform the normalization between [0,1] instead of the individual size of each landmark set for a given time t of the landmarks set st. The second variation consists on obtain the aspect ratio of the face $ar = \frac{vms}{hms}$, vms being the vertical size of the face and hms being the horizontal size of the face.

The proposed heuristics are shown in Algorithm 2.

Algorithm 2 Facial landmarks alignment

Require: A set of facial landmarks $st = \{x_i, y_i, x_2, y_2, ..., x_n, y_n, t\}$, facial reference points $r = \{x_1, , y_1, ..., x_n, y_n\} | r \subset s$

Ensure: An alignment set of facial landmarks $as = \{x_i, y_i, x_2, y_2, ..., x_n, y_n, t\}$

procedure AMAR(S,r)
    $s \leftarrow$ Rotation of landmarks to align horizontally eye canthus
    $AspectRatio \leftarrow ObtainAspectRatio(s, ns)$
    $ns \leftarrow LandmarksNormalization(s)$ ▷ Normalization in the range [0,1]
    $as \leftarrow$ Multiplication of $AspectRatio$ with ns ▷ Correction of face sizes
    $ns \leftarrow$ Superimposition of the neutral state by affine translations
end procedure

procedure AITNS(S,r)
    $s \leftarrow$ Rotation of landmarks to align horizontally eye canthus
    $sz \leftarrow$ Max size of the neutral state
    $ns \leftarrow LandmarksNormalization(s, sz)$ ▷ Normalization between [0,1]
    $as \leftarrow$ Superimposition of the neutral state by affine translations
end procedure

4.2 Feature extraction

The performance of a predictive model is often dependent on the chosen representation for data [15]. Choosing an appropriate representation is relevant to boosting classification rates. In this chapter, the proposed representation is described. The proposed representation consists of three steps.

First, the magnitude and orientation angle of each facial landmark between the facial landmark final frame where $t = f$ and the initial frame where $t = 0$ of a i-th sample are computed and concatenated in the following tuple $mo_i = [m_1, o_1, m2, m_2, ..., m_n, o_n]$. where f represents the final frame of the i-th image sequence, m_i and o_i are the magnitude and orientation of the movement of the i-th sample respectively, with n being the number of landmarks. Then a triangulated shape of the facial landmarks obtained from [24] was used to obtain a new vector with the area of each triangle $ac = [a_1, a_2, , ..., a_j]$ with j being the number of triangles. Finally, mo and ac are concatenated to obtain a raw 243-dimensional feature vector $br_i = [mo, ac]$. The feature vector is subsequently pooled to obtain a compact representation.

4.2.1 Methodology

After facial landmarks alignment, is possible to subtract the neutral state from the facial expression, this is called AU0 normalization and is related to obtaining the displacements of facial landmarks in an image sequence. Given a set of landmarks $st = \{\{x_i, y_i, x_2, y_2, ..., x_n, y_n\}, t\}$ AU0 normalization consists on the subtraction of the coordinates between the final frame t_f and the initial frame t_0, f being the number of frames of an image sequence is. Using the displacements of facial landmarks dx_i and dy_i the movement magnitude $m_i = \sqrt{(dx_i)^2 + (dy_i)^2)}$ and the movement direction $d_i = \tan^{-1}(\frac{dx_i}{dy_i})$ for each facial landmark i is calculated.

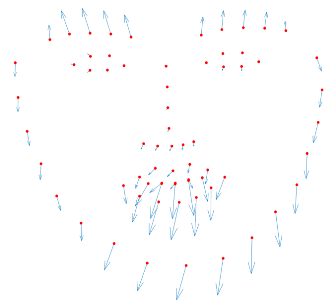
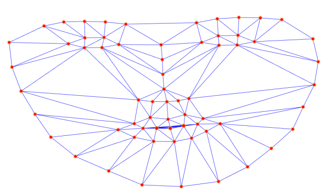
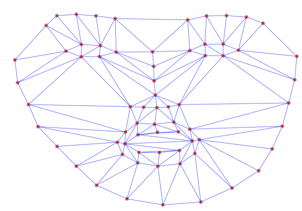


Figure 4.2: Movement magnitude and orientation of facial landmarks for a surprise representation

A triangulated shape $ts = \{a_1, b_1, c_1, a_2, b_2, c_2, ..., a_n, b_n, c_n\}$ was obtained from [24] and it describes facial shape in triangles. The triangulated shape is defined by where a, b and c are vertexes of a triangle. In Fig. 4.3 the used triangulated shape is shown. Using the triangulated shape the change in the size of facial distinctive areas is calculated. To calculate the change in size first the area of the i triangle for the neutral state is computed (where $t = 0$), then the same process is performed with the facial landmarks of the facial expression (where $t = f$), f being the number of the final frame of an image sequence.



(a) Neutral state



(b) Facial expression state

Figure 4.3: Triangulated shape shown in [24]

The number of possible rules generated using the basic feature representation can be obtained with the factorial of the number of features (which is 243!). Fuzzy models with a higher number of rules usually exhibit higher accuracies than one with a less number (a trivial consequence of increasing the model parameters), but in losing simplicity, they lose the ability to explain why the model is making a decision and are more prone to overfilling. We thus strive to reduce the dimensionality of the representation to generate a simpler model affording fewer rules. Given a chosen endpoint, many automatic strategies can search for optimal or suboptimal representations (like genetic algorithms). However, given our interest, we opted for a manual exploration of the data. Following this exploration, distinctive areas of the face were manually chosen. Magnitude and orientation: [Inner eyebrow, Outer eyebrow, eyelids, nose, upper lip, lower lip, right corner lip, left corner lip, jaw, Lips corners] Areas: [eyes, mouth]. A pooling operation then aggregates the selected local descriptors into a subset of the feature representation describing one facial distinctive area. The original 243-dimensional basic representation br is thus reduced to a 22-dimensional representation in which 20(=10x2) values are related to the magnitude and orientation of facial landmarks displacement mo and 2 to the size of the area change ac. Two pooling operations were taking into account to embed the features due its simplicity: average pooling $PooledCar_i = \frac{1}{p}\sum_{i=1}^{p} v_i$ and the max pooling $PooledCar_i = max(v_i)$ where i is the i-th described distinctive area and v_i is a subset of facial landmarks related to the facial area being embedded. Using the pooled vector $PooledCar$, a fuzzy model is trained for each AU. Then, the set of systems is used to describe the facial movements of each sequence in terms of AU's obtaining a feature vector μ_{au} with a membership value for each AU. Algorithm 3 describes the feature extraction process.

Algorithm 3 Feature representation

Require: A set of normalized facial landmarks $ns = \{\{x_i, y_i, x_2, y_2, ..., x_n, y_n\}, t\}$, a triangulated shape $ts = \{x_1, y_1, z_1, x_2, y_2, z_2, ..., x_n, y_n, z_n\}$

Ensure: A representation describing facial movements μ_{au}

procedure MAR(ns)
    $f \leftarrow$ cardinality of ns
    for $i \leftarrow 1$, number of landmarks **do**
        $dx_i \leftarrow x_{i,t=f} - x_{i,t=0}$
        $dy_i \leftarrow y_{i,t=f} - y_{i,t=0}$
        $m_i \leftarrow Distance(dx_i, dy_i)$ $\triangleright\ m_i = \sqrt{(dx_i)^2 + (dy_i)^2)}$
        $o_i \leftarrow Angle(dx_i, dy_i)$ $\triangleright\ d_i = \tan^{-1}(\frac{dx_i}{dy_i})$
        $mo \leftarrow$ concatenation of br, m_i and o_i
    end for
    for $i \leftarrow$ number of triangles in the triangulated shape ts **do**
        $at_{i,0} \leftarrow$ area of the i triangle formed by ns facial landmarks of time 0
        $at_{i,f} \leftarrow$ area of the i triangle formed by ns facial landmarks of time f
        $a_i \leftarrow at_{i,f} - at_{i,0}$
    end for
end procedure
$br \leftarrow$ concatenation of mo and a
for $i \leftarrow 1$, manually chosen facial distinctive areas **do**
    $PooledCar_i \leftarrow$ Pooling($v \subset br$) $\triangleright\ pc_i = \frac{1}{p}\sum_{i=1}^{p} v_i \vee max(v_i)$
end for
for each $action\_unit$ **do**
    $au\_models \leftarrow FuzzyModeling(PooledCar)$
end for
$\mu_{au} \leftarrow FuzzyModelsEvaluation(au\_models)$

4.3 Fuzzy rules generation

The final stage of the model is the classifier considering both the classification and explanation of the labeling. Knowledge is generated using granular fuzzy models in which the information is represented by hyperboxes. A hyperbox is a region of the decision space and can be viewed as a cluster obtained from a clustering operation. For the classification, a Takagi-Sugeno fuzzy inference system is used due to its extended flexibility in system design over the Mamdani fuzzy inference systems [16]. A Takagi-Sugeno model is generated using a modified rule generation algorithm [32] which consists in two steps: hyperboxes generation and rule generation. A Gaussian membership function was used for the hyperboxes. One fuzzy rule is obtained for each cluster obtained through a clustering operation and its standard deviation. The algorithm for fuzzy rule generation was modified here to assign semantics to each rule facilitating interpretation. Such semantics assignment was made by partitioning the membership range [0,1]. For each subset of the partition, semantics is assigned (very weak, weak, medium, strong, and very strong presence). Here we have parameters $\delta_a u$ and $\delta_e xp$ for the AU and facial expression models respectively. Usually, fuzzy rule generation algorithms include a rule selection phase. The rule selection phase is related to reducing the complexity of the model with less accuracy reduction. Usually, the set of fuzzy rules generated contains useless rules, counterproductive rules, or contradictory rules which are needed to extract from the fuzzy rules set. Due to a large number of possible rules, fuzzy rule selection, in general, takes a lot of time to perform. Our fuzzy rule generation algorithm previously the clustering and rule generation phase to perform a rule antecedent selection. With the rule antecedent selection, we intend to reduce the number of rules generated and the complexity of the model. By reducing the complexity of the model, it does not lose the ability to explain why it is making a decision. In the case of our model, two ways to select rule antecedents were used: 1) Manual exploration of the data for

AU model generation. In this case, the antecedents were selected for each AU (only the landmarks related to each AU. 2) Sequential Feature Selection to select features with a higher impact on classification rates.

Algorithm 4 Fuzzy rule generation algorithm with antecedent selection

Require: A vector of features $fv_i = [f_1, f_2, ..., f_n, c]$, where n is the number of features, c the sample class and i the sample number.

Ensure: A fuzzy model

```
procedure FRGAS( ns )
    for each distinct class c in fv do
        a_c ← AntecedentSelection(c)                ▷ Antecedent selection for each class
        cl_c =Clustering(sfv|sfv ⊂ fv, given by the selected parameters a_i )
        for each cluster cn with their standard deviation cd in cl_c do
            fr_cd ← FuzzyGaussianRule(cn,cl)
            fr_cd ← SemanticAssigment(fr_c d)
            A new rule fr_cd is added to FuzzyModel
        end for
    end for
end procedure
```

Chapter 5

Experiments and results

In this chapter, we will present the results obtained during the research. Section 5.1 describe the data sets used. The section , describes describes the model selection. The use of facial landmarks is a limitation of the model, landmarks were extracted and an experiment was performed in order to obtain the effect of extracting facial landmarks with two kinds of methods. Also, results obtained with another database are tested, looking for external validation. The model depends of obtain a neutral state from the subject, section 5.6 presents experiments to see the effect of generating a synthetic model to mitigate this limitation.

5.1 Datasets

5.1.1 Cohn-Kanade Plus

The Cohn-Kanade Plus dataset Cohn-Kanade Plus (CK+) was obtained from [24], it consists of 327 validated image sequences from 123 healthy subjects in which one of the seven basic facial expressions is represented: anger, contempt, disgust, fear, happiness, sadness OR surprise. CK+ is labeled by expert judges according to the FACS [13] and has become one of the most widely used for algorithm development and evaluation. The number of samples contained in the CK+ dataset for each facial expression is shown in table 5.1.

Table 5.1: Frequency for each facial expression from the CK+ dataset

Emotion	N
Angry (En)	45
Disgust (Dis)	59
Contempt (Con)	18
Happy (Hap)	69
Fear (Fe)	25
Sadness (Sad)	28
Surprise (Sur)	83

Each sequence begins with a subject in the neutral state and ends in a facial expression representation. An expert judge, encode manually each sequence using the facial action coding system [14] and labels which AU's are present in the sequence and their intensities. Coding is done in the following form $label = \{[au_1, i_1], [au_2, i_2], ..., [au_n, i_n]\}$, where au is the number of the AU presented in the sequence and i denotes the intensity of the AU . Below an example of the label is presented:

Table 5.2: Label example, each sequence in the dataset contains a similar one

Subject	Sequence	AU labels	Emotion label
01	04	{ [4,0], [7,5], [17,4], [23,4], [24,4] }	Angry

Using the labels contained in the CK+ dataset, a set of AU related to each facial expression were selected and rules for each one were developed (See Fig. 5.3).

The rules obtained from the selection of the AU were used to label each sequence with the represented facial expression in the sequence.

Table 5.3: Emotion description in terms of action units

Emotion	Action Units
Angry (An)	AU23 + AU24 must be present
Disgust (Dis)	AU9 o AU10 must be present
Contempt (Con)	AU14 must be present unilateral or bilateral
Happy (Hap)	AU12 must be present
Fear (Fe)	AU1+2+4 present unless AU5 be E then AU4 can be absent
Sadness (Sad)	AU 1 + 4 + 15 . with exception of AU6 + 15
Surprise (Sur)	AU1+2 o 5 must be present and y AU5 should not be stronger than B

Using Active Appearance Models [25] obtained a set of coordinates for describing the face shape in a set of n coordinates where each coordinate belongs to a vertex of the model that describes the face.

5.1.2 Radboud Faces Database

Set of pictures of 67 models in were Caucasian males and females, Caucasian children, both boys and girls and Moroccan Dutch males display 8 facial expressions including the neutral state. The database was obtained from [20]. Five camera angles were used to obtain the images: 180°, 135°, 90°, 45° and 0°. Although, in this research, only images with 90° were used. As CK+ database, Radboud database was validated using the FACS [13]. The sample number contained in the Radboud database is shown in table 5.7

Table 5.4: Frequency for each facial expression from the Radboud dataset

Emotion	N
Angry (En)	201
Disgust (Dis)	201
Contempt (Con)	201
Happy (Hap)	201
Fear (Fe)	201
Sadness (Sad)	201
Surprise (Sur)	201

5.2 Parameters definition

Most training algorithms or methodologies have some settings or parameters for which the user needs to choose. In the case of our model, the parameters defined for each step of our methodology were: alignment method (Explained in 4.1) (Without alignment, in terms of the neutral state, maintaining facial aspect ratio, Procrustes superimposition [24]), pooling operation (average or max pooling), clustering type (Fuzzy C-means, Subtractive Clustering), clustering parameters (Cluster number or influence radius for FCM and SC respectively); Being a total of four parameters. In the case of the number of clusters for the FCM algorithm a range between [2,5] was used, on the other hand, the cluster influence radius for subtractive clustering was between [0.2,0.8]. A 0.1 value for cluster influence of subtractive clustering algorithm was discarded due to the high computational cost at training the models and a value greater than 0.8 was not considered for simplicity. The possible number of cases using the selected parameters of the model cases was 840. The simplest approach is to run the proposed methodology with the same training data for each combination of parameters. The method was evaluated using leave-one-out replication allowing the use of the same training/validation data for each combination.

5.3 Model selection

As can be seen in Fig. 5.1, facial landmarks can be affected by orientation, location and size. The proposed methodologies [28] are sensitive to noise and shape deformation (such as those occurring in facial expressions). Two heuristics based on affine transformations (rotation for orientation, translation for location, and scale for size) were proposed in the methodology to align facial landmarks, alignment AITNS and AMAR (for a more detailed explanation see 4.1), of each state(neutral and facial expression).

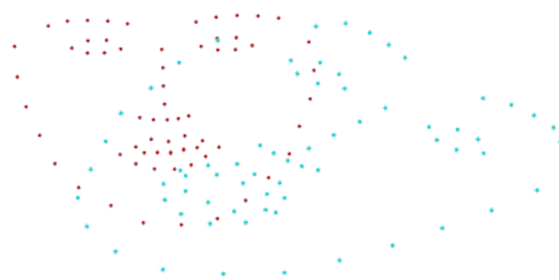


Figure 5.1: Facial landmarks affected by orientation, location, and size

All the possible settings for the parameters were evaluated for each alignment method using a leave-one-out replication method. The results of this experiment are shown in Fig. 5.2. The model Without alignment obtained a mean accuracy of 0.87±0.02. Procrustes analysis obtained the smaller accuracy and the higher deviation with a mean of 0.84±0.27. Alignment In terms of the neutral State obtained the higher mean accuracy of 0.90±0.014 followed by Alignment Maintaining the Aspect Ratio with a 0.88±0.01 mean accuracy.

The previous results show a statistical significance between the alignment methods (ANOVA: p=0). Guided by the higher accuracy, the Alignment in terms of the neutral state was selected for consequent experiments.

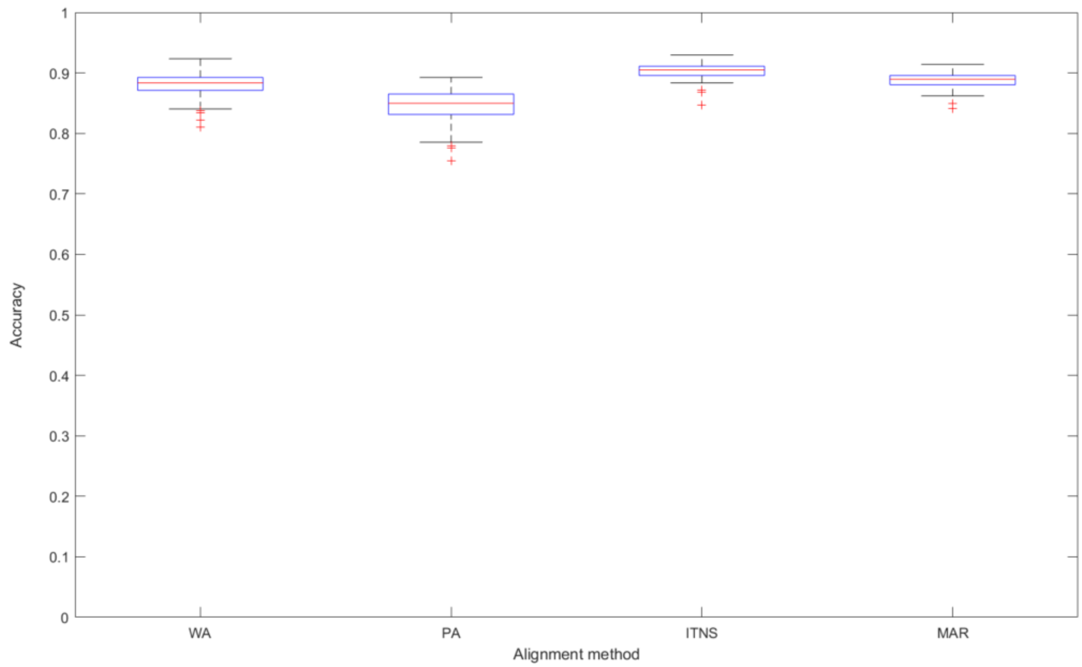


Figure 5.2: Box plot corresponding to the results for the 4 alignments (without Alignment (WA), Procrustes Analysis(PA), Alignment In terms of the Neutral state(AITNS), Alignment Maintaining Aspect Ratio(AMAR).

Later, the pooling operations used were analyzed using the proposed model in order to identify which one obtain the best accuracy with the minimum deviation. The average pooling obtained a mean accuracy of 0.90±0.01, which is over the max pooling which obtained a mean accuracy of 0.89±0.015. The results obtained show statistical significance (ANOVA: p=0.01).

Fuzzy rule generation begins with a clustering operation to generate the hyperboxes to model the data. Two clustering algorithms were analyzed to see the accuracy: fuzzy c-means and subtractive clustering. In the case of fuzzy c-means, the number of clusters generated varies in the range [2,5]. In the case of subtractive clustering, the radius influence varies from 0.2 to 0.8. Smaller values were not used due to the computational cost that implies.

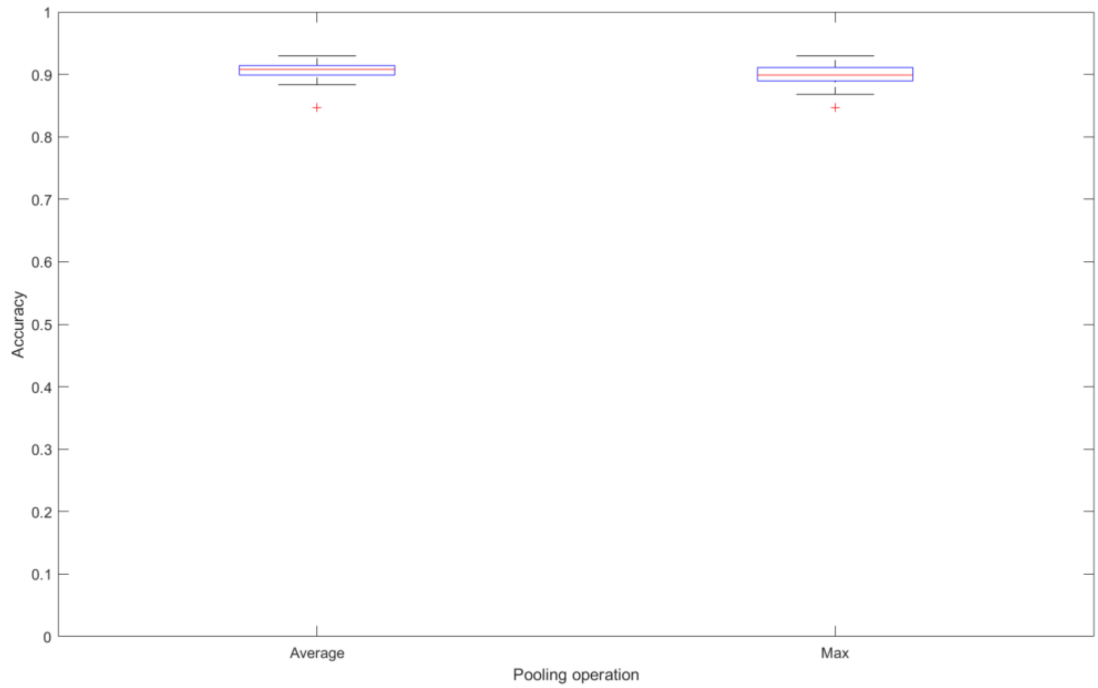


Figure 5.3: Box plot with results of the two pooling operations used: average pooling and max pooling.

The results obtained for the clustering algorithm are presented in Fig. 5.4, as can be seen, subtractive clustering obtained a higher accuracy than the fuzzy c-means algorithm, although also obtained a smaller standard deviation. Subtractive clustering has significantly better accuracy than Fuzzy C-means clustering (ANOVA: p=0.04).

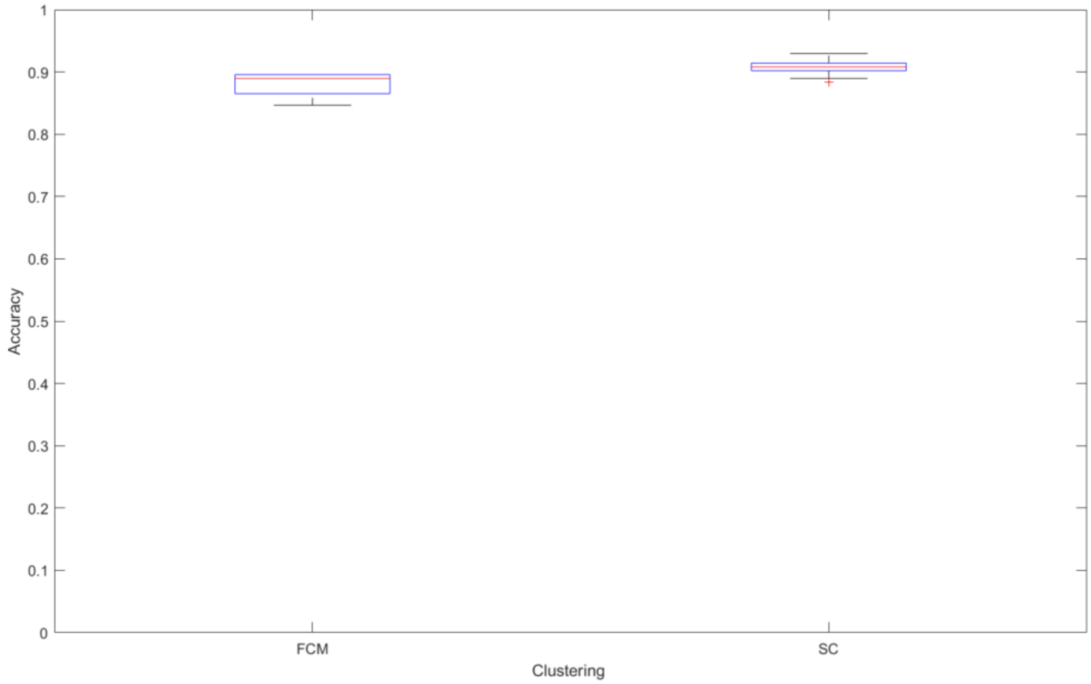


Figure 5.4: Box plot with results of the model varying the cluster clustering algorithm: Fuzzy C Means (FCM) and Subtractive Clustering (SC).

5.4 Model results

The selected parameters were the following, in the alignment, AITNS, in the pooling the average operation was selected and for modeling subtractive clustering is used. Below, are sensitivity, specificity, precision, f1 score. In order to provide the metrics that better describe the proposed method, also, the Receiver Operator Characteristics analysis is presented. Finally, an analysis of the rules generated can be seen.

In Table 5.5, the confusion matrix for the selected model is shown. In the case of the selected parameters, happiness, surprise and disgust obtained the highest precisions. Anger obtained a precision of 0.84 and is confused with contempt, disgust, and sadness, this can be because of the movements of these emotions are focused on the compression or deformation of the mouth. Contempt obtained the lowest precision (0.74) and its confused frequently with sadness, probably because bilateral dimple causes a similar deformation of landmarks when a lip corner depressor is found.

Table 5.5: confusion matrix for the selected model.

	An	Con	Dis	Fe	Hap	Sad	Sur
An	**0.84**	0.05	0.06	0.00	0.00	0.04	0.00
Con	0.01	**0.74**	0.00	0.01	0.07	0.14	0.02
Dis	0.02	0.04	**0.92**	0.01	0.00	0.01	0.00
Fe	0.00	0.01	0.03	**0.82**	0.06	0.07	0.01
Hap	0.00	0.00	0.00	0.01	**0.98**	0.00	0.01
Sad	0.06	0.05	0.00	0.06	0.00	**0.83**	0.00
Sur	0.00	0.01	0.00	0.01	0.00	0.01	**0.96**

The precision obtained for each facial expression is shown below in Fig. 5.5. In this case, the facial expression with greater precision is surprise and happiness, while contempt obtained the lowest precision with a value below 0.8.

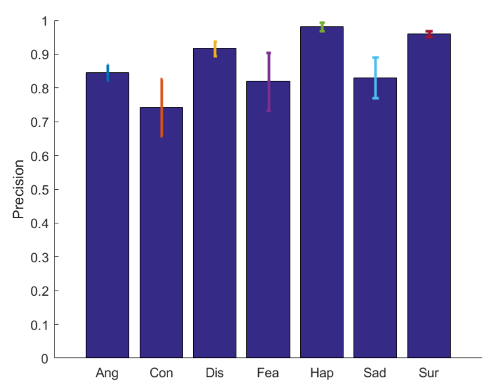


Figure 5.5: Mean precision by facial expression using the selected parameters.

The f1 score obtained by facial expression is shown in Fig. 5.8. Anger, disgust, fear, happiness, and surprise obtained a score over 0.8. On the other hand, contempt obtained the lowest score followed by sadness with values below 0.8.

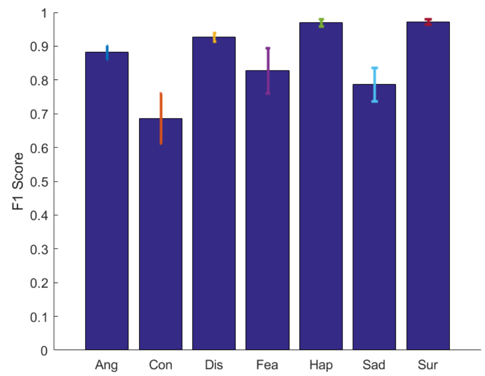


Figure 5.6: Mean f1 score obtained with selected parameters.

Finally, the ROC curve is presented in Fig. 5.7. In this case, the Area Under the Curve (AUC) for anger, disgust, fear, happiness, sadness, and surprise are all up to 0.9, except for contempt with an AUC of 0.873. The average AUC is 0.96.

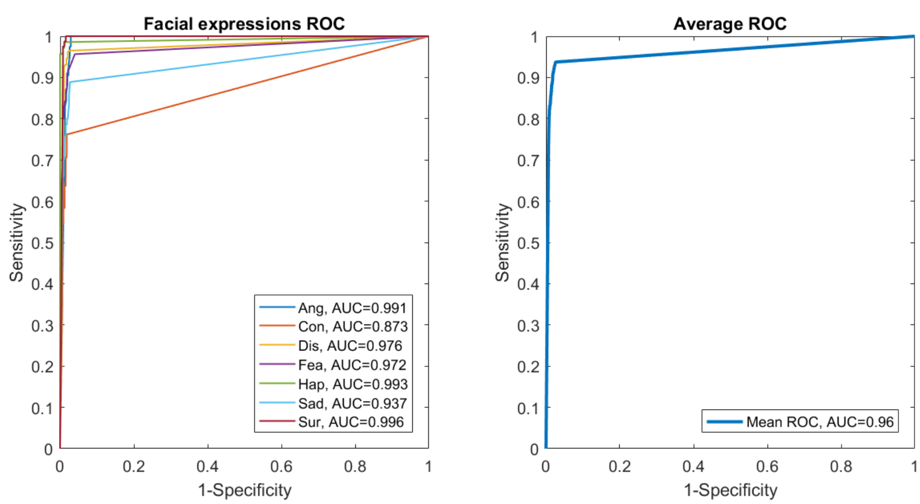


Figure 5.7: ROC curve for each facial expression with selected parameters, in the right the average ROC is presented.

5.5 Database testing and landmarks extraction method effect

The proposed method was tested in two databases to analyze its behavior with the selected parameters. First, the Cohn-Kanade dataset proposed in [24] was tested, this database includes 327 samples of the 7 basic facial expressions: anger, contempt, disgust, fear, happiness, and surprise. On the other hand, the Radboud database was also tested, it was proposed in [20] and also contains the 7 basic facial expressions, in addition to a neutral state. Radboud dataset does not contain any landmarks points, two algorithms were used to obtain them. To obtain the landmarks points a methodology proposed by Deva-Ramanan in [33] and Openface [4] were used. For a better explanation of the databases see Section 5.1. For the CK+ database, the Deva-Ramanan facial landmarks were not obtained due to the limitations of the algorithm.

The parameters selected were, for the alignment method the alignment in terms of the neutral state was selected, for the pooling the average pooling was used, for the clustering type Subtractive clustering with $\gamma au.02$ was considered. The dataset was partitioned into 70% for training and 30% for validation.

The classification results were variable among databases. for the CK+ database with the dataset landmarks a mean accuracy of 0.91 ± 0.03 was obtained, with a significant difference (ANOVA: $p<0.05$) between the same CK+ with the landmarks obtained with Openface, which obtained a mean accuracy of 0.76 ± 0.04. In the case of the raFD database, a mean accuracy of 0.81 ± 0.024 was obtained with the Openface landmarks contrasted with a 0.45 ± 0.03 of the Deva-Ramanan Landmarks showing that the landmarks extraction method can have a direct effect in the accuracy (ANOVA: $p< 0.05$)

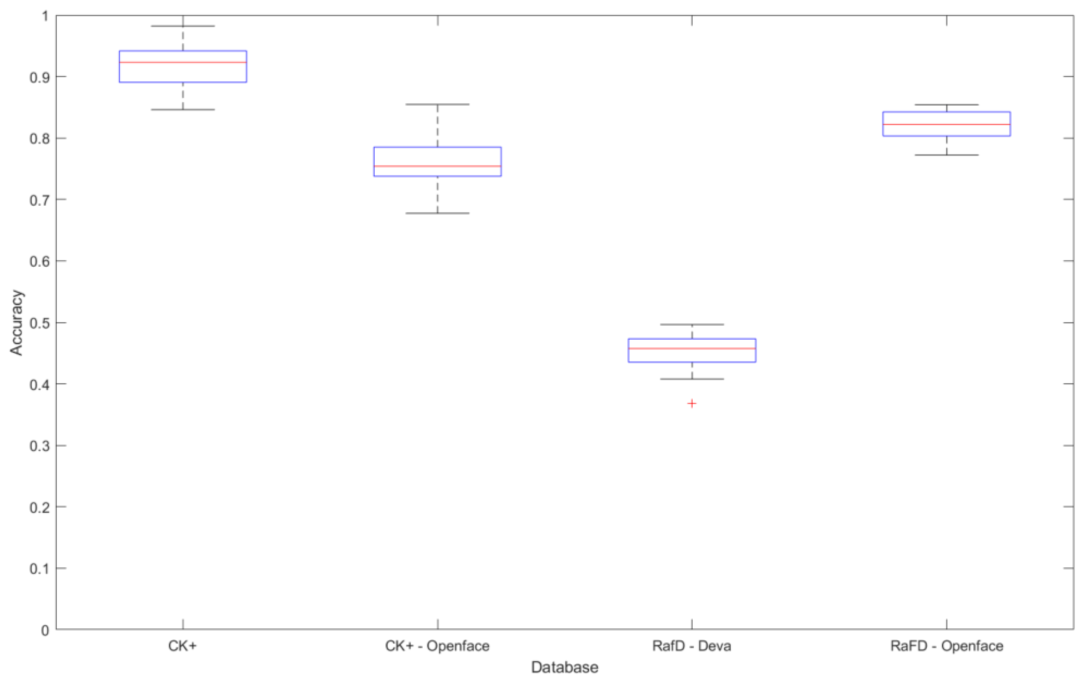
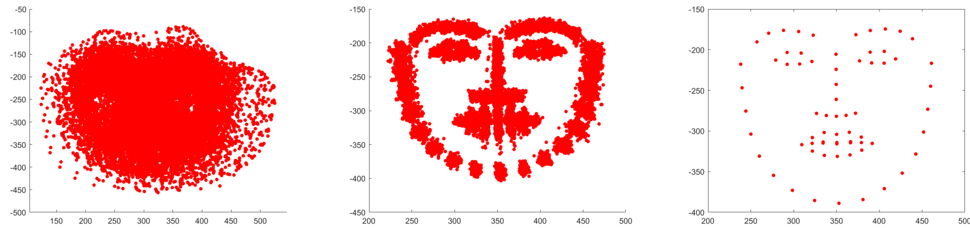


Figure 5.8: Box plot for the CK+ and the RaFD databases with the following parameters: alignment in terms of the neutral state, average pooling operation, and subtractive clustering varying γ between [0,1] with 70% of the samples for training and 30% for validation.

5.6 Synthetic model

One of the limitations of the model is the necessity of a neutral state of the subject to obtain the features. The AU0 normalization is the subtraction of the neutral state to a given facial expression. As an attempt to alleviate this limitation, a synthetic model was generated to perform the feature extraction method instead of using the neutral state shown in the dataset.

The neutral state model was generated using a random sample of a neutral state in the dataset selected, then Procrustes superimposition was used to fit the remaining samples to the selected one (Fig. 5.9). A mean operation for each landmark was performed in order to generate the model. Procrustes superimposition also can be used to obtain a model from landmarks set to describe shapes.



(a) All samples (b) After Procrustes (c) Generated model

Figure 5.9: Generation of the synthetic neutral state.

The experiment was performed with the CK+ and the RafD database with the selected model (Sec: 5.3). The results are shown in Fig. 5.10. The results vary, in three of the study cases, the mean accuracy decays from 0.91±0.03 to 0.70±0.07 for the CK+ database with the landmarks included in the dataset. Also for the CK+ but with facial landmarks obtained with Openface from 0.76±0.04 to 0.63±0.07. In the case of the RaFD database with the landmarks obtained with the method obtained from [33] the accuracy increased from 0.45±0.03 to 0.62±0.05. On the other hand for the RaFD database with landmarks obtained through Openface the accuracy decreased from 0.81±0.02 to 0.61±0.05.

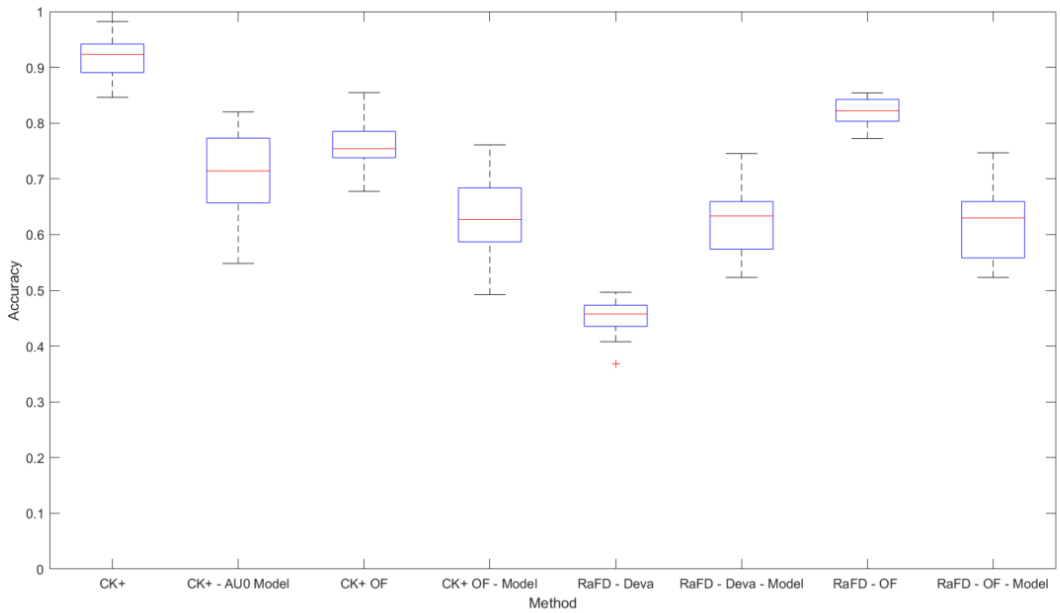


Figure 5.10: Mean accuracy obtained using the subject neutral state or a synthetic model.

5.7 Cross-Database facial expression recognition

Facial expression recognition methods have problems obtaining high accuracies when evaluated using the cross-database validation protocol. Even though the environment is controlled within databases (frontal faces, illumination, occlusions, among others) it is not controlled across databases. [41]. The experiments were organized in the following way. The selected model trained with the CK+ database [24] from Sec. 5.3 was used to evaluate each frontal sample in the RaFD database [20] like in a leave-one-out replication method with a variation in the clustering influence for the subtractive clustering algorithm γ_{exp} between [0.2,0.8]. The obtained accuracy was of 0.69±0.01.

Table 5.6: Comparison with related work. The six databases used are the following: JAFFE, CK+, MMI, RaFD, KDEF, BU3DFE, ARFace.

Method	Train	Target	Interpretable	AU based	Accuracy
[41]	6 databases	RaFD	no	no	0.85±0.04
[3]	JAFFE	RaFD	no	no	0.52±N/R
[3]	TFEID	RaFD	no	no	0.55±N/R
Proposed	CK+	RaFD	yes	yes	0.69±0.01

5.8 Fuzzy rules generated

The performance of the fuzzy rule generation algorithm is not only given by the accuracy of the prediction model. But also the number of rules generated for the models is required.

Table 5.7: Rules generated for each γ for the facial expressions models.

Cluster influence γ	Mean cluster number	Mean Accuracy
$\gamma = 0.2$	41.7±20.62	0.76±0.15
$\gamma =0.3$	31.1±8.1	0.79±0.14
$\gamma =0.4$	27.0±6.9	0.78±0.17
$\gamma =0.5$	23.5±4.7	0.77±0.17
$\gamma =0.6$	21.6±3.5	0.77±0.17
$\gamma =0.7$	18.9±3.3	0.79±0.17
$\gamma =0.8$	17.2±3.3	0.79±0.15

5.9 Discussion

Many efforts have been tackle facial expression recognition but current solutions have favored discriminative over explicative power. The results obtained with the proposed method are similar to recent works that have used a dynamic approach or facial landmarks for this task. [24] begins with a normalization step using Procrustes analysis, later a similar normalized shape is obtained. These are the normalized facial landmarks with the AU0 normalization obtaining an accuracy of 77±2.9 % for their models with a maximum of 88%. On the other hand, [17] proposed a descriptor that captures the changes of 560 angles obtained with the combination of facial landmarks. This descriptor can have three values, -1 when the angle decrease, 0 when the angle does not move, and 1 when the angle increase, and it is invariant to pose, for this reason, a normalization or alignment is not required.

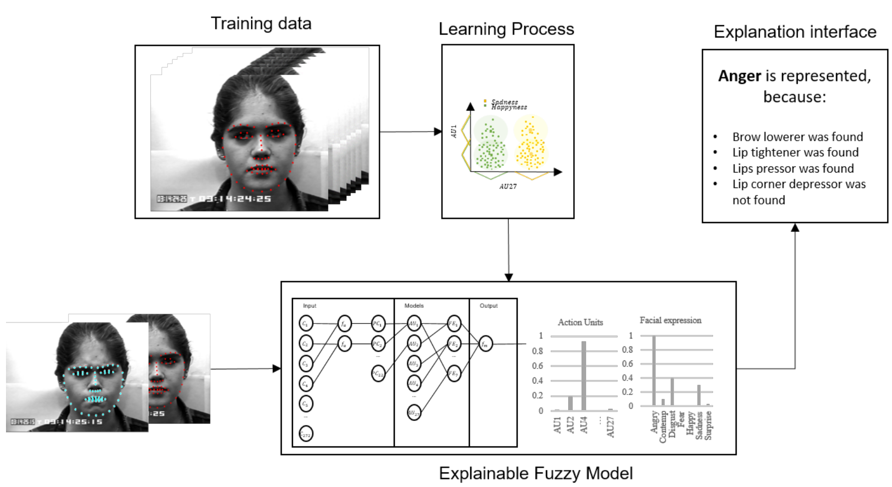


Figure 5.11: Fuzzy model for facial expression recognition

The proposed methodology starts with a heuristic based on affine transformations rather than a Procrustes analysis [24], as can be seen in Fig. 5.11. In our model, the use of these heuristics improves significantly the accuracy over Procrustes superimposition or without aligning the facial landmarks. The full feature representation

is embedded into a compact one by means of pooling, which reduce the complexity of the model obtaining a 22-dimensional representation that is smaller than the others presented in the literature [17],[3],
[42]. A Sugeno-type fuzzy inference system is used for each action unit to generate a description of the movements in the face that identifies the facial expression present in an image sequence. Obtaining a mean accuracy of 90.8±14 with a maximum of 92.9±28. Also, the results are interpretable which can lead to generating an interface that explains why the model is making a decision. The proposed method uses the subject neutral state as calibration (as the others). Experiments were realized in order to know the effects of using a synthetic model to avoid the dependency on the subject's neutral state. Experiments have shown that our model can obtain accuracies greater than 0.70 without calibration which is an initial step to alleviate the necessity of an initial neutral state image.

Table 5.8: Comparison with related work

Work	Data	Method	Interpretable	FACS	Mean Accuracy
[24]	CK+	CAPP	no	no	0.80±N/R
[24]	CK+	S+C	no	no	0.88±N/R
[17]	CK+	ORB	no	no	0.92±N/R
[18]	JAFFE	MFS	14 rules	no	0.87±N/R
[15]	JAFFE	FRM	565 rules	no	0.96±N/R
[33]	CK+	RM	no	no	0.85±N/R
[42]	CK+	no	no	no	0.97±N/R
Proposed	CK+	FERFAU0	yes	yes	0.70±0.07
Proposed	CK+	FERF	yes	yes	0.91±0.03
Proposed	RaFD	FERFAU0	yes	yes	0.61±0.05
Proposed	RaFD	FERF	yes	yes	0.81±0.61

In Table 5.12, the emotion description generated by the fuzzy model is shown.

A green dot indicates when an AU needs to be present in a sequence in order to accumulate evidence and obtain a bigger membership value of the emotion. A red dot indicates when an action unit must be absent in the sequence, these are called counterexamples. For example, if AU1 and AU25 are found, then the evidence is accumulated for fear and surprise (both need these AU to accumulate evidence) but if AU27 is found, then the membership value of fear is reduced so that the model does not confuse surprise with fear.

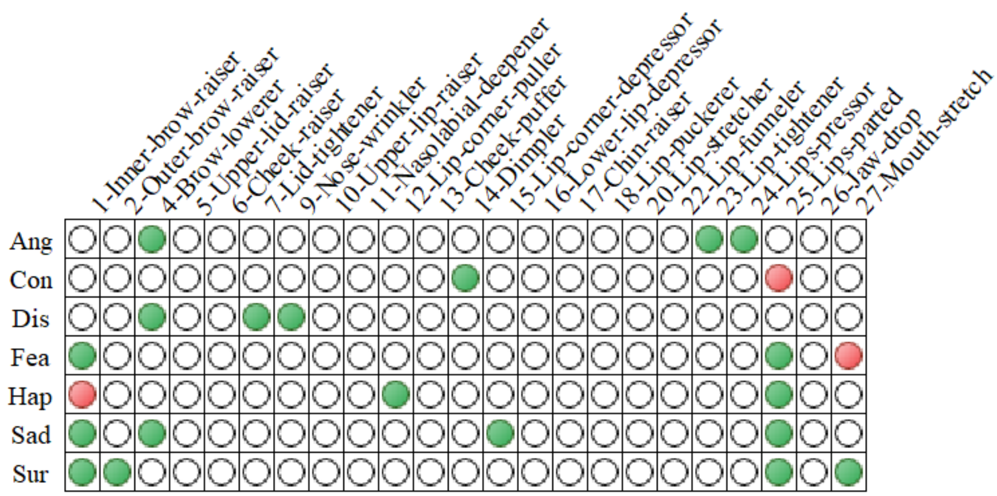


Figure 5.12: Emotion description in terms of action units.

Chapter 6

Conclusions and future work

6.1 Conclusions

In this work, the design, implementation, and experiments realized with a simple model for facial expressions recognition are based on the movements of facial distinctive areas presented. This model improved fuzzy models which explain through rules the appearance of the face, it differs from the models presented in the state of the art in the description given. The proposed model used an appearance system that describes facial expressions in terms of movements of facial distinctive areas giving a more detailed explanation of what movements are in the face and why are making a decision while keeping the semantic meaning of the facial action coding system.

The obtained results have shown that it was possible to pool a feature representation that encodes the movements presented in the face into a compact one, losing little information but keeping necessary to detect patterns in the facial landmarks after the AU0 normalization, obtaining an average accuracy of 90.8±14% with a maximum accuracy of 92.9±28%.

It was shown that the alignment process is an important step and has a primary impact on the accuracy, i. e. the noise introduced by orientation, size, and location of facial landmarks can be a weakness of the model if it is not extracted in a proper way.

The clustering algorithm for the generation of the fuzzy rules did not have a great impact by itself because a great selection of the clustering parameters can improve the effectiveness of the model due to the mixed effects present in this step. Results obtained in this work demonstrated that it is possible to develop a fuzzy explainable model using the description of emotions in terms of AU's based on the facial action coding system and expert knowledge. An algorithm of fuzzy rule antecedents selection can model the data using hyperboxes with a low number of rules (Mean of 25.9 ± 8.4 rules).

In an attempt to expand the method to another database a set of landmarks was extracted using two methods, the obtained results suggest that detailed training and parameter selection can improve the obtained results. Is important to remark that the results suggest that the facial landmarks selection method had a direct impact on the model efficiency.

A synthetic model synthesized using the training data was generated as an attempt of mitigating one of the model limitations, the need for the neutral state to extract the features. As expected in most cases the accuracy went lower than using the real neutral state of the subject. Even so the accuracy was over random. But, in order to obtain a significant improvement is need to use the calibration process or a better synthesis of the neutral state per subject.

6.2 Future work

After analyzing the obtained results, it has been considered that several issues of the proposed methodology can be improved in future work. Below, the proposals are presented:

- If facial landmarks are going to be used as a tool to detect facial distinctive areas a more reliable and generalized method for facial landmarks alignment based on the Procrustes analysis can be developed, a possible way to do this is to extend the analysis to keep distinctive deformations of the shape which is going to be superposed.

- A better pooling scheme in order to maintain the information captured by the feature representation, in the case of our pooling scheme, it is needed to experiment with using various combinations of features.

- If facial landmarks are not going to be used, a more sophisticated method for facial distinctive areas and movement detection such as optical flow can be used, using a pooling scheme to group the movement in the distinct facial areas and in this way extend the method to databases which not contains facial landmarks, also, extending it to a different angle in which the face can be observed.

Bibliography

[1] Fuzzy Logic. In *Fuzzy Control of Queuing Systems*, pages 15–25. Springer, London, 2005. DOI: 10.1007/1-84628-073-7_2.

[2] S. Agrawal and P. Khatri. Facial Expression Detection Techniques: Based on Viola and Jones Algorithm and Principal Component Analysis. pages 108–112, Feb. 2015.

[3] G. Ali, M. A. Iqbal, and T.-S. Choi. Boosted NNE collections for multicultural facial expression recognition. *Pattern Recognition*, 55(Supplement C):14–27, July 2016.

[4] B. Amos, B. Ludwiczuk, and M. Satyanarayanan. OpenFace: A general-purpose face recognition library with mobile applications. *hgpu.org*, July 2016.

[5] M. Berger, editor. *Geometry I*. Universitext. Springer Berlin Heidelberg, Berlin, Heidelberg, 1987. DOI: 10.1007/978-3-540-93815-6.

[6] J. C. Bezdek, R. Ehrlich, and W. Full. FCM: The fuzzy c-means clustering algorithm. *Computers & Geosciences*, 10(2):191–203, Jan. 1984.

[7] Y.-L. Boureau, J. Ponce, and Y. Lecun. A Theoretical Analysis of Feature Pooling in Visual Recognition. In *27th International Conference on Machine Learning, Haifa, Israel*, 2010.

[8] A. M. Burrows, B. M. Waller, L. A. Parr, and C. J. Bonar. Muscles of facial expression in the chimpanzee (Pan troglodytes): descriptive, comparative and phylogenetic contexts. *Journal of Anatomy*, 208(2):153–167, Feb. 2006.

[9] J. A. Coan and J. J. B. Allen. *Handbook of Emotion Elicitation and Assessment*. Oxford University Press, Apr. 2007. Google-Books-ID: ChiiBDGyewoC.

[10] J. F. Cohn, A. J. Zlochower, J. Lien, and T. Kanade. Automated face analysis by feature point tracking has high concurrent validity with manual FACS coding. *Psychophysiology*, 36(1):35–43, Jan. 1999.

[11] J. M. V. Cohn, J. F. Depression, smiling and facial paralysis. *Facial palsies. Amsterdam, the Netherlands: Lemma Holland*, 2005.

[12] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):681–685, June 2001.

[13] P. Ekman. Facial expression and emotion. *American Psychologist*, 48(4):384–392, 1993.

[14] P. Ekman and W. Friesen. *Facial Action Coding System: A Technique for the Measurement of Facial Movement.* Consulting Psychologists Press, 1978.

[15] R. Ghasemi and M. Ahmady. Facial expression recognition using facial effective areas and Fuzzy logic. In *2014 Iranian Conference on Intelligent Systems (ICIS)*, pages 1–4, Feb. 2014.

[16] A. Hamam and N. D. Georganas. A comparison of Mamdani and Sugeno fuzzy inference systems for evaluating the quality of experience of Hapto-Audio-Visual applications. In *2008 IEEE International Workshop on Haptic Audio visual Environments and Games*, pages 87–92, Oct. 2008.

[17] F. Iglesias, P. Negri, M. E. Buemi, D. Acevedo, and M. Mejail. Facial expression recognition: a comparison between static and dynamic approaches. In *International Conference on Pattern Recognition Systems (ICPRS-16)*, pages 1–6, Apr. 2016.

[18] A. Khanam, M. Z. Shafiq, and M. U. Akram. Fuzzy Based Facial Expression Recognition. In *Congress on Image and Signal Processing, 2008. CISP '08*, volume 1, pages 598–602, May 2008.

[19] S. Krig. Feature Learning and Deep Learning Architecture Survey. In *Computer Vision Metrics*, pages 375–514. Springer International Publishing, 2016. DOI: 10.1007/978-3-319-33762-3_10.

[20] O. Langner, R. Dotsch, G. Bijlstra, D. H. J. Wigboldus, S. T. Hawk, and A. v. Knippenberg. Presentation and validation of the Radboud Faces Database. *Cognition and Emotion*, 24(8):1377–1388, Dec. 2010.

[21] H. Li, C. C. L. Philip, and H.-P. Huang. *Fuzzy Neural Intelligent Systems: Mathematical Foundation and the Applications in Engineering*. CRC Press, Inc., Boca Raton, FL, USA, 1st edition, 2000.

[22] W. f. Liu, S. j. Li, and Y. j. Wang. Automatic Facial Expression Recognition Based on Local Binary Patterns of Local Areas. In *WASE International Conference on Information Engineering, 2009. ICIE '09*, volume 1, pages 197–200, July 2009.

[23] D. G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2):91–110, Nov. 2004.

[24] 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 *2010 IEEE Computer Society Conference on*

Computer Vision and Pattern Recognition - Workshops, pages 94–101, June 2010.

[25] I. Matthews and S. Baker. Active Appearance Models Revisited. *International Journal of Computer Vision*, 60(2):135–164, Nov. 2004.

[26] P. Mitteroecker, P. Gunz, S. Windhager, and K. Schaefer. A brief review of shape, form, and allometry in geometric morphometrics, with applications to human facial morphology. *Hystrix, the Italian Journal of Mammalogy*, 24(1):59–66, May 2013.

[27] M. R. Mohammadi and E. Fatemizadeh. Fuzzy local binary patterns: A comparison between Min-Max and Dot-Sum operators in the application of facial expression recognition. In *2013 8th Iranian Conference on Machine Vision and Image Processing (MVIP)*, pages 315–319, Sept. 2013.

[28] M. Pantic and L. J. M. Rothkrantz. Automatic analysis of facial expressions: the state of the art. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1424–1445, Dec. 2000.

[29] S. Park and D. Kim. Spontaneous facial expression classification with facial motion vectors. In *8th IEEE International Conference on Automatic Face Gesture Recognition, 2008. FG '08*, pages 1–6, Sept. 2008.

[30] K. M. Passino and S. Yurkovich. *Fuzzy Control*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1st edition, 1997.

[31] G. Peleg, G. Katzir, O. Peleg, M. Kamara, L. Brodsky, H. Hel-Or, D. Keren, and E. Nevo. Hereditary family signature of facial expression. *Proceedings of the National Academy of Sciences*, 103(43):15921–15926, Oct. 2006.

[32] A. Priyono, M. Ridwan, A. J. Alias, R. A. O. K. Rahmat, A. Hassan, and M. A. M. Ali. Generation of Fuzzy Rules with Subtractive Clustering. *Jurnal Teknologi*, 43(1):143–153, Feb. 2012.

[33] D. Ramanan. Face Detection, Pose Estimation, and Landmark Localization in the Wild. In *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, CVPR '12, pages 2879–2886, Washington, DC, USA, 2012. IEEE Computer Society.

[34] O. F. Reyes-Galaviz and W. Pedrycz. Granular fuzzy models: Analysis, design, and evaluation. 64:1–19, Sept. 2015.

[35] C. A. Reyes-García. Notal del curso: sistemas hibridos avanzados, 2016.

[36] C. Shan, S. Gong, and P. W. McOwan. Facial expression recognition based on Local Binary Patterns: A comprehensive study. *Image and Vision Computing*, 27(6):803–816, May 2009.

[37] S. Torcida and S. Ivan Perez. Análisis de Procrustes y el estudio de la variación morfológica/ Procrustes analysis and the study of morphological variation. 14, May 2012.

[38] D. M. Vo, A. Sugimoto, and T. H. Le. Facial expression recognition by re-ranking with global and local generic features. In *2016 23rd International Conference on Pattern Recognition (ICPR)*, pages 4118–4123, Dec. 2016.

[39] L. A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning—I. *Information Sciences*, 8(3):199–249, Jan. 1975.

[40] L. A. Zadeh. Is there a need for fuzzy logic? *Information Sciences*, 178(13):2751–2779, July 2008.

[41] M. V. Zavarez, R. F. Berriel, and T. Oliveira-Santos. Cross-Database Facial Expression Recognition Based on Fine-Tuned Deep Convolutional Network. In *2017 30th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, pages 405–412, Oct. 2017.

[42] H. Zhao, Z. Wang, and J. Men. Facial Complex Expression Recognition Based on Fuzzy Kernel Clustering and Support Vector Machines. In *Third International Conference on Natural Computation (ICNC 2007)*, volume 1, pages 562–566, Aug. 2007.

[43] K. Zhao, S. Yang, A. Wiliem, and B. C. Lovell. Landmark manifold: Revisiting the Riemannian manifold approach for facial emotion recognition. In *2016 23rd International Conference on Pattern Recognition (ICPR)*, pages 1095–1100, Dec. 2016.