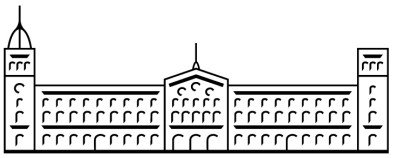


Treball final de grau

GRAU DE MATEMÀTIQUES

Facultat de Matemàtiques
Universitat de Barcelona

Anàlisi facial en entorns d'interacció home-màquina

Autor: Aitor Moreso Castellví

Director: Dr. Sergio Escalera,
Sr. Ciprian Corneanu, Sr. Marc Oliu
Realitzat a: Departament de Matemàtica Aplicada
i Anàlisi

Barcelona, 18 de gener de 2016

Abstract

This project analyzes the most recent state of the art methods for face fitting in images as a first step of the methodologies required to be applied in human-machine interaction environments. The project centers its evaluation on the Explicit Shape Regression (ESR) and the Robust Cascade Pose Regression (RCPR) methods, which stem from a cascade regressors methodology with the objective of finding a set of points of interests of the faces in the images. This methodology of cascade of regressors is an initial requirement in most human-machine interaction scenarios; gender, age, ethnicity, face and emotion recognition, just to mention a few.

To perform this analysis the public data base BOSHPORUS BD has been modified. The facial images of citizen with European ethnicity have been rotated and projected from different angular perspectives to carry out training of the previously appointed methods under different simulated conditions.

The recent development of new methods for the detection of faces in images under different conditions requires researchers to analyze and compare these tools with the objective of determining which method offers the highest performance under different conditions of use as well as computational restrictions. As the result of this project two of the most frequently used methods of the state of the art have been evaluated and compared on the generated data to reach conclusions on the pros that make them the most suitable for different applications and the cons that open the door to future lines of research

Keywords: facial analysis, cascade regresors, human-machine interaction, BOSHPORUS BD, Explicit Shape Regression (ESR), Robust Cascaded Pose Regression (RCPR).

Resum

En aquest treball s'analitzen els mètodes més actuals de l'estat de l'art per a l'anàlisi automàtic de cares en imatges, amb l'objectiu de ser aplicat en entorns d'interacció home-màquina. El projecte se centra en l'evaluació dels mètodes *Explicit Shape Regression* (ESR) i *Robust Cascaded Pose Regression* (RCPR), els quals parteixen d'una metodologia de regressors en cascada amb l'objectiu de trobar un conjunt de punts d'interès de les cares presents a les imatges. Aquesta metodologia de regressors de cascada són una etapa inicial de requeriment en la majoria d'aplicacions d'interacció home-màquina basats en anàlisi facial: reconeixement de gènere, edat, ètnia, identificadors d'individus, o emocions i expressions facials, entre d'altres.

Per tal de realitzar aquest anàlisi, s'ha modificat la base de dades pública BOSPHORUS BD, i s'han trencat i projectat imatges facials de persones d'ètnia europea des de diferents perspectives angulars, per així realitzar un entrenament baix diferents condicions simulades d'ús dels mètodes anteriorment esmentats.

Donat que diferents mètodes han sorgit recentment per la detecció de cares en imatges baix diferents condicions, aquest fet obliga els investigadors a analitzar i comparar aquestes eines amb l'objectiu de determinar quin mètode ofereix un millor rendiment baix diferents condicions d'ús, així com les seves restriccions computacionals. Com a resultat d'aquest treball, dos dels mètodes més usats en l'estat de l'art han estat avaluats i comparats sobre les dades generades, extraent diferents conclusions sobre els pros que els fan més adequats per a diferents aplicacions, i els contres que a la vegada obren les portes a futures línies d'investigació.

Paraules clau: anàlisi facial, regressors en cascada, interacció home-màquina, BOSHPORUS BD, *Explicit Shape Regression* (ESR), *Robust Cascaded Pose Regression* (RCPR).

Resumen

En este trabajo se analizan los métodos más actuales del estado del arte para el análisis automático de caras en imágenes, con el objetivo de ser aplicados en entornos de interacción hombre-máquina. El proyecto se centra en la evaluación de los métodos *Explicit Shape Regression (ESR)* y *Robust Cascaded Pose Regression* (RCPR), los cuales parten de una metodología de cascadas regresores con el objetivo de encontrar un conjunto de puntos de interés de las caras presentes en la imagen. Esta metodología representa una etapa inicial requerida en la mayoría de aplicaciones de interacción hombre-máquina basadas en análisis facial: reconocimiento de género, edad, etnia, identificación de individuos, o emociones y expresiones faciales, entre otros.

Para llevar a cabo este análisis, se ha modificado la base de datos pública BOSPHORUS BD, y se han rotado y proyectado imágenes faciales de personas de etnia europea desde diferentes perspectivas angulares, para así realizar un entrenamiento bajo diferentes condiciones simuladas de uso de los métodos anteriormente nombrados.

Dado que diferentes métodos han surgido recientemente para la detección de caras en imágenes bajo diferentes condiciones, este hecho obliga a los investigadores a analizar y a comparar estas herramientas con el objetivo de determinar qué método ofrece mejor rendimiento bajo diferentes condiciones de uso, así como sus restricciones computacionales. Como resultado de este trabajo, dos de los métodos más usados del estado del arte han sido evaluados y comparados sobre los datos generados, habiendo extraído diferentes conclusiones sobre los pros que los hacen más adecuados para diferentes aplicaciones, y los cons que a la vez abren la puerta a futuras líneas de investigación.

Palabras clave: análisis facial, regressores en cascada, interacción hombre-máquina, BOSHPORUS BD, *Explicit Shape Regression* (ESR), *Robust Cascaded Pose Regression* (RCPR).

Agraïments

El meu sincer agraïment a totes aquelles persones que, d'una manera o d'una altra, m'han ajudat i m'han proporcionat tot el seu suport en l'elaboració d'aquest Treball final de grau.

En primer lloc, cal agrair especialment al meu tutor Sergio Escalera i als co-directors Ciprian Corneanu i Marc Oliu per la seva dedicació, atenció i guiatge en aquesta darrera etapa del grau.

I, en segon i darrer lloc, m'agradaria agrair a la meva família per haver-me donat l'oportunitat de poder realitzar els meus estudis i motivar-me dia rere dia per tal d'assolir els meus objectius i el meu somni.

Índex

1 Introducció

L'anàlisi de cares, incloent la detecció i el reconeixement facial, així com l'alineament de la seva geometria, és un concepte relativament nou ja que es porta estudiant només des de fa 40 anys. A causa de la seva novetat, ens trobem amb molt recorregut per investigar sobre aquest camp. Aprofitant aquest fet s'ha volgut fer un estudi de la detecció de la geometria de la cara amb l'ajuda de dos codis de detecció de punts de referència. En primer lloc, tenim una base de dades amb milers d'imatges; aquestes es dividiran en dos grups. El primer grup formarà la part d'entrenament del programa per tal de tenir unes referències fixes; mentre que la segona part seran les imatges analitzades. Un cop es tinguin aquestes dades es portarà a terme un estudi dels resultats obtinguts.

1.1 Objectiu

Ànalisi facial en entorns d'interacció home-màquina és un treball que analitza diferents algorismes de detecció de la *geometria facial*[1] de qualsevol persona. S'han estudiat diversos algorismes, però aquest treball utilitza una família de mètodes anomenada regressió en cascada.

L'objectiu principal d'aquesta família de mètodes és trobar l'alineació de la cara de qualsevol persona detectant prèviament els *landmarks*[2] i ajustant-la amb els regressors. Aquests estudien la detecció dels diversos punts característics de la cara.

En aquest treball, s'analitzaran i es mostraran resultats per a dos mètodes de regressió en cascada. Tots dos consisteixen a aprendre un model definit per un conjunt de regressors, a partir de les dades d'entrenament de les quals es coneix la geometria. Els regressors preveuen una actualització d'una geometria inicial estimada que s'ha basat en informació visual extreta de la imatge. Aquest model servirà posteriorment per poder aplicar-ho a unes altres dades d'experimentació i, així, poder preveure la posició dels punts de referència en imatges noves.

[1] geometria facial: la geometria de l'objecte es defineix com un conjunt $L = L_i \ | L_i =< x_i, y_i, [z_i] >$ on cada marca correspon a una ubicació d'un punt. El problema de la localització de cada punt de referència es coneix com la geometria d'alineació.

[2] landmarks: punts de referència de la cara.

1.2 Diagrama de Gantt

A la Figura 1 es mostra el diagrama de Gantt on s'explica l'execució de les diferents parts del projecte respecte el temps de duració de cada secció.

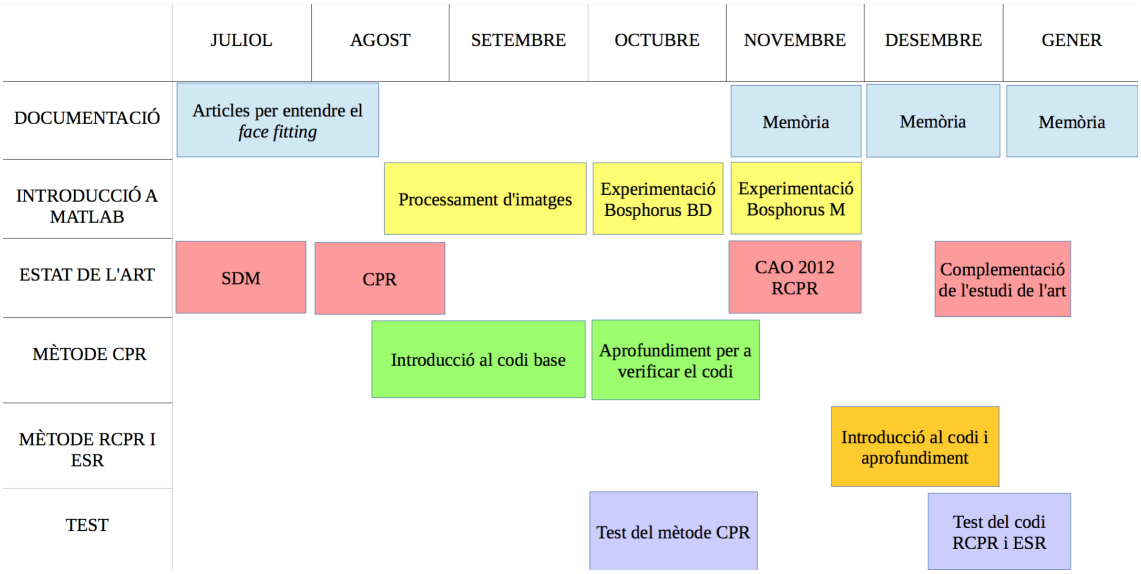


Figura 1: Diagrama de Gantt

1.3 Estructura

El treball es troba organitzat en tres apartats principals. Primerament, (secció 2) és de caràcter teòric i s'expliquen els conceptes bàsics que s'utilitzaran al llarg del treball i tenen a veure amb els mètodes utilitzats per realitzar els tests. D'altra banda, una segona part (secció 3, 4) on hi ha la part pràctica que és on es realitza l'experimentació i l'anàlisi dels resultats obtinguts. I, per finalitzar, hi ha una part on s'extreuen unes conclusions (secció 5) sobre l'elaboració del treball, les impressions personals i s'exposen algunes propostes de millora en aquest camp per a un futur proper.

2 Marc teòric

El procés posterior a la detecció de la cara [5] és la localització dels punts de referència. Aquest pas és necessari per a alguns algorismes de diferents tipus: reconeixement d'expressions facials, reconeixement de cares, generació de models 3D, augment de les dades mitjançant rotacions, etc.

Per molts mètodes, el primer pas en el procés de detecció de punts de referència és construir un model de la cara. Un conjunt de punts de referència es marquen a la cara (Figura 10). A partir d'aquest punts es busquen les direccions principals de variació sobre un conjunt de cares amb diverses persones, expressions facials, rotacions, la detecció de la mirada, interpretació del llenguatge de gestos, estimació de l'edat i lectura dels llavis, etc. Segons el model utilizat, hi ha diversos punts de referència. Hi ha models de 17 punts, de 29 o fins i tot de 68 punts. En aquest treball els models empren 22 punts. Alguns mètodes no requereixen d'un model de la cara (secció 2.1.2). Aquests regressen directament les coordenades de cada *landmark* partint de la forma mitja.

El mètode de *Facial feature point detection (FFPD)* [4] consta de dues fases. La primera és l'entrenament, la qual consisteix a aprendre les variacions de l'aparença i de la forma que posteriorment s'apliquen en una segona fase, la fase de prova. Amb un gran volum de dades, s'aprenen les variacions i s'inicialitzen pas a pas fins arribar a la convergència desitjada.

Dintre del FFPD es distingeixen quatre categories:

1) Constrained Local Model (CLM): és un mètode que es basa en un model de la fisonomia de la cara i una sèrie d'estudis locals, que s'utilizen per detectar un punt facial. Aquests mètodes consideren la variació de l'aparença de cada punt de referència. Es calcula un mapa de probabilitat al voltant de l'estimació actual de cada punt de referència.

Un exemple d'aquesta categoria és *Non-linear point distribution modeling using a multi-layer perceptron* [14]. Aquest mètode va canviar el PDM, un mètode lineal, a un mètode no lineal a partir de regressions polinomials i perceptron multicapa (Figura 2).

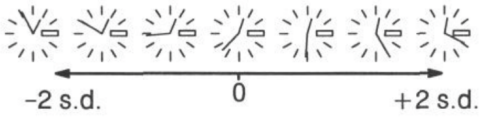


Figura 2: Aplicació del model *Non-linear point distribution modeling using a multi-layer perceptron* a rellotges.

2)Active Appearance Model (AAM): aquest és un dels mètodes més utilitzats en *Face Fitting*. Un mètode ampliat del *Active Shape Model* (ASM) [11] que

codifica la intensitat de les imatges i la textura de la cara. Aquest model consisteix a modelar la variació de l'aparença des d'una perspectiva holística reduint al mínim els errors en la textura. La variació de l'aparença i la forma s'aprèn mitjançant una combinació lineal de les formes que s'han provat en l'entrenament.

Un exemple d'aquesta categoria és *Real-time facial feature tracking on a mobile device* [26]. Aquest mètode va explorar característiques similars Haar per proporcionar una projecció lineal amb poc cost computacional per facilitar el rastreig dels punts de referència amb un aparell mòvil.

Un altre exemple és *Active appearance models revisited* [15]. Va proposar una millora de l'eficiència del procés de fitting utilitzant el model AAM com un problema d'alineament d'imatge i el va optimitzar a través d'un mètode invers composicional (Figura 3).

Figura 3: Exemple d'una aplicació del model *Active appearance models revisited.*

3)Mètodes basats en la regressió: aquests mètodes estimen la forma directament desde l'aparença, sense utilitzar cap model de forma ni cap model d'aparença. Aquests mètodes utilitzen una funció de regressió de l'aparença per ajustar la imatge.

Un exemple d'aquesta categoria és *Shape regression machine* [16]. Van proposar un mètode de regressió de formes basat en Freud and Schapire, el qual es divideix en dues parts. En la primera part, els paràmetres rígids es troben plantejant el problema com si fos de detecció d'objecte i aquest es soluciona amb un mètode basat en la impulsió. En la segona part, es troba una funció de regressió regularitzada a partir d'exemples d'entrenament per predir la forma no rígida (Figura 4).

Un altre exemple és *Fully automatic feature localization for medical images using a global vector concentration approach* [17]. Van proposar un vector per localitzar les característiques facials sense cap suposició específica a priori sobre la forma facial o la configuració dels punts de referència.

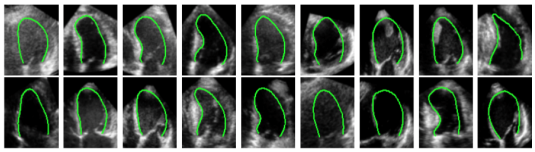


Figura 4: Aplicació del model *Shape regression machine* en ecografies.

4)Altres mètodes: aquests es divideixen en quatre subcategories més:

4.1)Mètodes gràfics: aquests estan basats en l'estructura d'un graf i el camp aleatori de Markov. Aquests mètodes seleccionen cada punt de referència com un node i tots els altres punts com a vèrtex del graf. Les ubicacions dels *landmarks* poden ser resoltes mitjançant programació dinàmica.

Un exemple d'aquesta categoria és *Finding deformable shapes using loopy belief propagation* [18]. Van desenvolupar un model generador de gràfics Bayesians que utilitzava models separats per descriure la variabilitat de la forma i l'aparença per trobar formes deformables.

Un altre exemple és *Accurate face alignment using shape constrained Markov network* [19]. Va proposar un mètode que incorporava una forma global a priori directament a la xarxa de Markov (Figura 5).

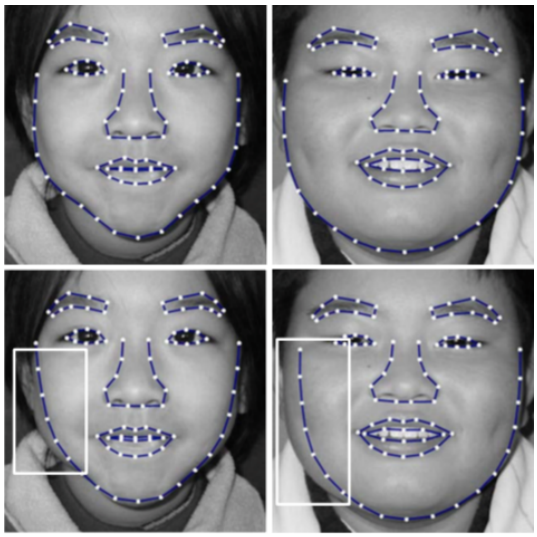


Figura 5: Aplicació del model *Accurate face alignment using shape constrained Markov network* on les imatges de sobre és l'estimació real i les de sota l'estimació del model.

4.2)Mètodes d'alineació del conjunt cara: aquest mètode utilitza un conjunt d'imatges que han estat sotmeses a una gran varietat de transformacions geomètriques de manera que per trobar els punts de referència només calgui resoldre un sistema no lineal. A més a més, proposa que el mètode per resoldre aquest pro-

blema sigui el dels multiplicadors de Lagrange que va donar resultats molt eficients en imatges degradades i amb oclusions.

Un exemple d'aquesta categoria és *Joint face alignment with a generic deformable face model* [20]. Van disenyar un AAM combinat a partir del mètode d'alineament *Congealing style joint alignment method* i *low-rank descomposition method*. Van assumir que les imatges de la mateixa cara s'havien de trobar en el mateix subespai lineal i l'espai específic a la persona havia de ser pròxim a l'aparença genèrica de l'espai facial genèric.

Un altre exemple és *Face detection, pose estimation, and landmark localization in the wild* [21]. A diferència del mètode anterior, proposaven un model d'enfocament en dos fases per alinear un conjunt d'imatges de la mateixa persona. En la primera fase, es fa una estimació inicial dels punts de referència amb un enfocament fora de la plataforma. En l'altra fase, es distingeixen alineaments bons dels dolents, on els bons s'utilitzen per millorar els dolents (Figura 6).

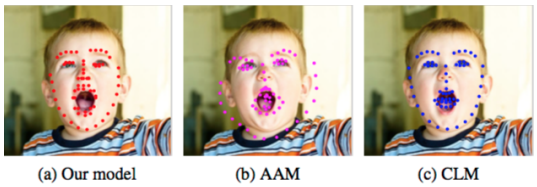


Figura 6: Comparació del model *Face detection, pose estimation, and landmark localization in the wild* amb AAM i CLM.

4.3)Detectors de punts facials independents: tots els mètodes presentats anteriorment preveuen les ubicacions de tots els punts de referència de forma simultània, mentre que aquests mètodes no es basen en les imatges etiquetades manualment. Utilitzen un classificador per als punts de referència que l'elegeixen en funció de la probabilitat del millor resultat. Un gran avantatge d'aquests mètodes és la lliure inicialització. En canvi, un problema important és l'ambigüitat a causa del gran nombre possible de resultats diferents.

Un exemple d'aquesta categoria és *Fully automatic facial feature point detection using Gabor feature based boosted classifiers* [22]. Van detectar cada punt per una tècnica focal com en els mètodes basats en CLM. Utilitza *Gavor feature based Boosted classifier* per classificar el grup d'imatges positives de les negatives. El mapa amb més respostes es marca com la resposta buscada (Figura 7).

Un altre exemple és *Cascaded shape space pruning for robust facial landmark detection* [23]. Van proposar estimar en conjunt les posicions correctes de tots els punts de referència d'alguns candidats obtinguts a partir de detectors de punts de referència independents.

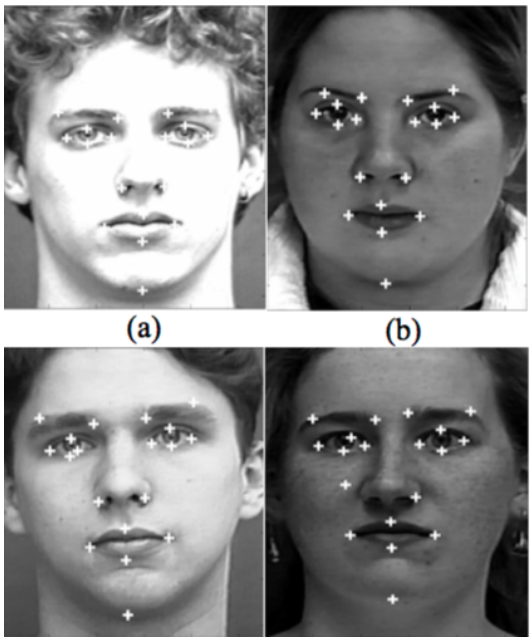


Figura 7: Apliacació del model *Fully automatic facial feature point detection using Gabor feature based boosted classifiers.*

4.4)Mètodes basats en l'aprenentatge profund: aquests mètodes operen de manera que aprenen conjuntament les característiques rellevants per resoldre aquests problemes, i, a la vegada, aprenen com s'han d'aplicar per resoldre'ls. Es pot formular com un problema de regressió o classificació amb un alt nivell de no linealitat, que agafa directament la imatge d'entrada i treu el valor de resposta.

Un exemple d'aquesta categoria és *Deep convolutional network cascade for facial point detection* [24]. Van proposar una xarxa convolucional de tres nivells de cascada per a la detecció de punts d'una manera gruix-fi, es a dir, cada vegada és té un pas amb millor precisió, que està compost per una sèrie de xarxes convolucionals. Millora els punts de referència, però amb molt de temps computacional.

Un altre exemple és *Facial feature tracking under varying facial expressions and face poses based on restricted Boltzmann machine* [25]. Van explorar una xarxa de creences profundes per capturar la forma de la cara a causa de canvis en les expressions facials i van utilitzar una màquina de Boltzamann restringida de tres camins per capturar les relacions entre les formes facials frontals i no frontals (Figura 8).

Figura 8: Aplicació del model *Facial feature tracking under varying facial expressions and face poses based on restricted Boltzmann machine.*

2.1 Estat de l'Art

L'estat de l'art [13] és un concepte anglès derivat de l'expressió *state of the art* i s'utilitza per designar la tecnologia punta, és a dir, les últimes i avançades tecnologies; i també es refereix al límit de la investigació científica.

Els mètodes utilitzats en aquest treball estan inclosos en l'estat de l'art ja que són els més actuals i avançats en el camp d'anàlisi de detecció de cares. A continuació, s'exposen els mètodes emprats en aquest projecte. Un cop s'hagin vist tots els mètodes s'escullirà els dos mètodes més adients als objectius planteijats en la part inicial del projecte.

2.1.1 Active Shape Model (ASM)

Active Shape Model (ASM) [11] va ser creat per Tim Cootes i Chris Taylor l'any 1995. Són models paramètrics deformables que, analitzant un conjunt de variacions d'una forma geomètrica deformable, construeixen un model estadístic de la variació global de la forma de l'objecte. Amb l'ajuda del model estadístic es pot ajustar a imatges no incloses a l'entrenament mitjançant descens de gradient.

Construcció del mètode

Amb l'ajuda de *Principal Component Analysis* (PCA), que té la finalitat de reduir la dimensionalitat de la imatge i generar formes similiars a la del conjunt d'entrenament, es construeix el mètode.

Per començar, es crea un polígon de n vèrtex per a cada geometria en el conjunt de dades:

$$X = (x_1, y_1, ..., x_n, y_n)^T$$

que posteriorment es normalitza respecte els paràmetres de posició t_x ,t_y,s , θ on

t_x i t_y és la translació, s l'escala i θ la rotació.

Es calcula la forma mitja ($\bar{x}$) i la variació de cada forma respecte la mitja (dx_i):

$$\bar{x} = \frac{1}{m}\sum_{i=1}^{m} x_i$$
$$dx_i = x_i - \bar{x}$$

Gràcies a la desviació típica, es pot construir la matriu de covariància que posteriorment s'utilitza per extreure els valors propis de la matriu:

$$\sum = \frac{1}{m}\sum_{i=1}^{m} dx_i dx_i^T$$

D'aquesta fórmula s'obté la relació següent:

$$\sum p_i = \lambda_i p_i$$

on λ_i es l'i-èssim valor propi. Es construeix P com la matriu de bases ortogonals ordenada en funció dels valors propis corresponents:

$$P = [p_i ... p_{2n}]$$

Gràcies a P es pot obtenir deformacions de la forma mitja:

$$\bar{x} = x + Pb$$

2.1.2 Supervised Descent Method (SDM)

La geometria de l'objecte es defineix com un conjunt $L = L_i | L_i =< x_i, y_i, [z_i] >$ on cada marca correspon a una ubicació d'un punt. El problema de la localització de cada punt de referència es coneix com la geometria d'alineació.

Supervised Descent Method (SDM) [7] [8] és un algorisme que implementa una cascada de regressors lineals estimant la direcció de descens de cada punt de referència. S'inicia amb la forma mitja de la cara i, per a cada pas, mitjançant un

descriptor SIFT, un algorisme que descriu i detecta les característiques locals de la imatge, els s'extreu una descripció de l'aparença de cada punt estimat, i, a través del PCA es, redueix la dimensió.

L'objectiu del SDM és aprendre una sèrie de direccions de descens de manera que es produeix una seqüència $x_{k+1} = x_k + \delta x_k$ a partir de x_0 fins que s'aproxima iterativament a $x*$, els punts de referència correctes.

Construcció del mètode

Donada una imatge $d \in R^{m+1}$ i els punts de referència $d(x) \in R^{p+1}$ per poder fer l'entrenament, es tracta de minimitzar h que és una funció d'extracció de característiques SIFT.

En l'entrenament es coneixen els punts de referència correctes $x*$, però per inicialitzar el mètode es forma una configuració inicial x_0, que és la forma mitja.

L'objectiu principal del mètode és minimitzar la següent funció, de la cual coneixem δx i ϕ_*:

$$f(x_0 + \delta x) = ||h(d(x_0 + \delta x)) - \phi_*||_2^2$$

on ϕ_* són els valors SIFT dels punts de referència que s'han marcat manualment.

SDM aprèn una seqüència de direccions de descens R_k i uns termes de polarització b_k. Per a cada imatge, a partir d'una estimació inicial x_0^i aleshores R_0, b_0 s'obté minimitzant la pèrdua esperada entre el que es preveu i el desplaçament òptim.

$$argmin_{R_0,b_0} \sum \int p(x_0^i)||\delta x^i - R_0\phi_0^i - b_0||^2 dx_0^i$$

on $\delta x^i = x_*^i - x_0^i$ i $\phi_0^i = h(d^i(x_0^i))$.

R_k i b_k s'aprenen recursivament utilizant la fórmula anterior. Després, s'obté un nou conjunt d'elements mitjançant l'actualització $\delta x_*^{ki} = x_*^i - x_k^1$ i $\phi_k^i = h(d^i(x_k^i))$ i aplicant novament la minimització de l'equació:

$$argmin_{R_k,b_k} \sum_{d^i} \sum_{x_k^{ki}} p(x_k^i)||\delta x^i - R_k\phi_k^i - b_k||^2$$

Es van obtenint R_k i b_k.

2.1.3 Cascaded Pose Regression (CPR)

Cascaded Pose Regression (CPR) [6] és un altre algorisme d'alineació facial 2D ràpid i precís. Es basa en una cascada de regressors de tipus *fern*, que és una composició de F característiques i límits que divideixen l'espai de característiques en contenidors de 2^F. Les característiques es mesuren com la diferència de píxels, és a dir, la diferència d'intensitat entre dos píxels. Aquestes diferències són molt bones i amb un cost computacional gens car.

Per poder construir un *fern* s'utilitzen 4 passos:

1) En primer lloc, s'obté l'increment de S, és a dir, la diferència entre la forma estimada i la forma real.

2)Entre P^2 característiques es selecciona la característica amb correlació més alta respecte l'increment.

3) Es repeteixen els passos 1 i 2 F vegades per tenir les F característiques corresponents.

4) Es construeix el *fern* per trets F amb límits aleatoris.

Aquest mètode utilitza regressió en cascada, al igual que SDM, però a diferència l'ús de regressors *fern* permet una regressió per parts, no lineal.

S'inicialitza amb una posició inicial i gràcies als regressors, diferents tots ells en cada pas, s'aproxima iterativament la geometria real de la cara. El que intenta el regressor R és minimitzar la distància entre la posició real i la posició estimada; i en l'algorisme cada regressor depèn dels regressors anteriors. En cada pas s'utilitza el *Random Fern Regressors* que genera aleatoriament N *ferns* i, en cada etapa de la cascada agafa el *fern* millor en termes d'error d'entrenament.

Construcció del mètode:

L'objectiu principal de l'algorisme és entrenar un regressor R. Es segueixen els següents passos:

Donada una entrada θ^0, $R(\theta^0, I)$ es calcula:

$$\theta^t = \theta^{t-1} \circ R^t(h^t(\theta^{t-1}, I)), t = 1, ..., T \tag{2.1}$$

Una vegada esta calculat θ^t s'intenta minimitzar l'error:

$$L = \sum_{i=1}^{N} d(\theta_i^t, \theta_i)$$

on $\theta^0 = \arg\ min_\theta \sum_i d(\theta, \theta_i)$, $\theta_i^0 = \theta^0$, $\forall i$. θ^0 és l'estimació de la posició que dóna l'error d'entrenament més baix sense dependre de cap regressor.

Els regressors R s'aprenen de la següent manera:

En cada pas t es generen aleatoriament les característiques h^t i $x_i = h^t(\theta^{t-1}, I)$ para cada entrenament I_i on $\theta_i^t = \theta_i^{t-1} \circ R^t(x_i)$:

$$R^t = argmin_R \sum_i d(R(x_i), \omega_i)$$

$$\omega_i = \theta_i^{t-1} \circ \theta_i$$

Després de cada entrenament s'aplica (2.1) per calcular θ_i^t.

Si l'error no es pot reduïr aleshores es calcula l'error com:

$$\epsilon_t = \frac{\sum_i d(\theta_i^t, \theta_i)}{\sum_i d(\theta_i^{t-1}, \theta_i)}$$

L'error ens dona la ratio entre l'error actualitzat i l'error en el pas anterior. Si l'error és més gran que 1, llavors ha deixat de decrementar i, per tant, s'acaba l'entrenament. Per tal de millorar el rendiment de l'algorisme s'utilitza *Pose Clustering*. Aquest algorisme consisteix a inicialitzar les imatfes K vegades diferents i agafar la posició on s'obté la regió de major densitat respecte la posició-espai.

2.1.4 Explicit Shape Regression (ESR)

Explicit Shape Regression (ESR) és un mètode que estima una forma S el més propera possible a la forma real S*. La forma estimada S és una combinació lineal de totes les formes obtingudes en l'entrenament.

La majoria dels enfocaments basats en l'alineació es classifiquen de dues maneres diferents: basats en optimització i basats en regressors.

El mètode ESR [2] és un mètode basat en regressors, però a diferència dels mètodes anteriors, aprèn una regressió vectorial que, de manera explícita, minimitza els errors d'alineació en l'entrenament. ESR també utilitza els *ferns*, ja comentats més detalladament en la secció 2.4.3.

Construcció del mètode

Donada una imatge i una forma inicial S^0, es calcula el regressor com l'increment δS, és a dir, la diferència entre la forma real i la forma estimada. La forma estimada s'actualitza amb la fórmula:

$$S^t = S^{t-1} + R^t(I, S^{t-1})$$

on S^t és el regressor R^t s'actualitza com:

$$R^t = \arg\min \sum_{i=1}^{N} ||S^* - (S_i^{t-1} + R(I_i, S_i^{t-1}))||$$

Les bases de dades que es van utilitzar per experimentar amb aquest mètode són: BioId, *Labeled Face Parts in the Wild* (LFPW) i *Labeled Faces in the Wild* (LFW).

2.1.5 Robust Cascaded Pose Regression (RCPR)

El mètode *Robust Cascaded Pose Regression* (RCPR) [1] és un mètode creat a partir de l'anterior mètode CPR. És una versió més sofisticada i robusta del CPR. La finalitat principal del mètode és minimitzar els errors, tot i afegint més dificultat a les imatges, així com oclusió, degradació, etc.

RCPR utilitza una base de dades anomenada COFW, la qual és més propera a la realitat amb un promig del 23% d'oclusió i amb una gran varietat de la forma. Uns clars exemples d'oclusions podrien ser les ulleres de sol o una part de l'extremitat superior ocultant alguna part de la cara.

RCPR utilitza els regressors i les estimacions igual que el CPR, però afegint un altre paràmetre que és l'oclusió dels punts de referència. Un punt de referència tindrà un 0 si està ocult i un 1 si està visible. Aquest vector de 0,1 donarà molta informació a l'entrenament. Primerament, es dividirà la imatge en una matriu 3x3 i en cada iteració tindrem el percentatge en cada quadrícula de l'oclusió dels punts de referència en cada parcel·la. La regressió es farà de tots els paràmetres a la vegada i les característiques s'obtindran del punt de referència més proper i no globalment.

RCPR és un mètode que treballa amb regressors. Els mètodes de regressió actuals són ràpids i troben una petita variació de la forma, tot i que quan es troben davant d'una oclusió o una variació bastant gran, el mètode deixa de ser efectiu.

Construcció del mètode

El CPR consta de T regressors, $R^{1,...,T}$ que s'inicialitzen en una forma S^0 i va canviant i millorant la forma fins arribar a l'estimació adequada S^t. Aquest canvi es produeix gràcies al regressor i l'estimació S^{t-1}. Els regressors R^t minimitzen l'estimació amb la forma real.

Gràcies a la interpolació lineal entre dos punts de referència, el rendiment del programa serà més ràpid i efectiu.

Donada una imatge i un nombre d'inicialitzacions proposa agafar un 10% de la cascada que s'aplica a cadascun. Seguidament, comprova la variació entre les prediccions. Si la variació està per sota d'una constant τ, el 90% restant s'aplica com en l'enfocament clàssic. En cas contrari, el procés torna al seu punt inicial i s'inicia amb un conjunt diferent. Després de fer les comprovacions necessàries, s'estima que el valor de la constant τ és de 0.15 i que es necessiten una mitja de 3 iteracions per equilibrar el rendiment amb la velocitat.

El RCPR s'ha experimentat amb tres bases de dades públiques: LFPW, LFW i Helen. Aquestes bases de dades s'han utilitzat per contribuir en l'entrenament d'altres mètodes.

Un cop estudiats els diferents mètodes, s'ha decidit que per a la part pràctica d'aquest projecte s'utilitzaran RCPR i ESR, perquè són els dos mètodes que compleixen els requisits que s'han demanat en la part d'objectius d'aquest projecte. A més aquest dos mètodes han obtingut els millors resultats en el camp dels models de regressió gràcies a un ajustament molt precís.

3 Marc pràctic

Per iniciar el treball, el primer pas a realitzar va ser trobar una base de dades d'imatges per tal de poder experimentar amb els mètodes. En aquest cas, es va seleccionar una base de dades pública: BOSPHORUS BD.

Aquesta base de dades consta d'un conjunt d'imatges amb 2D i 3D. Només s'utilitzen els *landmarks* i les imatges amb 2D. Les imatges estan amb arxius *.pnb i els *landmarks* amb arxius *.lm2

Cadascuna de les imatges té associat els seus *landmarks* corresponents, que varien entre 16 i 24 *landmarks*, segons la imatge. Com que cada imatge té un punt de vista diferent, a causa de l'auto-oclusió, no totes les imatges tenen el mateix nombre de punts de referència.

Les imatges poden incloure diferents aspectes com podria ser la perspectiva des d'on es pren la imatge, la inclinació de la cara o les expressions de la cara com podrien ser els estats d'ànim d'una persona.

Gràcies al codi prestat per la base de dades BOSHPORUS BD, es va poder dur a terme aquest treball.

Per començar, s'havia d'examinar bé el codi i observar quins eren els passos a seguir. BOSHPORUS BD no donava els *landmarks* d'una forma unificada. Per tant, el primer pas va ser compactar totes les dades creant un tipus d'estructura on a cada imatge se li associa els seus punts de referència corresponents. Per tal que les imatges fossin compactes, es van haver d'ajustar totes les imatges a una mida estàndard (300 x 300 píxels). Tanmateix, els punts de referència estaven en un format *.lm2 i, gràcies a una funció proporcionada per la base de dades, es podien obtenir els *landmarks* com a un vector de 2 x n. D'aquesta manera s'aconseguia la posició de tots els punts de referència.

La base de dades mostrava els diferents punts de referència que tenia cada imatge, però no n'hi havia prou amb aquesta informació ja que el codi no feia servir aquesta nomenclatura per poder entrenar les imatges.

El codi proporcionat treballa amb un vector de característiques per cada imatge. L'objectiu del codi és entrenar un conjunt d'imatges i extreure'n les característiques necessàries per poder construir un regressor. D'aquesta manera, aplicar-ho a un altre conjunt de dades d'on únicament coneixem la informació de la imatge. L'algorisme només treballa amb un triangle format entre els centres dels dos ulls i el centre de la boca (Figura 9). En conclusió, el programa té com a fita alinear aquest triangle de la millor manera possible.

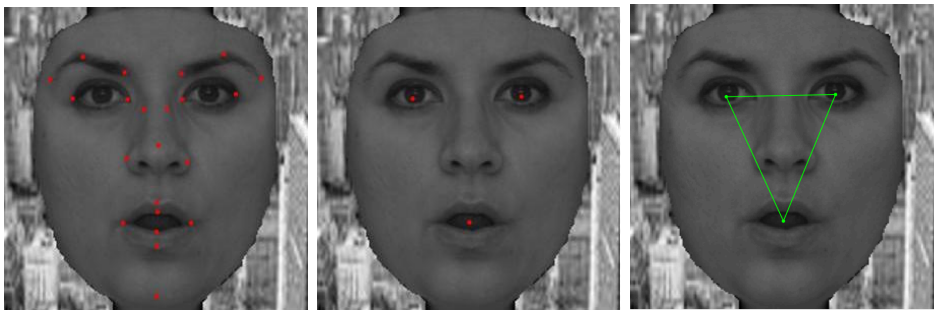


Figura 9: Evolució de tots els passos per arribar a l'alineació dels ulls i la boca.

Com que la base de dades proporciona entre 16 i 24 *landmarks*, per obtenir el centre dels ulls i de la boca es realitza una mitja aritmètica de tots els punts de referència que hi ha al voltant de cada ull i de la boca, per obtenir una aproximació del centre d'aquests.

Després d'experimentar amb les dades, es va poder esbrinar com s'obtenia aquest vector de característiques. El vector de característiques està format per cinc components:

1) La component x del punt mitjà entre els centres dels dos ulls.

2) La component y del punt mitjà entre els centres dels dos ulls.

3) L'angle que forma la perpendicular de la recta entre els dos ulls i la recta que uneix el vector mitjà i el centre de la boca.

4) L'amplitud del triangle, és a dir, l'escala.

5) El que anomenem l'aspecte del triangle, és a dir, la relació entre l'altura del triangle i la seva base.

Amb aquestes components es treballa tant la part de l'entrenament com la part del test.

D'aquesta manera, es va passar d'uns punts de referència a un vector de característiques amb els 5 components, tenint en compte que necessitàvem també la funció inversa per després poder plotejar els punts a sobre de la imatge.

El primer experiment va ser restringir el tipus d'imatge de la base de dades a causa de les extenses variacions de les imatges. Es va decidir que només es treballaria amb imatges sense cap tipus de rotació tenint en compte qualsevol tipus d'expressió facial. Per tant, només es va experimentar en les imatges amb més de 21 *landmarks*.

Es va treballar concretament amb un total de 2992 imatges. Però, per poder provar aquestes imatges, s'havia de separar les dades en dos parts: una part seria la

de l'entrenament i, l'altra, la del test. Però, la divisió no es va realitzar equitativament. Com més imatges teníem a l'abast per poder entrenar, millor era el resultat sobre la partició de test com a conseqüència d'una reducció del sobre-ajustament del model.

Per realitzar els experiments utilitzant la màxima quantitat de dades possible per l'entrenament, sense comprometre el nombre de dades disponibles durant el test, es va utilitzar una tècnica de particionament de les dades coneguda com a validació creuada [9]. Aquesta tècnica consisteix a avaluar els resultats d'un anàlisi estadístic i així garantitzar que els resultats són independents de les particions entre l'entrenament i el test. En aquest cas, es van realitzar cinc particions de les dades.

Posteriorment, es va procedir a la utilització de dos mètodes diferents a l'anterior. Com s'ha comentat anteriorment, el primer mètode procedia a fer una alineació de la cara basant-se amb l'alineació dels ulls amb la boca. Aquest dos mètodes posteriors treballaven de forma diferent i utilitzaven tots els punts de referència que es coneixien i, per tant, la detecció de la cara era més robusta ja que no només utilitzaven els punts de referència que envoltaven els ulls i la boca.

Aquest dos mètodes van ser una modificació i una millora del mètode comentat al principi. Es basaven amb el codi proporcionat per la base de dades, però modificat per poder obtenir més informació de la cara i detectar millor tots els punts de referència.

El segon mètode és l'anomenat ESR. Aquest utilitza les mateixes funcions que el CPR, però amb unes lleugeres modificacions. A diferència de l'anterior, aquest algorisme treballa en *folds*[5] o subgrups d'imatges per diferenciar entre la part d'entrenament i la part de test. Al primer mètode es va haver d'incorporar la divisió de *folds* perquè no constava en el codi proporcionat.

El tercer mètode és l'anomenat RCPR. Aquest algorisme és molt semblant al ESR, però amb la diferència que al mètode 2 només es fan 10 entrenaments i en aquest mètode s'utilitzen 100 iteracions i, per tant, l'entrenament té més precisió, però amb un increment de cost computacionalment molt més elevat.

[5]folds: divisió en seccions.

Un cop coneguts els dos mètodes es va procedir a fer les proves amb totes les rotacions. Es van distingir quatre casos possibles:

1) Rotació amb un angle de 45º.

2) Rotació amb un angle de 22.5º.

3) Tot tipus d'imatge frontal.

4) Una barreja dels tres anteriors.

En cap dels dos casos no es va plantejar que les parts de la cara poguessin estar amagades, ocluïdes o tancades. Un punt que seria bastant important d'estudiar posteriorment.

3.1 BOSPHORUS BD

En aquest treball s'ha experimentat amb una base de dades pública coneguda com BOSPHORUS BD [12]. Consta d'un total de 105 persones i 4666 cares, tenint en compte que de cada persona hi ha diferents postures i diverses expressions; 60 persones de les quals són homes i la resta, dones. La majoria de persones són d'ètnia europea.

L'adquisició de totes les imatges es va fer sistemàticament amb tots els individus a una distància de 1,5 metres de la càmera, amb la mateixa intensitat de llum i amb una resolució de 1200 x 1600 píxels.

BOSPHORUS BD és una base de dades d'imatges en 2D i 3D per a la investigació del reconeixement facial entre altres usos. La pàgina web d'aquesta base de dades proporciona les imatges amb el format .png i els seus *landmarks* en 2D i 3D. En aquest cas, només es treballa amb els *landmarks* 2D ja que el mètode es restringeix a les imatges sense profunditat. Per trobar els *landmarks* 2D, la pròpia pàgina web proporciona un algorisme per tal d'obtenir-los en forma de vectors.

Els punts de referència marcats en la imatge estan distribuïts tal com es mostra en la Figura 10.

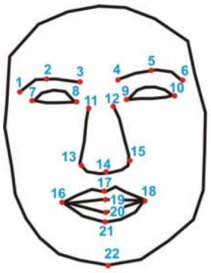


Figura 10: Punts de referència d'una imatge de la base de dades.

Tot i que s'analitzen imatges amb un grau de rotació, els *landmarks* anteriors estan descrits per imatges sense cap tipus de rotació. Si alguna imatge conté alguna rotació, aleshores hi ha menys *landmarks* ja que pot haver parts de la cara que no s'observin, el que es coneix com auto-oclusió.

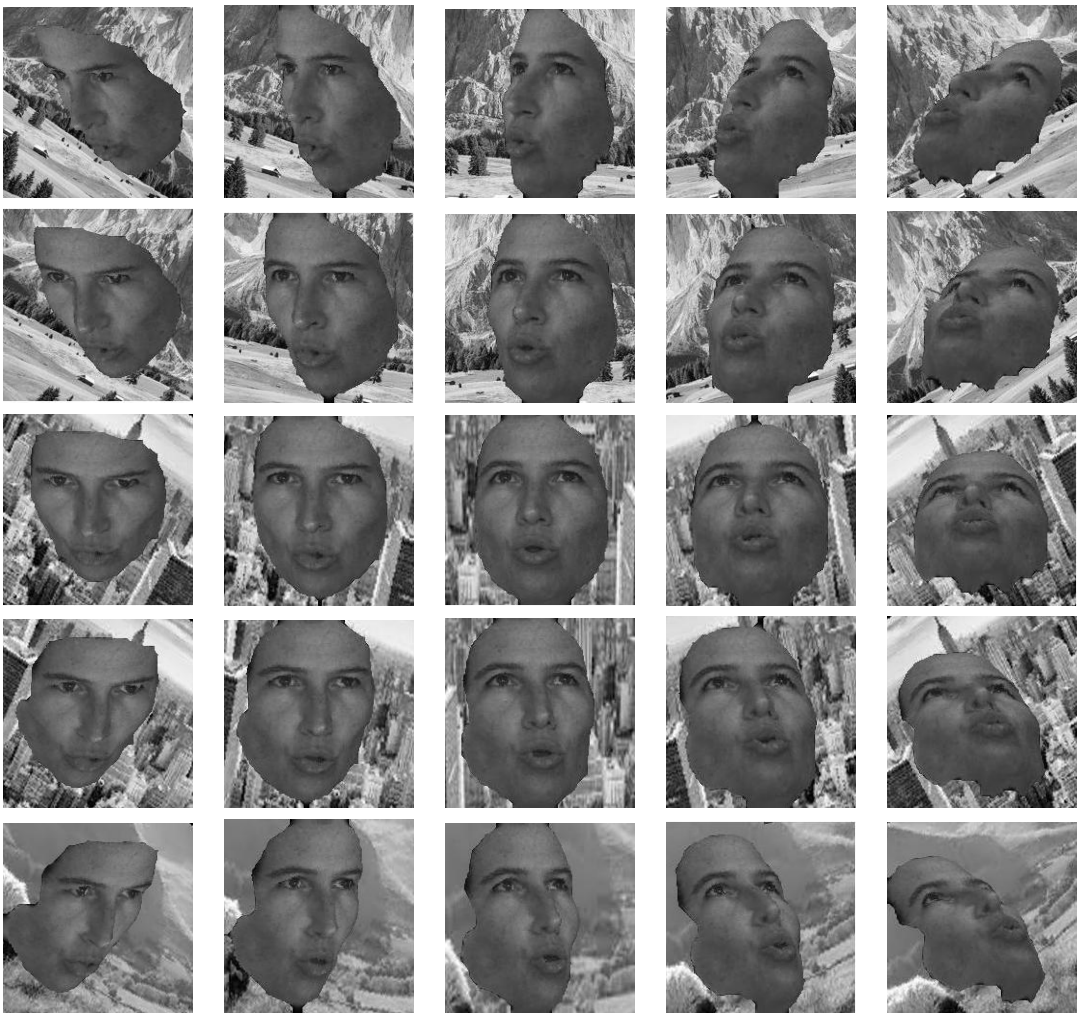
A més, els noms de les imatges descriuen l'expressió facial i tenen etiquetades les diferents unitats d'acció del sistema *Facial Analysis Action System* (FACS), un sistema per denominar els moviments facials humans, present en la imatge. Aquestes unitats d'acció donen una descripció muscular de la cara que permet discriminar la seva expressió facial. També conté descripcions de les emocions a partir de les imatges, com per exemple una cara de felicitat, de tristesa, de sorpresa o d'enuig.

3.2 BOSPHORUS M

BOSPHORUS M és una base de dades creada per poder experimentar en aquest treball. Aquesta és una modificació de totes les imatges de BOSHPORUS BD, la qual mostra una gran varietat d'expressions de la cara, però bastant incompleta en el camp de la rotació. La única modificació que s'ha realitzat ha estat incloure la rotació en totes les imatges.

La base de dades consta de 105 persones i aproximadament 80.000 imatges. Cada imatge està rotada amb angles de 22.5º i 45º de *pitch* i *yaw*.

Les rotacions que es van fer són les mostrades en la Taula 1.



Taula 1: Exemple de totes les combinacions de *pitch* i *yaw* considerades per una imatge de la base de dades BOSPHORUS BD.

3.3 Matlab

Matlab [10] és un *software* matemàtic amb llenguatge propi utilitzat en molts camps de la investigació. El seu nom pertany a un acrònim: Matrix Labortory. El llenguatge utilitzat és M, que fou creat al 1970 per tal de facilitar el *software* de matrius LINPACK i EISPACK sense la necessitat d'utilitzar el llenguatge Fortran. El *software* neix el 1984 gràcies al matemàtic i informàtic Cleve Moler, amb la finalitat de donar suport als cursos d'Àlgebra Lineal i Anàlisis Numèric i d'implementar paquets de subrutines escrites amb Fortran sense la necessitat d'utilitzar aquest llenguatge. Matlab permet utilitzar altres llenguatges de programació com per exemple C o Java. L'entorn permet treballar amb operacions de matrius i vectors; amb programació orientada a objectes; i, fins i tot, amb processament d'imatges en 2D i 3D.

Un inconvenient bastant important és el seu preu, ja que no és un *software* lliure. Matlab actualment pertany a Mathworks. Tot i haver alternatives gratuïtes com ara Octave, i d'altres llenguatges lliures que fan la mateixa funció, com per exemple R, Julia o Python, Matlab ofereix una gran varietat d'eines que el fan més flexible i fàcil d'utilitzar (Figura 11).

En el projecte, el *software* ha estat una peça fonamental per al processament d'imatges i l'entrenament del mètode, ja que amb altres llenguatges, com per exemple C o Python, hagués estat més costós d'implementar.

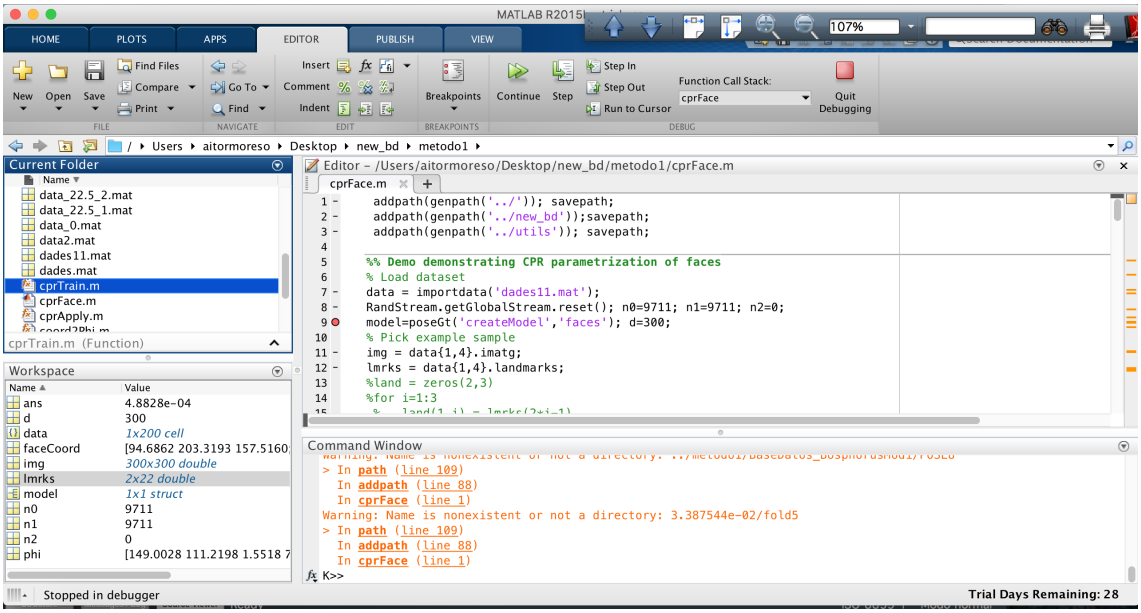


Figura 11: Entorn de Matlab.

4 Resultats

4.1 Anàlisi dels mètodes

En aquest apartat, es mostren els resultats obtinguts (Taula 2, Taula 3) a partir dels dos mètodes emprats en aquest treball ESR i RCPR amb l'objectiu de realitzar una anàlisi quantitativa global i específica de cadascun d'ells.

	Tot tipus	Rotació 22.5º	Rotació 45º	Frontal
Error-Entrenament	24.26%	20.81%	31.22%	9.5%
Temps-Entrenament	5.53 im/s	5.07 im/s	5.47 im/s	9.08 im/s
Error-Test	19.14%	16.23%	23.63%	8.77%
Temps-Test	44.6 im/s	40.90 im/s	44.12 im/s	50.10 im/s
Imatges totals	18225	19396	19396	2396

Taula 2: Taula de l'error mitjà i temps d'execució del mètode RCPR.

	Tot tipus	Rotació 22.5º	Rotació 45º	Frontal
Error-Entrenament	13.75%	12.73%	17.47%	4.2%
Temps-Entrenament	0.87 im/s	0.86 im/s	0.92 im/s	1 im/s
Error-Test	14.66%	13.35%	17.05%	8.7%
Temps-Test	6.17 im/s	5.94 im/s	6.25 im/s	6.05 im/s
Imatges totals	18225	19396	19396	2396

Taula 3: Taula de l'error mitjà i temps d'execució del mètode ESR.

Tot tipus de rotació

En l'estudi del mètode RCPR, es parteix d'una mostra de 18225 imatges on 14581 (4/5) es passen com entrenament i una mostra de 3644 imatges (1/5) com a part del test. Durant l'entrenament es diposa de 14581 imatges. A la taula 2 es pot veure que el temps mitjà d'entrenament és de 2634s i s'aconsegueix un error del 24.26%. En canvi, en el test amb 3644 imatges s'obté un temps mitjà de 81.7s amb un error del 19.14% (Figura 12).

Si s'estudia la variació entre l'entrenament i l'experimentació, s'observa que hi ha una relació 1:8, és a dir, per cada imatge entrenada es diposen 8 imatges al test.

$$v_{ent} = \frac{imatges}{temps} = \frac{14581}{2634} = 5.53\frac{imatges}{s}$$

$$v_{test} = \frac{imatges}{temps} = \frac{3644}{81.7} = 44.60\frac{imatges}{s}$$

$$relació = \frac{v_{test}}{v_{ent}} = \frac{44.60}{5.53} = 8.06$$

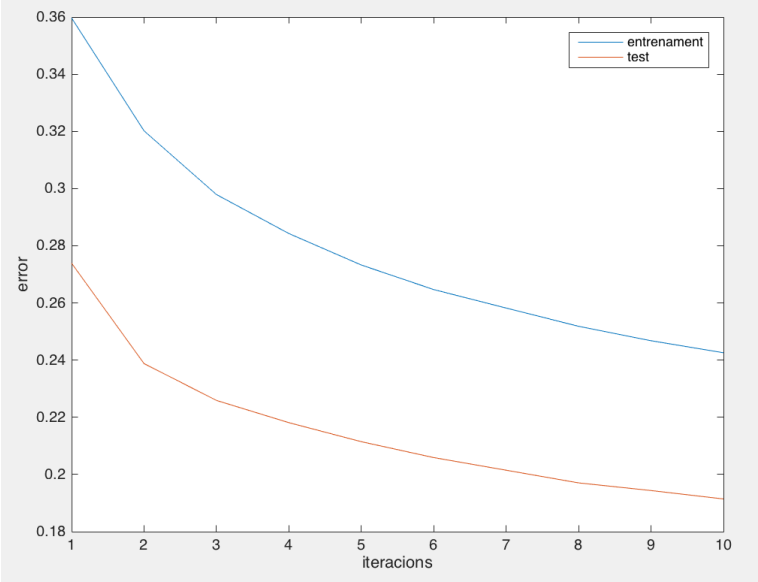


Figura 12: Gràfic d'error d'entrenament i d'error de test en el mètode RCPR.

En el mètode ESR es parteix d'una mostra de 18225 imatges on 14581 (4/5) es passen com entrenament i una mostra de 3644 imatges (1/5) com a part del test. Durant l'entrenament es disposa de 14581 imatges. A la taula 3 es pot veure que el temps mitjà d'entrenament és de 16674s amb un error del 13.75%. En canvi, en el test amb 3644 imatges es diposa d'un temps mitjà de 590.2s amb un error del 14.66% (Figura 13).

Si s'estudia la variació entre l'entrenament i l'experimentació, s'observa que hi ha una relació 1:7, és a dir, per cada imatge entrenada s'obtenen 7 imatges al test.

$$v_{ent} = \frac{imatges}{temps} = \frac{14581}{16674} = 0.87\frac{imatges}{s}$$

$$v_{test} = \frac{imatges}{temps} = \frac{3644}{590.2} = 6.17\frac{imatges}{s}$$

$$relació = \frac{v_{test}}{v_{ent}} = \frac{6.17}{0.87} = 7.09$$

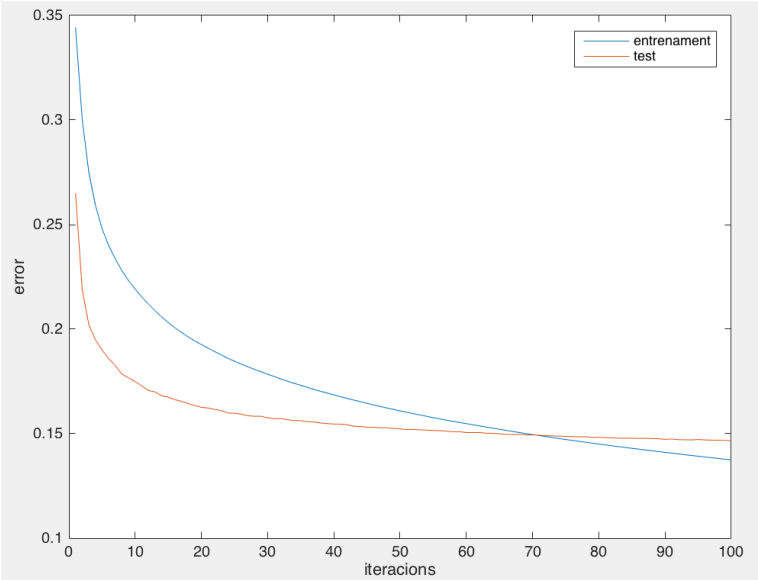


Figura 13: Gràfic d'error d'entrenament i d'error de test en el mètode ESR.

Rotació 22.5º

En el mètode RCPR es parteix d'una mostra de 19396 imatges on 15518 (4/5) passen a formar part de l'entrenament i la part restant, que són 3878 (1/5), formen part del test. Durant l'entrenament es disposa de 15518 imatges. A la taula 2 es pot veure que el temps mitjà d'entrenament és de 3060s i s'aconsegueix un error del 20.81%. En canvi, en el test amb 3879 imatges es disposa d'un temps mitjà de 94.8s amb un error del 16.23% (Figura 14).

Si s'estudia la variació entre l'entrenament i l'experimentació, s'observa que hi ha una relació 1:8, és a dir, per cada imatge entrenada s'obtenen 8 imatges al test.

$$v_{ent} = \frac{imatges}{temps} = \frac{15518}{3060} = 5.07 \frac{imatges}{s}$$

$$v_{test} = \frac{imatges}{temps} = \frac{3878}{94.8} = 40.91 \frac{imatges}{s}$$

$$relació = \frac{v_{test}}{v_{ent}} = \frac{40.91}{5.07} = 8.07$$

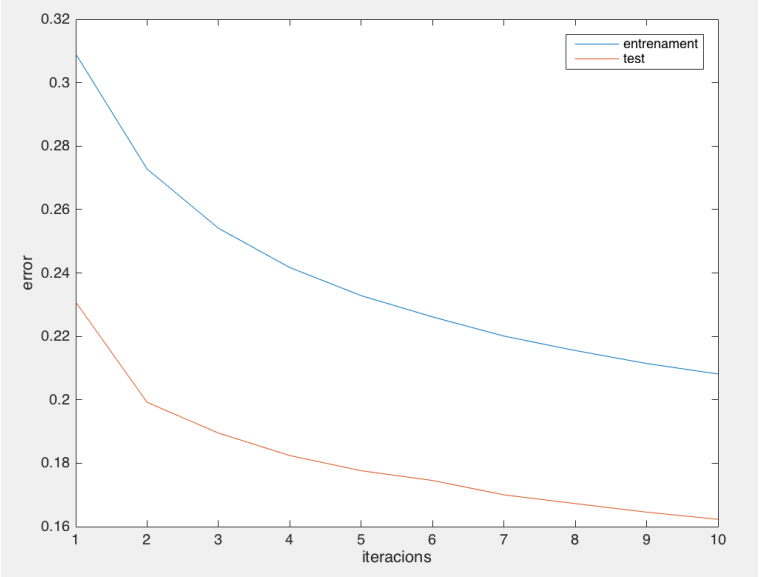


Figura 14: Gràfic d'error d'entrenament i d'error de test en el mètode RCPR.

En el mètode ESR s'obté una mostra de 19396 imatges on 15518 (4/5) passen a formar part de l'entrenament i la resta 3878 imatges (1/5), com a part del test. Durant l'entrenament es disposa de 15518 imatges. A la taula 3 es pot veure que el temps mitjà d'entrenament és de 18027.6s amb un error del 12.73%. En canvi, en el test amb 3878 imatges s'obté un temps mitjà de 652.8s amb un error del 13.35% (Figura 15).

Si s'estudia la variació entre l'entrenament i l'experimentació, s'observa que hi ha una relació 1:7, és a dir, per cada imatge entrenada es tenen 7 imatges al test.

$$v_{ent} = \frac{imatges}{temps} = \frac{15518}{18027.6} = 0.86 \frac{imatges}{s}$$

$$v_{test} = \frac{imatges}{temps} = \frac{3878}{652.8} = 5.94 \frac{imatges}{s}$$

$$relació = \frac{v_{test}}{v_{ent}} = \frac{5.94}{0.86} = 6.91$$

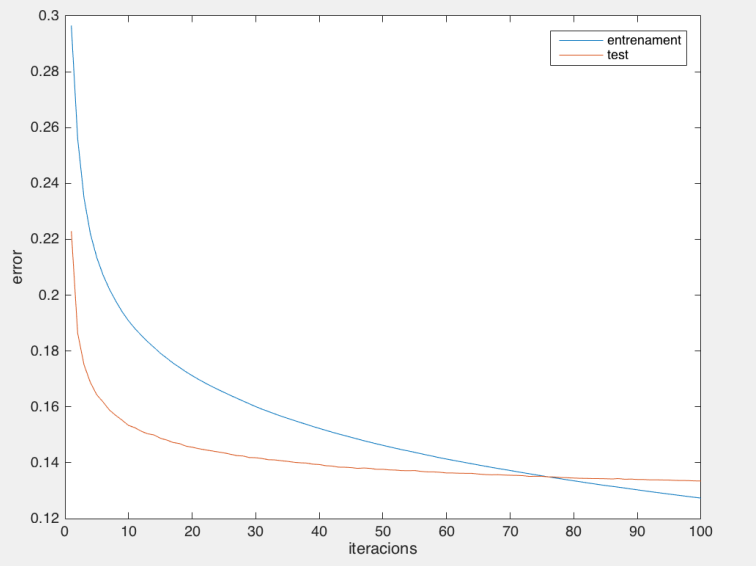


Figura 15: Gràfic d'error d'entrenament i d'error de test en el mètode ESR.

Rotació 45º

En el mètode RCPR es parteix d'una mostra de 19396 imatges on 15518 es passen com a part de l'entrenament i una mostra de 3878 imatges que es passen com a part del test. Durant l'entrenament es tenen 15518 imatges. A la taula 2 es pot observar que el temps mitjà d'entrenament és de 2836.8s amb un error del 31.22%. En canvi, en el test amb 3878 imatges s'obté un temps mitjà de 87.9s amb un error del 23.64% (Figura 16).

Si s'estudia la variació entre l'entrenament i l'experimentació, s'observa que hi ha una relació 1:8, és a dir, per cada imatge entrenada s'obtenen 8 imatges al test.

$$v_{ent} = \frac{imatges}{temps} = \frac{15518}{2836.8} = 5.47\frac{imatges}{s}$$

$$v_{test} = \frac{imatges}{temps} = \frac{3878}{87.9} = 44.12\frac{imatges}{s}$$

$$relació = \frac{v_{test}}{v_{ent}} = \frac{44.12}{5.47} = 8.07$$

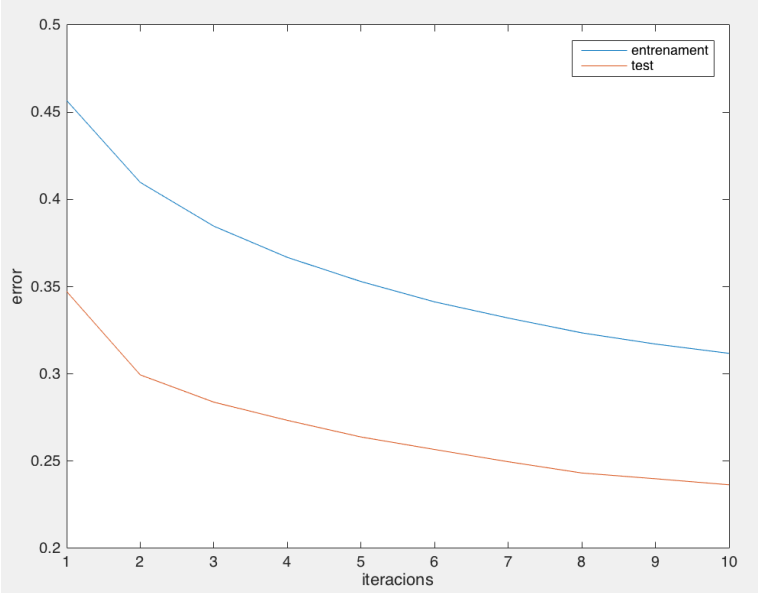


Figura 16: Gràfic d'error d'entrenament i d'error de test. en el mètode RCPR.

En el mètode ESR tenim una mostra de 19396 imatges on 15518 (4/5) es passen com entrenament i una mostra de 3878 imatges (1/5) com a part del test. Durant l'entrenament es tenen 14581 imatges. A la taula 3 es pot veure que el temps mitjà d'entrenament és de 16825.2s amb un error del 17.47%. En canvi, en el test amb 3878 imatges s'obté un temps mitjà de 620.4s amb un error del 17.05% (Figura 17).

Si s'estudia la variació entre l'entrenament i l'experimentació, s'observa que hi ha una relació 1:7, és a dir, per cada imatge entrenada s'obtenen 7 imatges al test.

$$v_{ent} = \frac{imatges}{temps} = \frac{15518}{16825.2} = 0.92\frac{imatges}{s}$$

$$v_{test} = \frac{imatges}{temps} = \frac{3878}{620.4} = 6.25\frac{imatges}{s}$$

$$relació = \frac{v_{test}}{v_{ent}} = \frac{6.25}{0.92} = 6.79$$

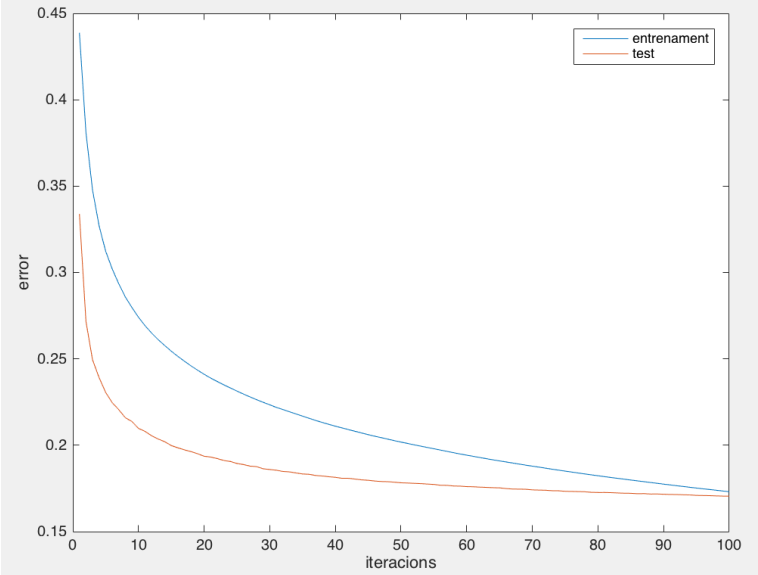


Figura 17: Gràfic d'error d'entrenament i d'error de test en el mètode ESR.

Cap tipus de rotació

En el mètode RCPR es diposa d'una mostra de 2396 imatges on 1917 es passa com entrenament i una mostra de 479 imatges que es passen com a part del test. Durant l'entrenament es disposa de 1917 imatges. A la taula 2 es pot veure que el temps mitjà d'entrenament és de 211.04s amb un error del 9.5%. En canvi, en el test amb 479 imatges s'obté un temps mitjà de 9.56s amb un error del 8.77% (Figura 18).

S'estudia la variació entre l'entrenament i l'experimentació, s'observa que hi ha una relació 1:5, és a dir, per cada imatge entrenada s'obtenen 5 imatges al test.

$$v_{ent} = \frac{imatges}{temps} = \frac{1917}{211.04} = 9.08\frac{imatges}{s}$$

$$v_{test} = \frac{imatges}{temps} = \frac{479}{9.56} = 50.10\frac{imatges}{s}$$

$$relació = \frac{v_{test}}{v_{ent}} = \frac{50.10}{9.08} = 5.52$$

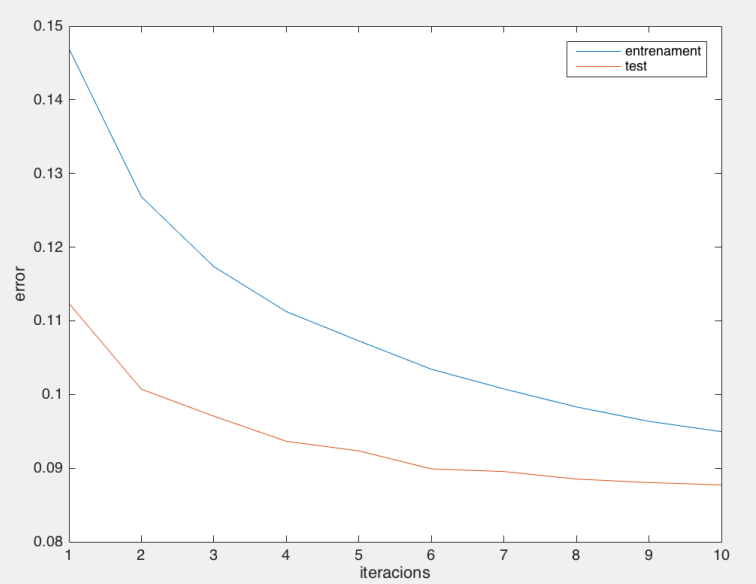


Figura 18: Gràfic d'error d'entrenament i d'error de test en el mètode RCPR.

En el mètode ESR es disposa d'una mostra de 2396 imatges on 1917 (4/5) es passen com entrenament i una mostra de 478 imatges (1/5) com a part del test. Durant l'entrenament es tenen 19181 imatges. A la taula 3 es pot veure que el temps mitjà d'entrenament és de 1914s amb un error del 4.2%. En canvi, en el test amb 479 imatges s'obté un temps mitjà de 78.9s amb un error del 8.69% (Figura 19).

Si s'estudia la variació entre l'entrenament i l'experimentació, s'observa que hi ha una relació 1:6, és a dir, per cada imatge entrenada s'obtenen 6 imatges al test.

$$v_{ent} = \frac{imatges}{temps} = \frac{1917}{1914} = 1.002 \frac{imatges}{s}$$

$$v_{test} = \frac{imatges}{temps} = \frac{479}{78.9} = 6.07 \frac{imatges}{s}$$

$$relació = \frac{v_{test}}{v_{ent}} = \frac{6.07}{1.002} = 6.04$$

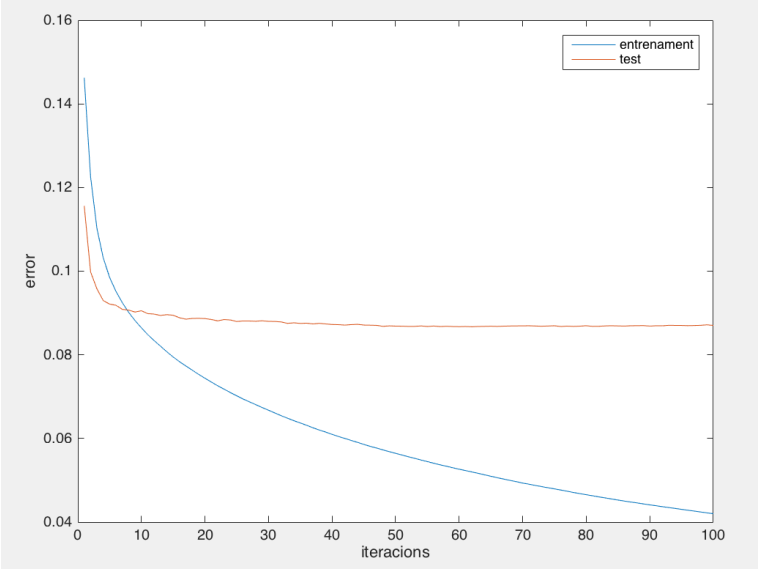


Figura 19: Gràfic d'error d'entrenament i d'error de test en el mètode ESR.

4.2 Comparació dels dos mètodes

4.2.1 Comparació quantitativa

En la comparació quantitativa s'avaluaran els mètodes respecte l'error d'entrenament i de test.

En el mètode RCPR, si s'estudien ara els errors, a la taula 2 es pot observar que l'error en l'estudi d'imatges frontals és el més petit amb un error d'entrenament del 9.50% i només un 8.77% al test. En canvi, en les rotacions de 22.5º s'obté un error 20.81% i un 16.23% en l'entrenament i en el test, respectivament, que és més del doble de l'error en una frontal. Succeix el mateix si es realitza una comparació frontal-rotació 45º. La diferència no és el doble, sinó el triple ja que s'està parlant d'un error en l'entrenament del 31.22% i en la part del test d'un 23.63%.

Si es fa una valoració global, és a dir, un estudi que inclou tot tipus d'imatges s'obté un error d'entrenament del 24.26% i un error en test del 19.14%. Es pot observar que aquests valors estan compresos entre els valors dels dos tipus de rotació, ja que les imatges frontals disminueixen quantitativament aquest error.

Finalment, s'ha comprovat que el RCPR té entre 1.5 i 4 vegades més rapidesa que altres mètodes de la competència i que redueix el nombre de casos d'error a quasi bé la meitat.

En el mètode ESR s'observa que, on no s'inclou rotació en cap tipus d'imatge, el mètode és més precís i s'obtenen millors resultats. En canvi, la relació d'error entre les imatges frontals-rotació 22.5º varien, de ser el doble passen a ser el triple en

l'entrenament. Per contra, en les imatges frontals-rotació 45º, l'error s'incrementa el quàdruple. No obstant, en la part del test s'observa que l'error és més gran que l'error d'entrenament tant en imatges amb una rotació de 22.5º com de 45º.

Si es fa una valoració global, és a dir, un estudi que inclou tot tipus d'imatges s'obté un error d'entrenament del 13.75% i un error en test del 14.66%. Es pot observar que aquests valors estan compresos entre els valors dels dos tipus de rotació, ja que les imatges frontals disminueixen quantitativament aquest error.

En els gràfics d'error, s'aprecia que l'increment del nombre d'iteracions, a partir d'un nombre elevat d'itereacions en l'entrenament, no garantitza un millor resultat en el test. Per tant, es pot dir que el fet de tenir un major nombre d'iteracions per entrenar el model, són casos particulars de les imatges i no rellevants.

S'observa que els dos mètodes son molt precisos per imatges frontals, però en canvi no podem dir el mateix per imatges amb algun tipus de rotació. El mètode RCPR dóna errors més grans a les imatges amb rotació, peró no té un temps computacional elevat, mentre que amb l'ESR s'observa que les imatges amb gir tenen menor error, però per contra s'obté un augment de temps important.

4.2.2 Comparació qualitativa

En aquesta secció es mostren unes comparacions visuals dels dos mètodes juntament amb l'alineació real de les imatges. Com es pot apreciar en les imatges frontals els errors són gairebé mínims (Figura 20).

Cap tipus de rotació

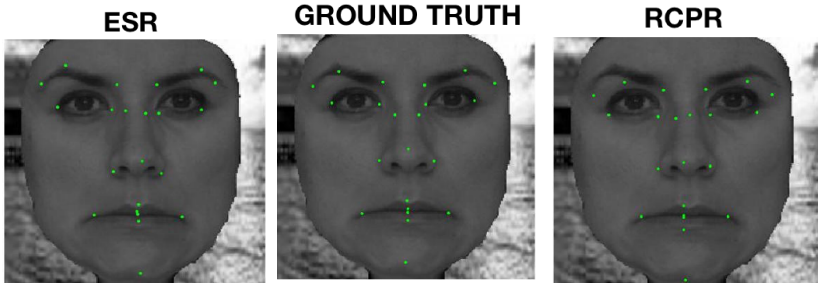


Figura 20: Comparació d'una imatge amb els dos mètodes i el *ground truth*

Tant en les rotacions de 22.5º com en les de 45º, ja es veu perfectament com els mètodes no acaben d'ajustar-se a la imatge. Es destaquen alguns punts amb conflictes, com podrien ser la barbeta o les parts interiors del llavi. Aquesta inestabilitat pot ser causada probablement perquè la regió al voltant del punt no és molt discriminativa, és a dir, no hi ha gaires canvis d'intensitat. Si s'observen les

imatges de cadascun dels dos mètodes (Figura 21, Figura 22 i Figura 23) es pot veure que en el mètode ESR els punts són més precisos en punts com les celles, el contorn dels ulls o la barbeta i el RCPR mostra alguns desviaments majors.

Rotació 22.5º

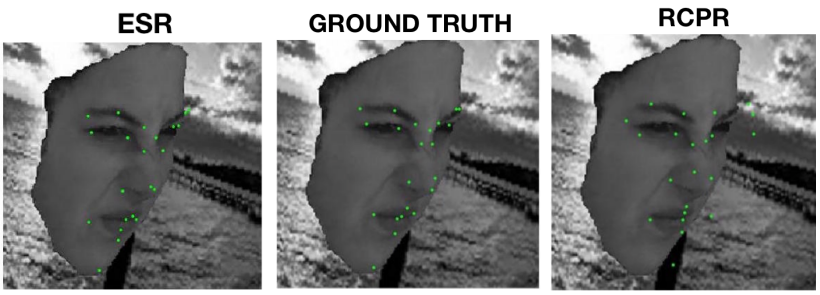


Figura 21: Comparació d'una imatge amb els dos mètodes i el *ground truth*

Rotació 45º

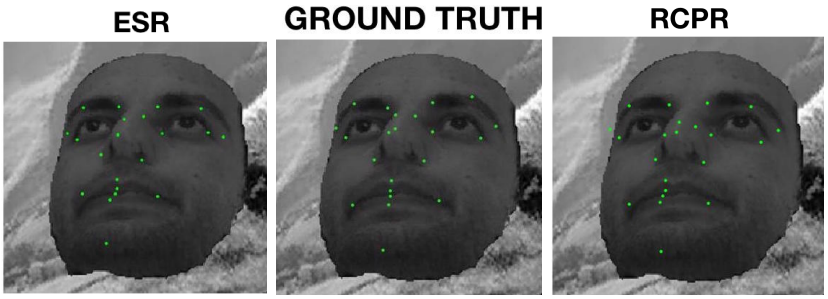


Figura 22: Comparació d'una imatge amb els dos mètodes i el *ground truth*

Tot tipus de rotació

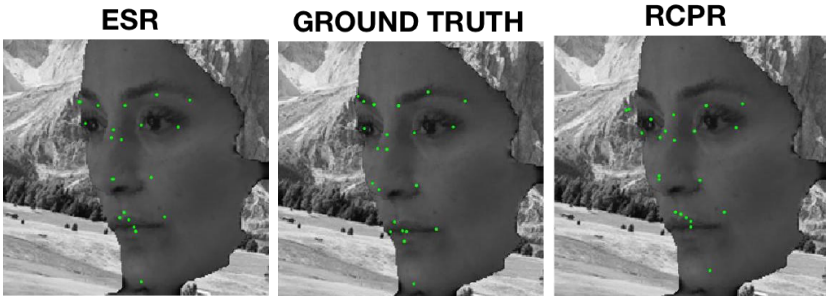
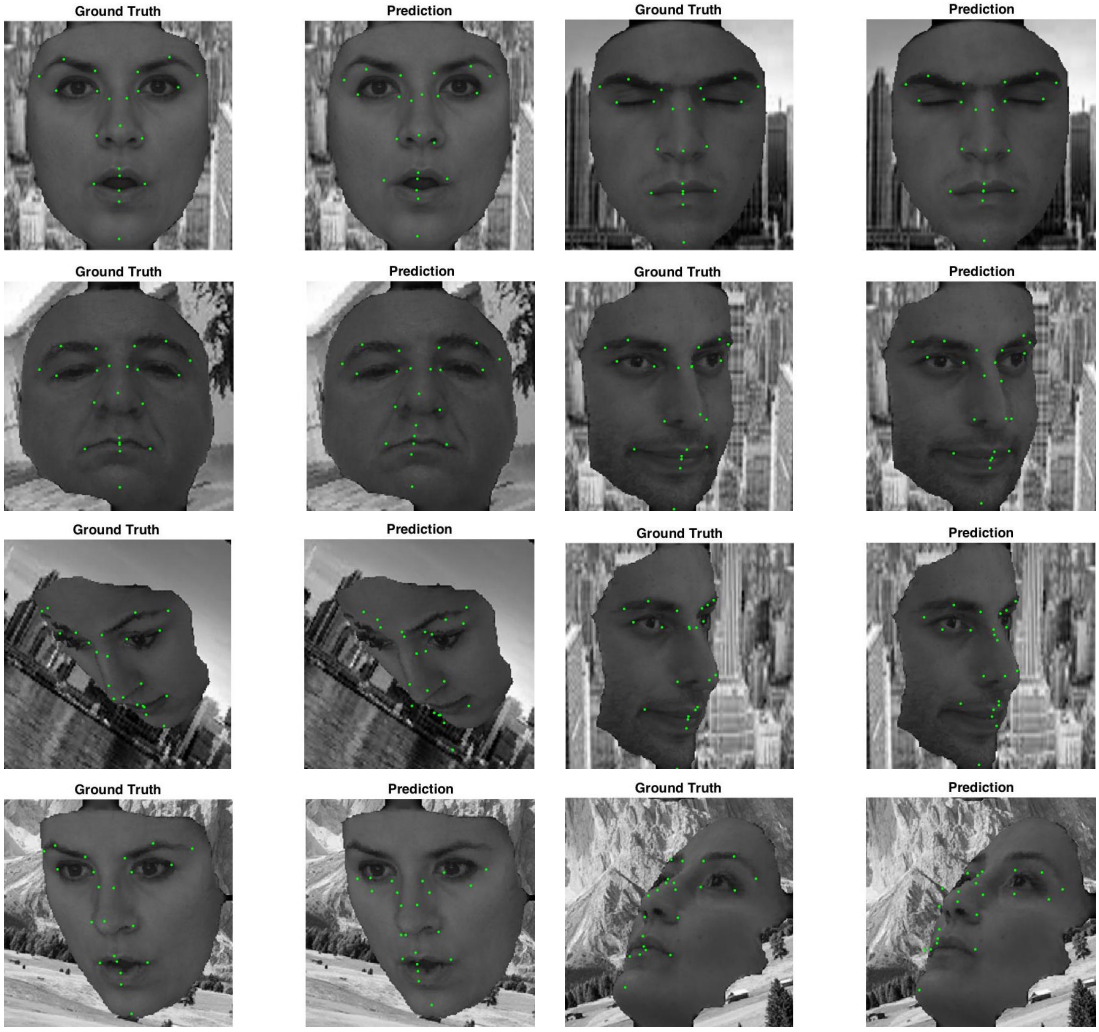
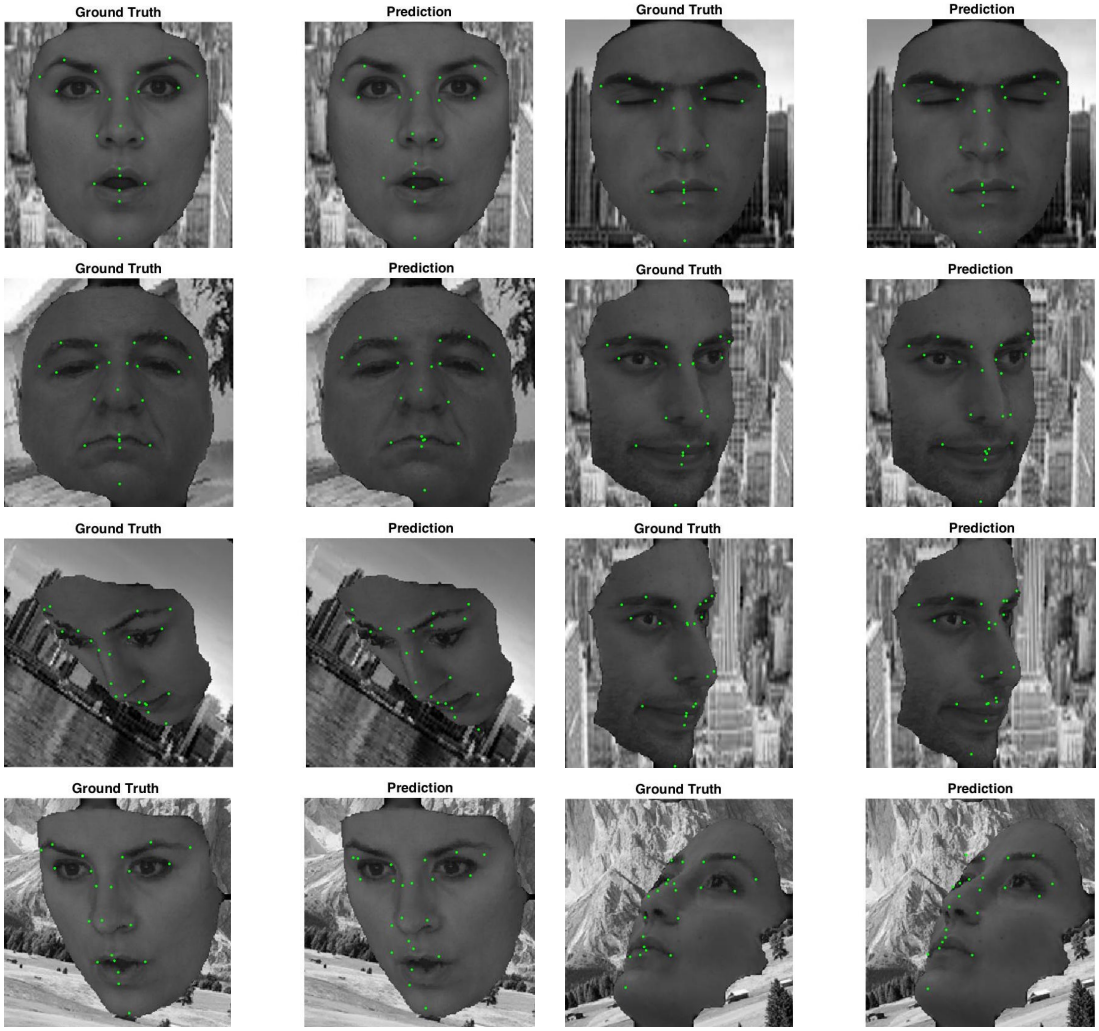


Figura 23: Comparació d'una imatge amb els dos mètodes i el *ground truth*

A continuació, es mostren més exemples dels mètodes RCPR i ESR (Taula 4 i Taula 5).



Taula 4: Imatges aleatòries en el mètode RCPR



Taula 5: Imatges aleatòries en el mètode ESR

5 Conclusions

Després d'haver vist l'estudi de la fisonomia facial, es conclou que l'estudi dels punts representatius d'una cara és un estudi complicat i amb una precisió millorable notablement. Encara que s'ha observat com funcionaven dos mètodes diferents però amb el mateix nombre de punts, els resultats obtinguts són bastant dispars.

S'observa que els dos mètodes donen millors resultats quan es tracta d'imatges frontals, en canvi les imatges amb gir, especialment les de 45º, presenten uns errors més grans; tant en els entrenaments com en el tipus test.

Tot i així, no podem ser capaços de decidir quin dels dos mètodes és millor. En les imatges frontals tots dos mètodes són força eficients, però mostren resultats de diferent presició quan imposem imatges amb gir; és a dir, el mètode RCPR dóna errors més grans a les imatges amb gir, però el temps no augmenta excessivament. Mentre que amb l'ESR s'observa que les imatges amb gir tenen menor error, però en contra partida d'un augment de temps important.

Els resultats que s'observen estan donats per a un nombre d'iteracions igual a 10 o bé igual a 100. El primer cas podríem dir que es queda curt d'iteracions. En canvi, el segon potser en són massa. Una millora que es podria fer és experimentar amb 50 iteracions per veure si la relació 10-50-100 dóna realment una millora important. En el moment de doblar el nombre d'iteracións de 50 a 100, possiblement la millora sigui inferior a la de 10 a 50.

D'altra banda, existeixen varies vies d'investigació a partir d'aquest treball. Una d'elles seria la possibilitat d'intentar trobar el nombre òptim d'imatges per tal de no entrenar la màquina amb més imatges de les necessàries. Una altra, seria repetir l'estudi amb altres bases de dades de diferent nombre de punts de referència. I per últim, la millora del nombre d'iteracions per obtenir el resulat més precís amb el menor nombre d'iteracions. Evidentment, la unió d'aquestes tres vies seria la més efectiva.

Una segona part d'aquesta mateixa investigació que es deixa oberta per al lector és el fet de l'estudi facial amb imatges que tinguin algun tipus d'oclusió.

Referències

[1] Burgos Artizzu, X.P.; Dollár, P.; Perona, P.: Robust face landmark estimation under occlusion, *California Institute of Technology*, `http://nubr.co/2rkorI`, 2013.

[2] Cao, X.; Wei, Y.; Weu, F.; Sun, J.: Face Alignment by Explicit Shape Regression, *Microsoft Research Asia*, 2012.

[3] Martinez, B.; Valstar, M.P.; Binefa, X.; Pantic, M.: Local Evidence Aggregation for Regression Based Facial Point Detection, *IEEE*, `http://nubr.co/aMyf2x`, 2013.

[4] Wang, N.; Gao, X.; Tao, D.; Li, X.: Facial Feature Point Detection: A Comprehensive Survey, *International Journal of computer vision manuscript No*, `http://nubr.co/t6b68M`, 2014.

[5] Corneanu, C.A.; Oliu, M.; Cohn, J.F.; Escalera, S.: Survey on RGB, 3D, Thermal, and Multimodal Approaches for Facial Expression Recognition: History, Trends, and Affect-related Applications, 2015.

[6] Dóllar, P.; Weilender, P.; Perona, P.: Cascaded Pose Regression, *California Institute of Technology*, `http://nubr.co/lVL5XP`, 2010.

[7] Xiong, X.; De la Torre, F.: Supervised Descent Method and its Applications to Face Alignment, *The Robotic Institute, Carnegie Mellon University, Pittsburgh PA*, `http://nubr.co/XcZbsF`, 2013.

[8] Oliu, M.: Head pose recovery and shape estimation in still images, *Master in Artificial Intelligence (UPC, URV, UB)* `http://nubr.co/BqqjFx`, 2014.

[9] Aler, R.; evaluación de técnicas de aprendizaje, *Universidad Carlos III, Madrid* `http://nubr.co/lR3VyA`

[10] MATLAB-The Language of Technical Computing: `http://nubr.co/GFhFCM`

[11] Cootes, T.F; Cooper, D.H; Graham, J.: Active Shape Models-Their Training and Aplication, *Department of Medical Biophysics, University of Manchester* `http://nubr.co/p9nJni`, 1994.

[12] Savran, A.; Alyüz, N.; Dibeklioglu, H.; Çeliktutan, O.; Gökberk, B.; Sankur, B.; Akarun, L.: Bosphorus Database for 3D Face Analysis, *Bogaziçi University, Electrical and Electronics Engineering Department, Bogaziçi University, Computer Engineering Department, Philips Research, Eindhoven, The Netherlands* `http://nubr.co/Vl7CKB`, 2008.

[13] Molina, N.P: Herramientas para investigar, ¿Qué es el estado del arte?, *Universidad de la Salle*
http://nubr.co/9sX4l8, 2005.

[14] Sozou, P.D; Cootes, T.F; Taylor, C.J; Di Mauro, E.C: Non-linear Point Distribution Modelling using a Multi-layer Perceptron, *Department of Medical Biophysics University of Manchester*
http://nubr.co/r8zt5j, 1995.

[15] Matthews, I.; Baker, S.: Active Apperance Models Revisited, *The Robotic Institute Carnegie Mellon University*
http://nubr.co/MMDB2U, 2004.

[16] Zhou, K.; Comaniciu, D.: Shape Regression Machine, *Integrated Data Systems Department, Siemens Corporate Research*
http://nubr.co/853LhP, 2007.

[17] Kozakaya, T.; Shibata, T.; Takeguchi, T.; Nishiuara, M: Fully Automatic Feature Localization for Medical Images using a Global vector concentration Approach, *Corporate Research and Developpment Center*
http://nubr.co/Kx14WW, 2010.

[18] Coughlan, J.M.; Ferreiro, S.J.: Finding Deformable Shapes using Loopy Beliet Propagation, *Smith-Kettlewell Institute Department of Stadistics, Federal University of Minas Gerais UFMG*, 2002.

[19] Liang, L.; Wen, F.; Xu, Y.-Q.; Shum, H.-Y.; Tang, X.: Accurate Face Alignment using Shape Constrained, *Microsoft Research Asia*
http://nubr.co/Eo6W7W, 2006.

[20] Zhao, C.; Cham, W.-K.; Wang, X.: Joint Face Alignment with a generic deformable Face Model, *Department of Electronic Engineering The Chinese University of Hong Kong*
http://nubr.co/wPq25i, 2011.

[21] Zhu, X.; Ramanan, D.: Face Detection, Pose Estimation, and Landmark Localization in the wild *Dept. of Computer Science, University of California*
http://nubr.co/zTLE3l, 2012.

[22] Vukadinovic, D; Pantic, M.: Fully automatic facial feature point detection using Gabor feature based boosted classifiers , *Electrical Engineering, Mathematic and Computer Science Delft, The Netherlands*
http://nubr.co/Jqj8N3, 2005.

[23] Zhao, X.; Shan, S.; Chai, X.; Chen, X.: Cascaded shape space pruning for robust facial landmark, *Key Lab. of Intell. Inf. of Comput. Technol., Bejing*
http://nubr.co/0EbHyT, 2013.

[24] Sun, Y.; Wang, X.; Tang, X.: Deep convolutional network cascade for facial point detection, *Department of Informatic and Electronic Engineering, The Chinese University of Hong Kong*
`http://nubr.co/jQMWv2`, 2013.

[25] Wu, Y.; Wang, Z.; Ji, Q.: Facial feature tracking under varying facial expressions and face poses based on restricted Boltzmann machine, *ECSE Dept. Rensselaer Polytech. Inst, Troy, NY, USA*
`http://nubr.co/jTdKOP`, 2013.

[26] Tresadern, P.A.; Ionita, M. C,: Real-time facial feature tracking on a mobile device, *Internacional journal of computer vision*, 2012.