



Escola Tècnica Superior d'Enginyeria  
de Telecomunicació de Barcelona

UNIVERSITAT POLITÈCNICA DE CATALUNYA

## **PROJECTE FINAL DE CARRERA**

# **SISTEMA DE REONEIXEMENT DE L'EXPRESSIÓ FACIAL BASAT EN MODELS ACTIUS D'APARENÇA** (FACIAL EXPRESSION RECOGNITION SYSTEM BASED ON ACTIVE APPEARANCE MODELS)

*Estudis: Enginyeria de Telecomunicació*

*Autor: Elia Nebot Gil*

*Directora: Montse Pardàs Feliu*

*Any: Decembre 2010*



## **Resum del projecte**

L'objectiu d'aquest projecte és el desenvolupament d'un sistema automàtic de reconeixement de les expressions facials. El sistema consta de dues parts: la detecció i extracció dels trets facials i la classificació de l'expressió utilitzant els trets extrems.

Se'ns va proposar basar la detecció i extracció dels trets facials en la teoria dels models actius d'aparença i en la implementació d'aquesta feta per Mikkel Stegmann. El primer pas, doncs, és l'estudi dels models actius d'aparença i de la implementació AAM-API.

En aquest punt proposem un sistema per a la classificació de les expressions facials. Aquest sistema utilitza els models AAM de la següent forma: es generaran un conjunt de models AAM on cadascun d'ells contindrà la informació de l'expressió facial d'una emoció. D'aquesta manera, donada una imatge d'entrada, sintetitzant-la per a cada model i avaluant quina d'elles s'ajusta millor podrem estimar quina serà la seva expressió facial i, per tant, la seva emoció.

La nostra feina, doncs, és trobar la millor opció per avaluar quin dels models s'ajusta millor a la imatge original. Això ho farem, primer, proposant i estudiant una sèrie de paràmetres que quantifiquin l'error entre la imatge original i la imatge sintetitzada. I, després, proposant i estudiant diversos mètodes de decisió basats en aquests paràmetres per a cada model AAM utilitzat.

El resultat és un mètode de classificació que aconsegueix taxes d'error no menyspreables, degudes als error inherents al sistema de models AAM així com a l'ambigüitat entre les diferents expressions analitzades.

## **Resumen del proyecto**

El objetivo de este proyecto es el desarrollo de un sistema automático de reconocimiento de las expresiones faciales. El sistema consta de dos partes: la detección y la extracción de las características faciales y la clasificación de la expresión utilizando las características extraídas.

Se nos propuso basar la detección i extracción de las características faciales en la teoría de modelos activos de apariencia y en la implementación de ésta hecha por Mikkel Stegmann. El primer paso, así, es el estudio de los modelos activos de apariencia y de la implementación AAM-API.

En este punto proponemos un sistema para la clasificación de las expresiones faciales. Este sistema utiliza los modelos AAM de la siguiente forma: se generaran un conjunto de modelos AAM donde cada uno de ellos contendrá la información de la expresión facial de una emoción. De este modo, dada una imagen de entrada, sintetizándola por cada modelo y evaluando cuál de ellas se ajusta mejor podremos estimar cual será su expresión facial y, por tanto, su emoción.

Nuestro trabajo, así, es encontrar la mejor opción para evaluar cual de los modelos se ajusta mejor a la imagen original. Esto lo haremos, primero, proponiendo y estudiando una serie de parámetros que cuantifiquen el error entre la imagen original y la sintetizada. Y, después, proponiendo y estudiando diversos métodos de decisión basados en estos parámetros para cada modelo AAM utilizado.

El resultado es un método de clasificación que consigue tasas de error no despreciables, debidas a los errores inherentes al sistema de modelos AAM, así como a la ambigüedad que existe entre las diferentes expresiones analizadas.

**Abstract**

The aim of this project is the development of an automatic facial expression recognition system. The system has two parts: the detection and extraction of the facial features and the expression classification based on the extracted features.

It is suggested to base the detection and extraction of the facial features on the Active Appearance models theory and its implementation done by Mikkel Stegmann. The first step we have to do is to study the Active appearance models and the AAM-API implementation.

At this moment, a facial features classification system is proposed. This system uses the AAM models as follows: a set of AAM models will be created, each model will contain the facial expressions information that belongs to a specific emotion. Then, having an input image, it is synthesized for each model and then each model synthesis is evaluated to find the one that fits better with the original image. This way, we could estimate which facial expression has the input image and so its emotion.

So, our work is find the best evaluation mode, so we could know which model fits better with the original image. We will do this, first, proposing and studying several parameters that could quantify the error between the original and synthesized image. And, then, proposing and studying several decision methods, based on these parameters, for each used AAM model.

The result is a classification method that obtains not insignificant error taxes, due to the inherent errors in the AAM models system, as well as the ambiguity between the analyzed facial expressions.



# Índex

<b>Col·laboracions .....</b>	<b>3</b>
<b>Agraïments .....</b>	<b>4</b>
<b>1 Introducció .....</b>	<b>5</b>
<b>2 Estat de l'art.....</b>	<b>7</b>
2.1 Adquisició del rostre.....	8
2.2 Extracció dels trets facials .....	8
2.3 Classificació de l'expressió facial .....	10
<b>3 Esquema general del sistema .....</b>	<b>11</b>
3.1 Objectiu del sistema .....	11
3.2 Esquema general.....	11
3.2.1 Classificació a partir de les característiques extretes.....	12
3.2.2 Classificació amb models AAM.....	12
3.3 Sistema basat en la classificació per models AAM.....	13
3.4 Parts del sistema .....	14
<b>4 Teoria AAM .....</b>	<b>15</b>
4.1 Introducció.....	15
4.2 Teoria Models Estadístics d'Aparença .....	15
4.2.1 Models Estadístics de forma (Statistical Shape Models).....	15
4.2.1.1 La forma i la seva representació.....	15
4.2.1.2 Alineació de les formes (Shape Alignment).....	16
4.2.1.3 Modelant la variació de la forma.....	17
4.2.1.4 Generar formes plausibles .....	18
4.2.1.5 Trobar la forma plausible més propera.....	18
4.2.1.6 Ajustar el model a nous punts .....	19
4.2.1.7 Estimació p(forma).....	20
4.2.2 Models Estadístics d'aparença.....	21
4.2.2.1 Models Estadístics de textura .....	21
4.2.2.2 Models Combinats d'Aparença .....	22
4.2.2.3 Aproximant un nou exemple .....	23
4.2.3 Interpretació d'imatges amb Models .....	23
4.2.3.1 Escollir una funció d'ajust.....	24
4.2.3.2 Optimitzar l'ajust del model.....	24
4.2.4 Model actiu de forma (Active Shape Models).....	25
4.2.5 Model Actiu d'aparença (Active Appearance Models).....	25
4.2.5.1 Visió general de la cerca amb AAM .....	26
4.2.5.2 Aprenent a Corregir els Paràmetres del Model .....	26
4.2.5.3 Refinament del model iteratiu (Iterative Model Refinement) .....	27
4.3 Interfície AAM .....	28
4.3.1 AAM-API.....	28
4.3.1.1 Motor AAM-API.....	30
4.3.1.2 Interfície Consola aamc .....	31
<b>5 Sistema de reconeixement d'emocions.....</b>	<b>32</b>
5.1 Funcionament general del sistema.....	32
5.2 Detecció de forma.....	33
5.2.1 Caracterització de les imatges .....	33
5.2.2 Operacions amb l'aplicació de Mikkel B. Stegmann .....	34
5.2.2.1 Generació dels models de cada emoció $S_i$ .....	34

5.2.2.2	Síntesi d'imatge mitjançant models AAM .....	37
5.3	Decisió .....	41
5.3.1	Paràmetres .....	41
5.3.1.1	Pre processat de les dades.....	42
5.3.1.2	Proposta paràmetres decisió .....	47
5.3.1.3	Estudi Paràmetres Decisió .....	57
5.3.1.4	Paràmetre de Decisió Òptim.....	65
5.3.2	Mètode de Decisió .....	68
5.3.2.1	Anàlisi dels paràmetres.....	69
5.3.2.2	Proposta criteris de decisió .....	78
5.3.2.3	Estudi dels criteris de decisió .....	84
5.3.2.4	Criteri de decisió final .....	90
5.4	Implementació .....	91
5.4.1	Reconeixement d'emocions (AAM).....	91
5.4.1.1	Execució del reconeixement d'emocions .....	92
5.4.1.2	Generació asf .....	92
5.4.1.3	Càlcul de paràmetres .....	93
5.4.1.4	Generació Excel.....	93
5.4.2	ENG PFC: Estudi Reconeixement d'emocions .....	93
5.4.2.1	Importar dades des de Excel.....	95
5.4.2.2	Inicialitzar els tipus d'imatges i models .....	96
5.4.2.3	Configuració de l'estudi de dades .....	96
5.4.2.4	Generar els estudis de dades.....	96
5.4.2.5	Decisió per zones.....	97
5.4.2.6	Criteris de decisió final.....	97
5.4.2.7	Resum de les decisions .....	98
<b>6</b>	<b>Resultats .....</b>	<b>99</b>
<b>7</b>	<b>Conclusió .....</b>	<b>117</b>
<b>8</b>	<b>Referències .....</b>	<b>119</b>
	<b>Annex Configuració AAM-API.....</b>	<b>123</b>



**Col·laboracions**



Departament de Teoria de Senyal i Comunicacions

## **Agraïments**

Aquest treball no hauria estat possible sense l'ajuda de la meva mare, germana, avis i tiet que sempre van animar-me a estudiar.

També els meus anys d'escola i institut, ja que van influir a la meva decisió d'estudiar una enginyeria i en concret l'enginyeria de telecomunicació.

Agraïr també a professors i companys de la UPC amb els que vaig compartir anys i hores de classe i d'estudi.

En aquest punt vull agrair de forma especial a Montse Pardàs per la seva paciència i dedicació en aquests anys en que les circumstàncies demoraven l'execució d'aquest treball.

I, per últim, agrair al meu marit pels ànims i paciència tots aquests anys.

Sense tots ells aquest treball no hauria estat possible.

## 1 Introducció

Al llarg de les tres últimes dècades, l'anàlisi de l'expressió facial ha esdevingut un àrea de recerca potencialment aplicable a àrees com interfícies home-màquina, caps que parlen, recuperació d'imatges i anàlisi de les emocions humanes. Això és així, ja que les expressions facials no només reflecteixen les emocions sinó que també mostren altres activitats mentals, interaccions socials i senyals psicològics.

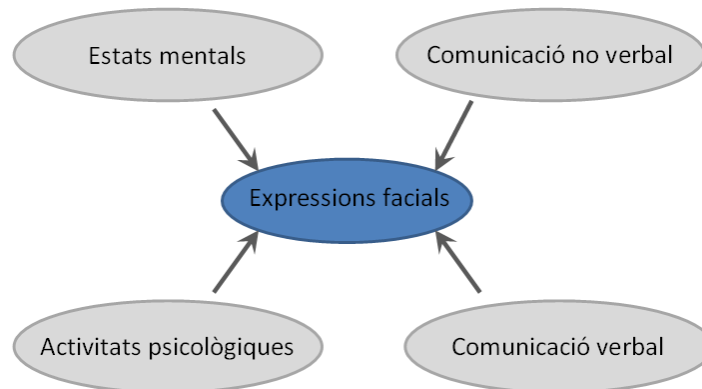
La comunicació cara a cara dels éssers humans és el model a seguir per dissenyar les interfícies entre l'home i les màquines. Les característiques principals de la comunicació humana són la multiplicitat i la diversitat de sentits i de canals de comunicació. Un canal és un mitjà de comunicació mentre un sentit és una eina utilitzada per percebre senyals del món exterior. Els exemples de canals de comunicació són el canal auditiu que comporta la parla i l'entonació vocal, el canal visual que comporten les expressions facials i els moviments corporals.

En una comunicació cara a cara estàndard s'utilitzaran diversos canals i s'activaran diversos sentits. Com a resultat, la comunicació esdevé elevadament flexible i robusta. L'error en un canal es recupera amb un altre canal i un missatge en un canal es pot explicar pel d'un altre.

Mitjançant el rostre podem identificar membres d'altres espècies, interpretar el que s'ha dit mitjançant la lectura dels llavis, i entendre l'estat emocional d'algú i les seves intencions basant-nos en les expressions facials mostrades. En aquest context, el terme "cara a cara" indica que el rostre ocupa un paper essencial en la comunicació interpersonal. És així com un HCI (*Human-Computer Interface*) s'hauria de desenvolupar per facilitar una interacció entre home i màquina de forma robusta, natural, eficient i efectiva.

Els avenços recents en l'anàlisi d'imatges i reconeixement de patrons obren la possibilitat a la detecció automàtica i a la classificació dels senyals facials emocionals i de conversa. L'automatització de l'anàlisi de les expressions facials podria convertir les expressions facials en el marc de la interacció entre l'home i la màquina en un nou sentit de comunicació i fer que la interacció fos ajustada i més eficient.

Sovint succeeix a la comunitat de visió per computador que el reconeixement de l'expressió facial es confon amb el reconeixement de l'emoció humana. La diferència entre el reconeixement de l'expressió facial i de l'emoció humana és la següent: el reconeixement de l'expressió facial s'encarrega de la classificació dels moviments facials i de la deformació de les característiques facials en classes abstractes, basant-se únicament en informació visual, mentre que les emocions humanes són el resultat de diferents factors i es detecta en base als canals com la veu, la postura, els gestos, la direcció de la mirada i les expressions facials. A més, les emocions no són la única font de les expressions facials com es pot veure a la Figura 1-1. En contrast amb el reconeixement de l'expressió facial, el reconeixement d'emocions és un intent d'interpretació i sovint requereix la comprensió d'una situació, juntament amb la disponibilitat d'informació completa del context.

*Figura 1-1*

Els humans tenen la capacitat de detectar i interpretar les expressions facials de forma natural. En aquests moments, és difícil desenvolupar un sistema automàtic que aconseguixi implementar aquesta capacitat. Es troben diversos problemes: la detecció del rostre com a una zona de la imatge, l'extracció de la informació de l'expressió facial i la classificació de l'expressió (per exemple, en categories d'emocions). Un sistema que implementi aquestes operacions de forma acurada i precisa i en temps real, podria constituir un gran pas en la consecució d'una interacció més humana entre l'home i la màquina.

La detecció i extracció de l'expressió facial són els passos indispensables per poder dur a terme la classificació de l'expressió. La classificació de l'expressió facial és el pas del procés sobre el que es treballarà en aquest projecte. S'hauran d'intentar superar certs reptes, el primer és la correcta detecció i extracció de la informació pel sistema AAM, el segon és la correcta classificació de les expressions facials i el tercer és que la proposta per al mètode de decisió d'aquest treball pugui assegurar una taxa d'error acceptable.

## 2 Estat de l'art

L'anàlisi de l'expressió facial es remunta al segle XIX, quan, al 1872, Darwin [4] va demostrar la universalitat de les expressions facials i la seva continuïtat en els homes i en els animals i va afirmar també que hi ha emocions innates que es van originar en els hàbits útils associats. Al 1971 Ekman i Friesen [5] van postular sis emocions primàries. Cada una d'elles posseeix un contingut distintiu d'una única expressió facial. Aquestes representacions d'emocions prototipus se les denomina també emocions bàsiques. Sembla que són universals per a totes les ètnies i cultures humanes i són alegria, tristesa, por, fàstic, sorpresa i enuig.

En el passat, l'inici de la recerca en l'anàlisi de les expressions facials es va dur a terme pels psicòlegs, i no va ser fins al 1978 quan Suwa [6] va presentar la seva investigació preliminar sobre l'anàlisi automàtic de l'expressió facial en una seqüència d'imatges. En els anys 90, la recerca sobre l'anàlisi automàtic de l'expressió facial va guanyar empena gràcies al treball pioner de Mase i Pentlans [7]. Les raons d'aquest renovat interès en les expressions facials són diverses, però sobretot va ser degut als èxits aconseguits en diverses àrees com detecció facial, seguiment i reconeixement de rostres així com la millora de l'eficiència computacional.

Relativament pocs estudis combinen diversos sentits en un mateix sistema per analitzar les reaccions comunicatives humanes. Uns exemples són els treballs de Chen [8] i De Silva [9] que van estudiar els efectes d'una detecció combinada de les expressions facials i vocals de les emocions.

La majoria dels estudis tracten els canals de comunicació humans per separat, així ho va indicar Nakatsu [10]. Uns exemples del sistema presentat són: interpretació emocional de les veus humanes [11], [12], [13], [14], reconeixement d'emocions mitjançant el reconeixement de patrons dels trets fisiològics [15], detecció i interpretació dels gestos de les mans [16], reconeixement dels moviments del cos [17], [18], [19] i anàlisi de l'expressió facial.

Les expressions facials es generen a partir de contraccions dels músculs facials, que resulten en la deformació temporal dels trets facials com les parpelles, celles, nas, llavis i textura de la pell. Els canvis en els músculs són ràpids, la seva duració és només d'uns pocs segons, rarament més de cinc segons o menys de 250 ms.

Volem quantificar de forma acurada l'expressió facial i d'aquesta manera necessitarem una terminologia per a descriure-la. La localització de l'acció facial és important, la seva intensitat o força així com el seu moviment dinàmic. La intensitat de l'expressió es pot mesurar determinant les deformacions geomètriques dels trets facials o la densitat d'arrugues que apareixen en certes zones del rostre.

Tenint en compte que la intensitat de l'acció entre persones pot variar, incrementa la dificultat per determinar la intensitat en valor absolut sense haver de fer referència a un subjecte conegut.

L'anàlisi automàtic de les expressions facials és una tasca complexa, així com ho és la variació de les fisonomies entre un individu i un altre degut a les diferències d'edat, ètnia, gènere, cabell facial, productes cosmètics i objectes que tapen com ulleres o cabell. A més, els rostres apareixen de forma dispar degut als canvis en la posició i en la il·luminació. Aquestes variacions s'han de tractar als diferents passos del sistema d'anàlisi automàtic de l'expressió facial, veure Figura 2-1.

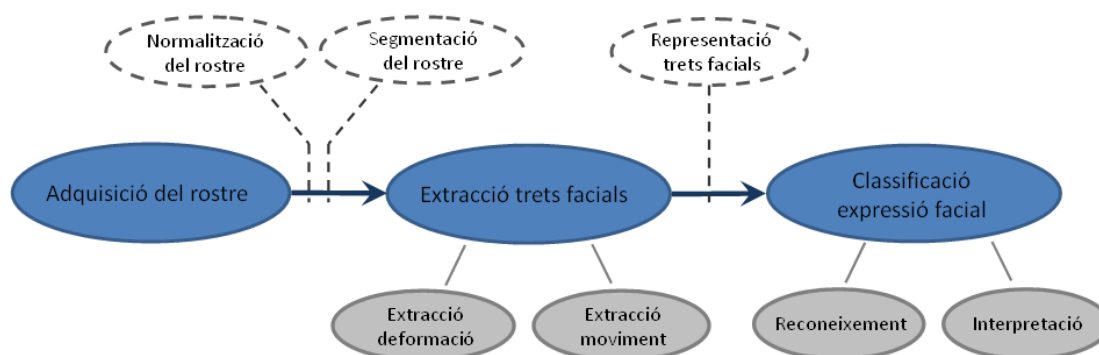


Figura 2-1

## 2.1 Adquisició del rostre

L'adquisició del rostre idealment implementa un detector automàtic de rostre que permet localitzar rostres en escenaris complexos amb fons d'imatge ple de coses. Certs mètodes d'anàlisi del rostre necessiten la posició exacta del rostre per extreure els trets facials que interessin, mentre altres feines només necessiten una posició aproximada del rostre. Aquest es el cas dels models actius d'aparença (*active appearance models*) [22]. Hong [23] va utilitzar el sistema *PersonSpotter* d'Steffens [24] per implementar el seguiment de rostres en temps real. Essa i Pentland [33] localitzaven rostres utilitzant el view-based i el mètode modular d'autoespais de Pentland [25].

Com abans s'ha esmentat un dels principals problemes és la variació en la posició del rostre dins la imatge i en la il·luminació d'aquestes. Així que pot ser una bona idea normalitzar les imatges adquirides abans del seu anàlisi.

La normalització dels rostres no és obligatòria en alguns mètodes d'anàlisi facial, com els models basats en l'aparença [26] i els basats en el moviment local [27]. Aquests mètodes aconsegueixen localitzar el rostre dins de la imatge utilitzant rotacions fora de pla sense haver d'executar la normalització facial.

## 2.2 Extracció dels trets facials

Després de la detecció d'un rostre en una escena, el següent pas es extreure la informació sobre l'expressió facial de forma automàtica. Un dels temes fonamentals és la representació de la informació visual que un rostre pot revelar. Els resultats dels experiments duts a terme per Johansson [54][55] suggereixen que les propietats visuals del rostre, centrats en la informació de l'expressió facial mostrada, es poden extreure seguint els moviments de punts que pertanyin a trets facials (celles, ulls i boca) i analitzant la relació entre ells.

Els mètodes d'extracció dels trets es poden categoritzar en funció de l'objecte sobre els que es centren. Si ho fan en el moviment o en la deformació dels rostres i dels seus trets o si ho fan de forma local o global. La Taula 2-1 dóna una visió general dels mètodes emprats per la comunitat de visió per computador per dur a terme la tasca de l'anàlisi de l'expressió facial per a l'extracció de moviment d'una manera global.

Extracció moviment	Mètode global
Basat en imatge	Filtres Gabor
Basat en model	Models actius d'aparença Model de punt distribuït Grafs etiquetats

Taula 2-1

*Mètodes d'extracció de trets facials*

La **deformació dels trets facials** es caracteritza pels canvis en la forma i la textura i ens porta als gradients espacials que són bons indicadors de les accions facials i poden ser analitzats tant a la imatge com en el domini freqüencial. Aquests es poden calcular amb el gradient pas alt o pels filtres Gabor [28],[29]. Permeten detectar els contorns per múltiples escales i orientacions. Els filtres Gabor eliminen la majoria de les variacions de les imatges degudes als canvis en la il·luminació. Han demostrat ser bones implementacions per l'anàlisi de les expressions facials i són utilitzats en propostes basades en les imatges [30],[31],[32] així com en combinacions de grafs etiquetats [23],[33],[34].

Les propostes basades en models constitueixen una alternativa a l'extracció de la deformació basada en la imatge. Les propostes del model basat en l'aparença permeten separar de forma fiable diferents fonts d'informació com els canvis en la il·luminació i la deformació del rostre. Lanitis [22] va interpretar imatges de rostres mitjançant la utilització dels models actius d'aparença (AAM – Active Appearance Models) [35],[36]. Els rostres s'analitzen des d'un doble enfocament, utilitzant tant models de contorn com de textura. Els models actius de forma (ASM – Active Shape Models) permeten determinar simultàniament la forma, l'escalat i la posició ajustant a l'objecte d'interès un model de punt distribuït apropiat (PDM – Point Distribution Model).

Un altre tipus de model integral de rostre és el de grafs etiquetats, que estan constituïts per punts característics fiducial poc distribuïts [23],[33],[34]. Els nodes d'aquestes gràfiques característiques consisteixen en *Gabor jets*, on cada component del *jet* és la resposta del filtre a una senyal Gabor estreta d'un punt donat de la imatge. Una graf etiquetada és ajustada a un rostre de test variant la seva escala i posició.

Kobayashi i Hara [37] van utilitzar un model de rostre geomètric que consisteix en 30 punts característics del rostre (FCP – Facial Characteristic Points). Van mesurar la intensitat de la distribució al llarg de 13 FCPs verticals travessant les línies facials amb l'ajuda d'una xarxa neuronal. Finalment, Pantic i Rothkrantz [38] van utilitzar un model 2D basat en punts compostat de vista frontal i lateral.

### 2.3 Classificació de l'expressió facial

La classificació de les característiques s'executa a l'últim pas d'un sistema d'anàlisi de l'expressió facial. Podem distingir entre propostes de classificació espacials o espai temporals:

- Propostes espai temporals

Els *Hidden Markov models (HMM)* són utilitzats habitualment en el camp del reconeixement de la parla, però també són útils per l'anàlisi de l'expressió facial ja que permeten modelar el moviment de les accions facials. Es poden trobar diverses propostes de classificació basada en HMM en la literatura [39][40][41] i són emprades conjuntament amb mètodes d'extracció del moviment en imatges.

Les xarxes neuronals recurrents constitueixen una alternativa als models HMM i es van utilitzar també per la tasca de la classificació de l'expressió facial [42][43]. Un altre camí per tenir en compte l'evolució temporal de l'expressió facial són les denominades plantilles espai temporals d'energia del moviment. En aquests el moviment facial es representa per camps de moviment 2D i es pot utilitzar doncs la distància euclidiana entre dues plantilles per estimar l'expressió facial predominant [25].

- Propostes espacials

Les xarxes neuronals s'utilitzen també sovint per la classificació de l'expressió facial [44][32][11][37][33][45][46]. Aquestes es van aplicar tant directament en imatges de rostres [30][44] com combinades amb mètodes d'extracció i representació dels trets facials com PCA Independent component analysis ICA o els filtres de Gabor [30][31].

Dailey i Cottrell [31] van aplicar tant PCS local i Gabor jets per a l'aplicació del reconeixement de l'expressió facial i van obtenir resultats quantitativament indistingibles per a les dues representacions.

Desafortunadament, les xarxes neuronals són difícils d'entrenar si s'utilitzen per la classificació d'emocions que no són les bàsiques. El problema és el gran nombre de possibles combinacions d'accions facials, s'han identificat a l'espai FACS [47].

Un altre grup d'investigadors han tingut en compte esquemes de codificació de l'expressió facial com MPEG-4 [19][48] o FACS [32][38][49][50][51][52][40][53]. Essa i Pentland [25] van proposar una extensió de FACS denominada FACS+, que consisteix en un conjunt de paràmetres de control que utilitzen observacions basades en la visió.

Barlett [82] va proposar un sistema que integra extracció del moviment entre diferents imatges de forma integral amb PCA.

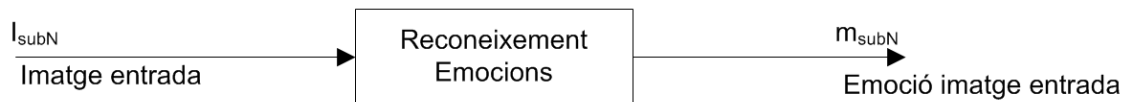


### 3 Esquema general del sistema

#### 3.1 Objectiu del sistema

L'objectiu del sistema és determinar l'emoció representada per una expressió que presenta el rostre de la persona d'una imatge donada.

L'objectiu del sistema es mostra en el següent esquema



*Figura 3-1*

A partir de la imatge,  $I_{subN}$ , el sistema haurà de ser capaç d'interpretar la figura de la imatge d'entrada i proposar l'expressió més significativa de la cara humana que es mostra en ella.

Així, la sortida del sistema serà una estimació de l'expressió de la persona representada a la imatge d'entrada.

#### 3.2 Esquema general

El sistema doncs haurà de ser capaç de dur a terme els següents processos:

- detectar la forma de la cara humana de la imatge i les seves característiques,
- extreure la informació sobre la seva expressió significativa,  $i$
- classificar l'expressió detectada per poder estimar a quina correspon.

La informació sobre la que ens podem basar per classificar l'expressió dependrà dels sistema amb el que extraurem la informació de l'expressió de la imatge.

En el nostre cas el sistema que utilitzem està basat en la teoria dels Models Actius d'Aparença (AAM), implementat per Cootes<sup>1</sup> i Mikkel B. Stegmann<sup>2</sup>.

La informació que ens proporciona el sistema AAM és la localització, forma i textura del contorn dels trets facials i la relació entre ells.

Així, a l'hora de plantejar l'esquema del classificador de l'expressió haurem de tenir en compte aquesta informació de contorn i textura (AAM).

Se'ns plantegen dos camins per resoldre el procés de classificació:

- Classificador a partir de l'entrenament amb les característiques extretes per AAM.
- Classificador a partir de models AAM de cada expressió (que correspondrà amb una emoció coneguda).

<sup>1</sup> T.F. Cootes. Consultar articles [35], [36], [56], [57].

<sup>2</sup> Mikkel B. Stegmann. Consultar l'article [58].

### 3.2.1 Classificació a partir de les característiques extretes

El sistema de classificació basat en la generació de models de les característiques extretes per cada emoció tindrà el següent esquema (utilitzant per exemple xarxes neuronals per aprendre els models de cada emoció):

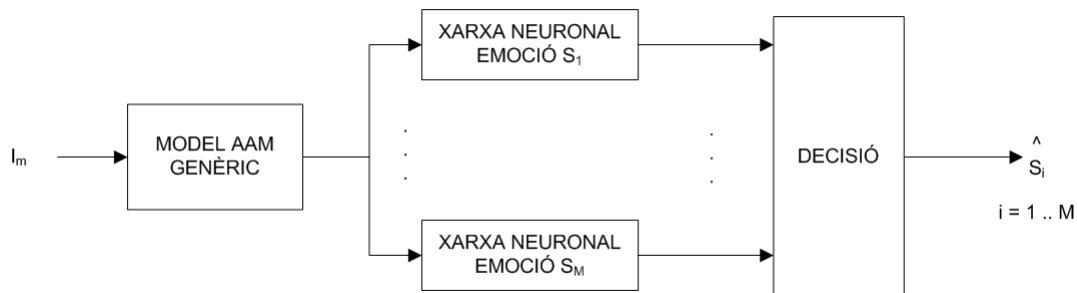


Figura 3-2

El model AAM genèric, serà un model AAM que contindrà la informació de totes les emocions desitjades i que, per tant, podrem suposar que ens donarà la informació de característiques (contorn i textura) correctes per la imatge d'entrada.

Les xarxes neuronals s'entrenaran per reconèixer els patrons en els contorns de les diferents emocions. Cada xarxa modela una emoció diferent.

D'aquesta manera, la idea és que es decideixi per una emoció o una altra en funció de la similitud que el contorn d'aquesta tingui amb els contorns de cada model.

### 3.2.2 Classificació amb models AAM

El sistema de classificació basada en models AAM entrenats amb expressions d'emocions concretes tindrà el següent esquema:

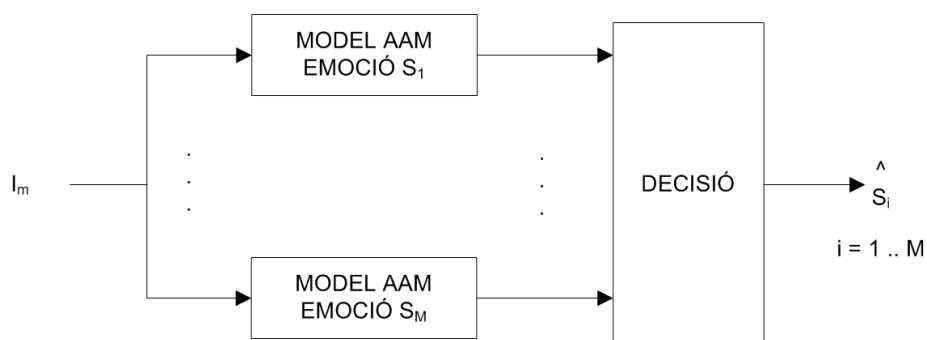


Figura 3-3

Els models AAM s'entrenaran per a contenir la informació de l'expressió d'una emoció concreta.

D'aquesta manera, la idea és que es decideixi per una expressió o una altra en funció de la similitud de la imatge d'entrada amb l'expressió d'entrenament d'un dels diferents models.

És en aquest punt on hem d'escollir un dels camins. La nostra feina s'ha dut a terme basant-nos en la classificació amb models AAM. El sistema de classificació basat en models de característiques va ser implementat per l'Àngel Gutierrez [59].

Estudiem doncs més a fons el sistema basat en la classificació per models AAM.

### 3.3 Sistema basat en la classificació per models AAM

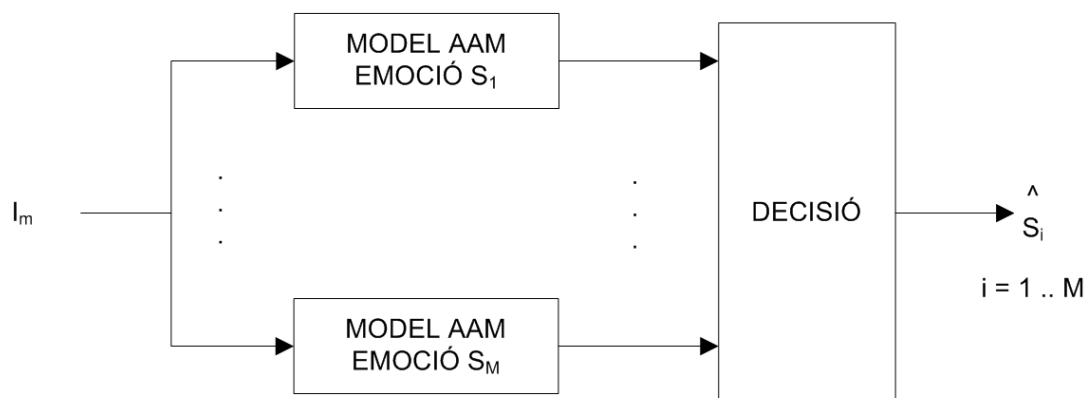


Figura 3-4

L'esquema del sistema de classificació per models AAM recorda a un sistema de decisió amb criteri MAP en l'entorn de transmissió de les senyals digitals.

El sistema de decisió amb criteri MAP es basa en el següent esquema genèric:

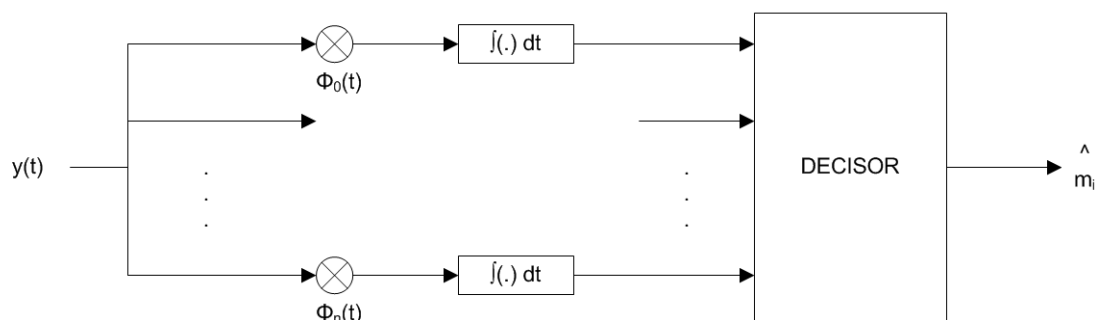


Figura 3-5

Rebuda una senyal \$y(t)\$, el criteri MAP consisteix en decidir quin és el símbol més probable que s'hagi enviat, és a dir:

$$\hat{m}_i \Rightarrow p\left(\frac{m_i}{y(t)}\right) > p\left(\frac{m_j}{y(t)}\right) \text{ per } \forall i, j \text{ tal que } i \neq j$$

Donades  $M$  emocions, representades per  $M$  models AAM (model  $S_1$  .. model  $S_M$ ), processarem una imatge d'entrada  $I_m$  amb els  $M$  models AAM, obtenint així  $M$  imatges síntesi de cada model,  $Im_s^1$  ..  $Im_s^M$ .

El mètode de decisió consistirà en comparar les  $M$  imatges de síntesi,  $Im_s^1$  ..  $Im_s^M$ , amb la imatge d'entrada i trobar la imatge síntesi de major similitud. O el que és el mateix, la síntesi del model amb menor error.

### 3.4 Parts del sistema

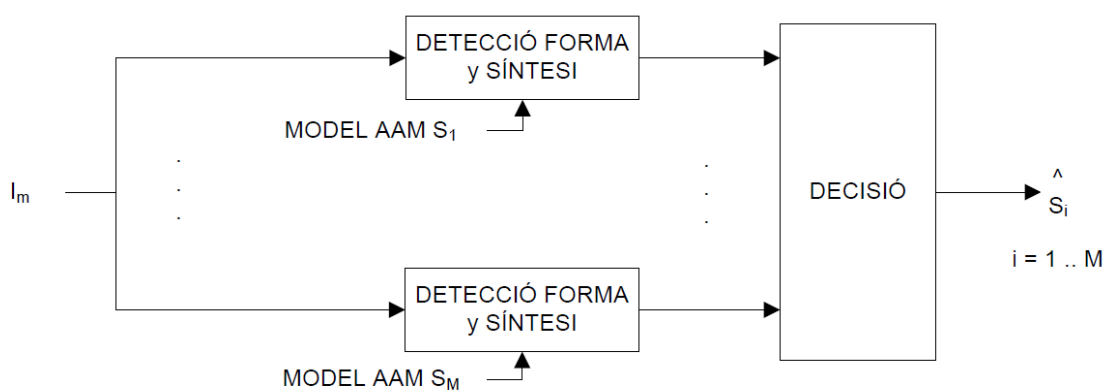


Figura 3-6

El sistema de reconeixement d'emocions té dues parts abans enunciades:

- detecció dels trets facials de la imatge,  $I_m$ , mitjançant la implementació de la teoria AAM per a un model concret,  $S_i$ , i síntesi de la imatge original utilitzant el model
- la decisió de l'emoció de la imatge d'entrada,  $I_m$ , mitjançant la comparació de les diverses formes de les imatges síntesi obtingudes,  $I_s^1$ .

La primera part del sistema que haurem d'estudiar serà l'aplicació basada en la teoria AAM que detecta la forma de la cara humana d'una imatge, computa les seves característiques tals com la col·locació dintre de la imatge, la forma de la cara (punts de control) i la textura (en nivells de gris) i permet sintetitzar una aproximació a la imatge original utilitzant aquest model.

La segona part del sistema que haurem d'analitzar i implementar són els paràmetres i mètode de decisió per poder comparar les imatges síntesi i estimar l'emoció de la imatge d'entrada.

## 4 Teoria AAM

### 4.1 Introducció

El nostre sistema té la necessitat de detectar la forma de la cara humana d'una imatge i les seves característiques i extreure informació sobre la seva emoció. Aquesta necessitat la resolrem mitjançant l'utilització del sistema implementat pel Mikkel B. Stegmann<sup>3</sup>, basat en la teoria dels Models Actius d'Aparença.

Aquest sistema implementat per Mikkel B. Stegmann ens proporciona les eines per definir contorns i textures de diverses formes. Tanmateix, es podran crear unes entitats anomenades models, que contindran la informació de contorns i textures d'un conjunt inicial de formes. Aquests models s'utilitzen tant per caracteritzar conjunts de formes com per generar noves formes amb les mateixes característiques que el conjunt inicial o d'entrenament.

En aquest punt, doncs, aprofundirem en la teoria de Models Actius d'Aparença (AAM). I també, posteriorment, en les eines implementades per Mikkel B. Stegmann.

### 4.2 Teoria Models Estadístics d'Aparença

#### 4.2.1 Models Estadístics de forma (Statistical Shape Models)

En aquest punt descriurem la construcció de models estadístics de forma (*shape*) i aparença (*appearance*) que seran utilitzats per representar formes en imatges.

##### 4.2.1.1 La forma i la seva representació

La forma d'un objecte es representa amb un conjunt de  $n$  punts. Aquests punts poden tenir qualsevol dimensió.

Segons T. F. Cootes, la forma es defineix habitualment com la configuració dels punts que són invariants sota una certa transformació. En el cas de dues o tres dimensions es considera que aquesta transformació és la transformació de Similitud (*Similarity transform*). Així, la forma d'un objecte no canvia si el rotem, escalem o apliquem una translació.

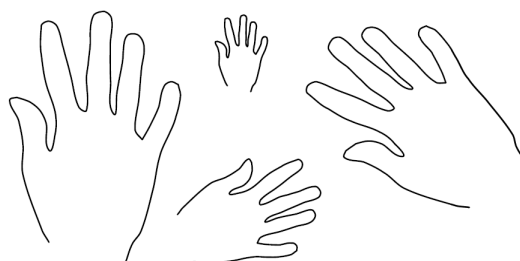


Figura 4-1

Quatre còpies de la mateixa forma, però sota diverses transformacions euclidianes<sup>4</sup>

<sup>3</sup> Stegmann.

<sup>4</sup> M. B. Stegmann, D. Delgado Gomez. A Brief Introduction to Statistical Shape Analysis. pàgina 2.

El desenvolupament teòric dels models de forma es farà en un espai  $d$ -dimensional arbitrari sota una transformació de Similitud (*Similarity transform*)  $T_\theta$  ( $\theta$  representa els paràmetres de la transformació). En dues dimensions els paràmetres de transformació són translació, escalat i rotació.

Els punts que defineixen la forma es denominen punts de referència o punts de control (*landmarks*). La definició que en dóna Mikkel B. Stegmann és la següent:

Punt de referència és un punt de la forma d'un objecte que té correspondència entre les formes d'altres objectes i també amb el mateix objecte sota una certa variació.

Dryden & Mardia<sup>5</sup> divideixen els punts de referència entre 3 grups:

- **Punts de referència anatòmics.** Punts assignats per un expert que són iguals entre organismes d'una mateixa forma biològica.
- **Punts de referència matemàtics.** Punts localitzats en un objecte d'acord amb alguna propietat matemàtica o geomètrica.
- **Pseudo-Punts de referència.** Punts de construcció d'un objecte que estan al contorn o entre altres punt de referència.

Si una forma es defineix per  $n$  punts en  $d$  dimensions, la seva representació serà un vector de  $nd$  elements format per la concatenació de les posicions dels punts de referència individuals de la forma.

Per exemple, en una imatge de 2-D podem representar els  $n$  punts de referència,  $\{(x_i, y_i)\}$ , per una única forma, com el vector de  $2n$  elements:

$$x = (x_1, \dots, x_n, y_1, \dots, y_n)^T \quad (4.1)$$

Tenint  $s$  exemples d'entrenament per construir el model, generarem els seus corresponents  $s$  vectors  $\mathbf{x}_j$  ( $j=1..s$ ). Abans de dur a terme un anàlisi estadístic d'aquests vectors és important que les formes estiguin representades en el mateix espai de coordenades. Voldrem eliminar la variació que pugui ser atribuïda a la transformació global,  $T$ .

#### 4.2.1.2 Alineació de les formes (Shape Alignment)

El mètode per eliminar la variació deguda a la transformació global,  $T$  (posició, escalat i rotació) és establir un espai comú de coordenades (coordenades de referència).

El procés més comú per obtenir aquest espai de coordenades s'anomena *Procrustes analysis*. Aquest mètode alinea cada forma minimitzant la suma de les distàncies de la forma respecte la mitjana. La magnitud a minimitzar es defineix per l'equació:

$$D = \sum |x_i - \bar{x}|^2$$

<sup>5</sup> Dryden & Mardia

Per a formes en 2-D o 3-D una aproximació és centrar cada forma a l'origen, escalar cadascuna de manera que  $|\mathbf{x}|=1$  i aleshores escollir la orientació que ens minimitza  $D$ .

#### 4.2.1.3 Modelant la variació de la forma

Suposem ara que tenim un conjunt  $s$  de vectors  $\mathbf{x}_j$  alineats en un espai de coordenades comú. Aquests vectors formen una distribució a l'espai  $nd$  dimensional en el que estan. Si podem modelar aquesta distribució, podrem generar nous exemples, similars als del conjunt original d'entrenament i podrem analitzar noves formes per decidir fins a quin punt són exemples plausibles.

El nostre objectiu és trobar un model parametritzat del tipus:  $\mathbf{x}=M(\mathbf{b})$ , on  $\mathbf{b}$  és un vector de paràmetres del model. El model també podrà ser utilitzat per generar nous vectors  $\mathbf{x}$ . Si podem modelar la distribució dels paràmetres,  $p(\mathbf{b})$ , podrem generar nous  $\mathbf{x}$  similars als del conjunt d'entrenament jugant amb els valors dels paràmetres.

Per simplificar el problema, primer reduïrem les dimensions de les dades de  $nd$  a alguna cosa més manejable. Una aproximació efectiva és aplicar *Principal Component Analysis* (PCA) a les dades.

Les dades formen un núvol de punts en un espai de  $nd$ -D. PCA troba els eixos principals d'aquest núvol, i ens permet aproximar qualsevol punt original utilitzant un model amb menys de  $nd$  paràmetres. L'aproximació es la següent:

1-calquem la mitjana de la següent forma:

$$\bar{x} = \frac{1}{s} \sum_{i=1}^s x_i \quad (4.2)$$

2-calquem la covariança de les dades:

$$S = \frac{1}{s-1} \sum_{i=1}^s (x_i - \bar{x})(x_i - \bar{x})^T \quad (4.3)$$

3-calquem els autovectors,  $\phi_i$  i els corresponents autovalors  $\lambda_i$  de  $\mathbf{S}$  (ordenats de manera que  $\lambda_i \geq \lambda_{i+1}$ ).

4-Si  $\Phi$  conté els  $t$  autovectors corresponents als autovalors més grans, aleshores podem aproximar qualsevol forma del conjunt d'entrenament,  $\mathbf{x}$ , utilitzant

$$x \approx \bar{x} + \Phi b \quad (4.4)$$

on  $\Phi = (\Phi_1 | \Phi_2 | \dots | \Phi_t)$  i  $\mathbf{b}$  es un vector  $t$  dimensional donat per

$$b = \Phi^T (x - \bar{x}) \quad (4.5)$$

El vector  $\mathbf{b}$  defineix els paràmetres d'un model deformable. Variant els elements de  $\mathbf{b}$  podem variar la forma,  $\mathbf{x}$  utilitzant la Equació 4.4.

La variança en el conjunt d'entrenament per al paràmetre  $i$ -èssim,  $b_i$ , ve donada per  $\lambda_i$ . Aplicant límits de  $\pm 3\sqrt{\lambda_i}$  al paràmetre  $b_i$ , ens assegurem que la forma generada serà similar a les del conjunt d'entrenament.

El nombre d'autovectors a retenir  $t$ , es poden escollir de manera que el model representi una porció del 98% de la variació de les dades, o de manera que els termes residuals puguin ser considerats soroll.

Per escollir  $t$  de manera que representi el 98% de la variació de les dades seguirem el següent càlcul.

La variació total de les dades d'entrenament és la suma de tots els autovalors,  $V_T = \sum \lambda_i$ .

Podem escollir doncs els  $t$  autovalors majors de manera que

$$\sum_{i=1}^t \lambda_i \geq f_v V_T \quad (4.6)$$

on  $f_v$  defineix la proporció del total de la variació que es vol obtenir.

#### 4.2.1.4 Generar formes plausibles

Si volem utilitzar el model  $x = \bar{x} + \Phi b$  per generar exemples similars als del conjunt d'entrenament, haurem d'escollir els paràmetres,  $\mathbf{b}$ , a partir d'una distribució obtinguda del conjunt d'entrenament. D'aquesta manera hem d'estimar la distribució,  $p(\mathbf{b})$ , del conjunt d'entrenament.

Definirem un conjunt de paràmetres com a "plausibles" si  $p(\mathbf{b}) \geq p_t$ , on  $p_t$  és un llindar adequat de la p.d.f (funció de densitat de probabilitat).  $p_t$  s'escollix de manera que hi hagi un cert percentatge del conjunt d'entrenament que passi el llindar (per exemple 98%).

Si assumim  $b_i$  com independents i gaussianes, aleshores

$$\log p(b) = -0.5 \sum_{i=1}^t \frac{b_i^2}{\lambda_i} + const$$

Per trobar valors plausibles de  $\mathbf{b}$  podem aplicar límits a cada element,  $b_i$  (per exemple,  $|b_i| \leq 3\sqrt{\lambda_i}$ ) o podem fixar  $\mathbf{b}$  per a que sigui una hiper el·lipse.

#### 4.2.1.5 Trobar la forma plausible més propera

Quan ajustem un model, a un nou conjunt de punts d'un objecte, tenim el problema de trobar la forma plausible del model més propera,  $\mathbf{x}$ , a la forma objectiu,  $\mathbf{x}'$ . La primera aproximació és projectar-la dins de l'espai dels paràmetres, resultant els seus paràmetres  $b' = \Phi^T (x' - \bar{x})$ .

Definim un conjunt de paràmetres plausibles si  $p(\mathbf{b}) \geq p_t$ . Si  $p(\mathbf{b}) < p_t$  desitjarem moure  $\mathbf{x}$  al punt més proper que es pugui considerar plausible.



4.2.1.6 Ajustar el model a nous punts

Una síntesi d'un model en una imatge ve descrita pels paràmetres de forma,  $\mathbf{b}$ , combinat amb la transformació des de l'espai de coordenades del model a l'espai de coordenades de la imatge. Típicament aquesta transformació serà la transformació de Similitud que defineix la posició,  $(X_t, Y_t)$ , orientació,  $\theta$ , i escala,  $s$ , de la síntesi del model a la imatge.

Les posicions dels punts del model a la imatge,  $\mathbf{x}$ , venen donades per

$$\mathbf{x} = T_{X_t, Y_t, s, \theta}(\bar{\mathbf{x}} + \Phi \mathbf{b}) \quad (4.11)$$

On la funció  $T_{X_t, Y_t, s, \theta}$  representa la rotació per  $\theta$ , l'escalat per  $s$  i la translació per  $(X_t, Y_t)$ . Per exemple, si ho apliquem a un punt  $(xy)$ , obtindrem

$$T_{X_t, Y_t, s, \theta} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} X_t \\ Y_t \end{pmatrix} + \begin{pmatrix} s \cos \theta & s \sin \theta \\ -s \sin \theta & s \cos \theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \quad (4.12)$$

Suposem que volem trobar els paràmetres de col·locació (*pose*) i forma que millor encaixin una síntesi del model  $\mathbf{x}$  amb un nou conjunt de punts d'una imatge,  $\mathbf{Y}$ . Minimitzar la suma quadràtica de la distància entre la síntesi i els punts de la imatge és equivalent a minimitzar l'expressió

$$\left| \mathbf{Y} - T_{X_t, Y_t, s, \theta}(\bar{\mathbf{x}} + \Phi \mathbf{b}) \right|^2 \quad (4.13)$$

Una simple aproximació iterativa per aconseguir-ho és la següent:

- 1 Inicialitzar els paràmetres de forma,  $\mathbf{b}$ , a zero.
- 2 Generar la síntesi del model  $\mathbf{x} = \bar{\mathbf{x}} + \Phi \mathbf{b}$
- 3 Trobar els paràmetres de col·locació  $(X_t, Y_t, s, \theta)$  que millor ajustin  $\mathbf{x}$  a  $\mathbf{Y}$
- 4 Invertir els paràmetres de col·locació i utilitzar-los per projectar  $\mathbf{Y}$  a l'espai de coordenades del model:

$$\mathbf{y} = T_{X_t, Y_t, s, \theta}^{-1}(\mathbf{Y}) \quad (4.14)$$

5. Projecta  $\mathbf{y}$  al plà tangent  $\bar{\mathbf{x}}$  escalant per  $\frac{1}{(\mathbf{y} \cdot \bar{\mathbf{x}})}$
- 6 Actualitza els paràmetres del model que s'ajustin a  $\mathbf{y}$

$$\mathbf{b} = \Phi^T (\mathbf{y} - \bar{\mathbf{x}}) \quad (4.15)$$

- 7 Aplicar les restriccions a  $\mathbf{b}$  segons les Equacions 4.6, 4.7.
- 8 Si no convergeix, tornar al pas 2

S'arriba a la convergència si no hi ha canvis significatius als paràmetres de col·locació i forma després d'una iteració.

4.2.1.7 Estimació p(forma)

Donada una configuració de punts,  $\mathbf{x}$ , voldrem ser capaços de decidir fins a quin punt  $\mathbf{x}$  és un exemple plausible de la classe de formes descrites pel nostre conjunt d'entrenament.

El conjunt d'entrenament original, una vegada alineat, es pot considerar com el conjunt de mostres d'una funció de densitat de probabilitat,  $p(\mathbf{x})$ , que hem d'estimar.

Qualsevol forma es pot aproximar pel punt del subespai més proper definit pels autovectors,  $\Phi$ . Un punt en aquest subespai es defineix per un vector de paràmetres de forma,  $\mathbf{b}$ .

Definim  $dx = x - \bar{x}$ . Així, la millor aproximació ve donada per  $x' = \bar{x} + \Phi b$  on  $b = \Phi^T dx$ .

L'error residual és, doncs,  $r = dx - \Phi b$

El quadrat d'aquesta magnitud és

$$\begin{aligned} |r|^2 &= r^T r = dx^T dx - 2 dx^T \Phi b + b^T \Phi^T \Phi b \\ |r|^2 &= |dx|^2 - |b|^2 \end{aligned}$$

Aplicant un PCA es generen dos subespais (definitos per  $\Phi$  i els seu nucli (null-space)) que divideixen el vector de forma en dos components ortogonals amb coeficients descrits per elements de  $\mathbf{b}$  i  $\mathbf{r}$ , que assumim independents.

D'aquesta manera,

$$p(x) = p(r) \cdot p(b) \quad (4.17)$$

$$\log p(x) = \log p(r) + \log p(b) \quad (4.18)$$

Si assumim que cada element de  $\mathbf{r}$  es independent i amb una distribució gaussiana de varianza  $\sigma_r^2$ ,

$$p(r) \propto \exp(-0.5|r|^2 / \sigma_r^2) \quad (4.19)$$

$$\log p(r) = -0.5|r|^2 / \sigma_r^2 + const$$

La distribució dels paràmetres,  $p(\mathbf{b})$ , es pot estimar com es descriu al punt **4.2.4. Generar formes plausibles**.

Donat això, podem estimar la funció densitat de probabilitat d'una nova forma,  $\mathbf{x}$ , utilitzant

$$\log p(x) = \log p(b) - 0.5(|dx|^2 - |b|^2) / \sigma_r^2 + const \quad (4.20)$$

#### 4.2.2 Models Estadístics d'aparença

Per sintetitzar una imatge completa d'un objecte o estructura, hem de modelar tant la seva forma com la seva textura (el patró de intensitat o color al llarg de la regió de l'objecte). A continuació descriurem com els models estadístics poden representar la variació de la forma, la variació de la textura i la correlació entre elles.

Els models es generen mitjançant la combinació d'un model de variació de forma amb un model de variacions de textura en un espai normalitzat a la forma.

##### 4.2.2.1 Models Estadístics de textura

Per construir un model estadístic de textura (intensitat o color d'una zona de la imatge) deformarem cada imatge del conjunt d'entrenament de manera que els seus punts de control o *punts de referència* encaixin amb els de la forma mitjana. Això extraurà la variació espúria de la textura deguda a les diferències entre formes.

Mostregem, ara, la informació de la intensitat a la regió de la imatge (normalitzada a la forma) limitada per la forma mitjana i generarem el vector de textura  $\mathbf{g}_{im}$ .

Per minimitzar l'efecte de la variació de la il·luminació global, normalitzarem les mostres de cada imatge aplicant un escalat,  $\alpha$ , i un offset,  $\beta$ ,

$$g = (g_{im} - \beta) / \alpha \quad (5.1)$$

Els valors  $\alpha$  i  $\beta$  els escollirem per tal que el vector  $\mathbf{g}$  s'apropi més a la mitjana normalitzada. Definim  $\bar{g}$  com la mitjana de les dades normalitzades, escalades i amb un offset de manera que la suma dels seus elements sigui zero i la variància dels seus elements sigui la unitat. Els valors  $\alpha$  i  $\beta$  necessaris per normalitzar  $\mathbf{g}_{im}$  venen donats per

$$\alpha = g_{im} \bar{g} \quad , \quad \beta = (g_{im} \cdot 1) / n \quad (5.2)$$

on  $n$  és el nombre d'elements dels vectors.

Aplicant PCA a les dades normalitzades obtindrem el model lineal:

$$g = \bar{g} + P_g b_g \quad (5.3)$$

on  $\bar{g}$  és el vector mitjana dels nivells de gris normalitzat,  $\mathbf{P}_g$  és un conjunt de modes de variació ortogonals i  $\mathbf{b}_g$  és un conjunt de paràmetres de nivells de gris.

La textura en l'espai de la imatge es pot generar a partir dels paràmetres de textura,  $\mathbf{b}_g$ , i els paràmetres de normalització  $\alpha$ ,  $\beta$ . Per linealitat representarem aquests últims en el vector  $u = (\alpha - 1, \beta)^T$ . La textura en l'espai de la imatge ve donada, doncs, per

$$g_{im} = T_u(\bar{g} + P_g b_g) = (1 + u_1)(\bar{g} + P_g b_g) + u_2 1 \quad (5.4)$$

4.2.2.2 Models Combinats d'Aparença

La forma i la textura de qualsevol imatge pot, així, resumir-se pels vectors de paràmetres  $\mathbf{b}_s$  i  $\mathbf{b}_g$ . En el moment en que podem trobar correlacions entre les variacions de forma i textura, aplicarem PCA a les dades de la següent manera.

Per cada imatge d'entrenament generarem el vector concatenat

$$b = \begin{pmatrix} W_s b_s \\ b_g \end{pmatrix} = \begin{pmatrix} W_s P_s^T (x - \bar{x}) \\ P_g^T (g - \bar{g}) \end{pmatrix} \quad (5.5)$$

on  $W_s$  és la matriu diagonal de pesos per cada paràmetre de la forma, degut a la diferència en les unitats entre els models de forma i textura.

Apliquem PCA a aquests vectors, obtenim el següent model

$$b = P_c c \quad (5.6)$$

on  $P_c$  son els autovectors i  $c$  es el vector de paràmetres d'aparença que controlen tant la forma com els nivells de gris del model.

Notem que la natura lineal del model ens permet expressar la forma i els nivells de gris directament en funció de  $c$

$$x = \bar{x} + P_s W_s^{-1} P_{cs} c \quad , \quad g = \bar{g} + P_g P_{cg} c \quad (5.7)$$

on

$$P_c = \begin{pmatrix} P_{cs} \\ P_{cg} \end{pmatrix} \quad (5.8)$$

Si volem simplificar-ho mes

$$\begin{aligned} x &= \bar{x} + Q_s c \\ g &= \bar{g} + Q_g c \end{aligned} \quad (5.9)$$

on

$$\begin{aligned} Q_s &= P_s W_s^{-1} P_{cs} \\ Q_g &= P_g P_{cg} \end{aligned} \quad (5.10)$$

Una imatge qualsevol donada que no pertanyi al conjunt d'entrenament pot ser sintetitzada per uns paràmetres  $c$  donats, generant la imatge de nivells de gris independent de la forma a partir del vector  $g$  i deformant-la utilitzant els punts de control o punts de referència descrits per  $x$ .

#### 4.2.2.3 Aproximant un nou exemple

Donada una imatge, etiquetada amb un conjunt de punts de control o *punts de referència*, podem generar una aproximació amb el model. Seguirem els passos abans explicats per obtenir  $\mathbf{b}$ , combinant la forma i els paràmetres de nivell de gris que s'ajusten a l'exemple.

Com  $\mathbf{P}_c$  és ortogonal, els paràmetres del model combinat d'aparença,  $\mathbf{c}$  venen donats per

$$\mathbf{c} = \mathbf{P}_c^T \mathbf{b} \quad (5.11)$$

La reconstrucció completa ve donada aplicant l'Equació 5.7, invertint la normalització de nivell de gris, aplicant la col·locació apropiada als punts i projectant el vector de nivell de gris a la imatge.



Figura 4-2

*Representació d'un model combinat (dreta) d'una imatge d'un rostre prèviament desconeguda (esquerra)*

### 4.2.3 Interpretació d'imatges amb Models

Per interpretar una imatge utilitzant un model, haurem de trobar el conjunt de paràmetres amb els que millor ajustem el model a la imatge. Aquest conjunt de paràmetres defineixen la forma, posició i la possible aparença de l'objecte desitjat, i poden ser utilitzats per posteriors processats, així com per mesurar o classificar l'objecte.

Hi ha diverses aproximacions que es poden utilitzar per ajustar una síntesi del model a una imatge, però totes es basen en la optimització d'una funció de cost. Per a un conjunt de paràmetres de model,  $\mathbf{c}$ , podem generar una síntesi del model que projectarem a la imatge. Podem comparar aquesta síntesi amb la imatge, per trobar una funció d'ajust  $F(\mathbf{c})$ . El conjunt de paràmetres que millor interpretin l'objecte de la imatge seran els que optimitzaran aquesta funció.

Així, en teoria el que hem de fer és escollir una funció d'ajust apropiada, i utilitzar un mètode d'optimització per trobar el mínim. El mínim només es defineix per l'elecció de la funció, el model i la imatge, i és independent del mètode d'optimització escollit. No obstant això, a la pràctica, s'ha d'escollir la funció d'optimització amb cura, que sigui ràpida i robusta, i un mètode d'optimització que convergeixi.

#### 4.2.3.1 Escollir una funció d'ajust

Idealment voldrem escollir una funció d'ajust que representi la probabilitat de que els paràmetres del model representin l'objecte de la imatge,  $P(\mathbf{c}|\mathbf{I})$  (on  $\mathbf{I}$  representa la imatge). Escollirem, doncs, els paràmetres que maximitzin aquesta probabilitat.

En el cas dels models de forma descrits anteriorment, els paràmetres que podrem variar són els paràmetres de forma,  $\mathbf{b}$ , i els paràmetres de col·locació  $X_t, Y_t, s, \theta$ . Per als models d'aparença podrem variar els paràmetres d'aparença,  $\mathbf{c}$  i els paràmetres de col·locació.

La qualitat de l'ajust del model d'aparença es pot calcular mesurant la diferència entre la imatge i la imatge sintètica generada a partir del model.

La forma de calcular-ho només per als models de forma és més difícil de determinar. Si suposem que el model de forma representa la delimitació de l'objecte, una mesura útil és la distància entre els punts donats pel model i el punt de delimitació més proper de l'objecte de la imatge.

Si les posicions dels punts donats pel model venen donades al vector  $\mathbf{X}$ , i cada punt més proper dels límits de l'objecte són  $\mathbf{X}'$ , la mesura de l'error és

$$F(\mathbf{b}, X_t, Y_t, s, \theta) = |\mathbf{X}' - \mathbf{X}|^2 \quad (6.1)$$

#### 4.2.3.2 Optimitzar l'ajust del model

No tenint coneixement sobre on està posicionat l'objecte dins de la imatge, trobar els paràmetres que optimitzen l'ajust és un problema difícil. Això ho podem abordar amb tècniques generals d'optimització global, com *Genetic Algorithms* o *Simulated Annealing*.

Si, no obstant això, tenim una aproximació inicial a la solució correcta (sabem on estarà l'objecte dins de la imatge, degut a un pre-processament), podem utilitzar tècniques d'optimització local com el mètode *Powell* o *Simplex*.

Desenvoluparem dos algorismes que porten a una cerca directa a l'espai dels paràmetres – Model Actiu de Forma (*Active Shape Models*) i Model Actiu d'Aparença (*Active Appearance Model*)

#### 4.2.4 Model actiu de forma (Active Shape Models)

Donada una aproximació inicial vasta, la síntesi d'un model es pot ajustar a una imatge. Escollint un conjunt de paràmetres de forma,  $\mathbf{b}$ , per al model definim la forma de l'objecte en un espai de coordenades centrat en l'objecte. Podrem crear una síntesi del model,  $\mathbf{X}$ , en l'espai de la imatge definint la posició, orientació i escala, utilitzant l'equació 4.11.

Una aproximació iterativa per millorar l'ajust de la síntesi,  $\mathbf{X}$ , a una imatge és la següent:

- 1 Examinar una regió de la imatge al voltant de cada punt  $X_i$  per trobar el següent candidat més proper o que millor s'ajusta (match) a la forma de la imatge  $X_i'$
- 2 Actualitzar els paràmetres ( $X_t, Y_t, s, \theta, \mathbf{b}$ ) que millor s'ajustin als nous punts trobats  $\mathbf{X}$
- 3 Repetir fins a convergir

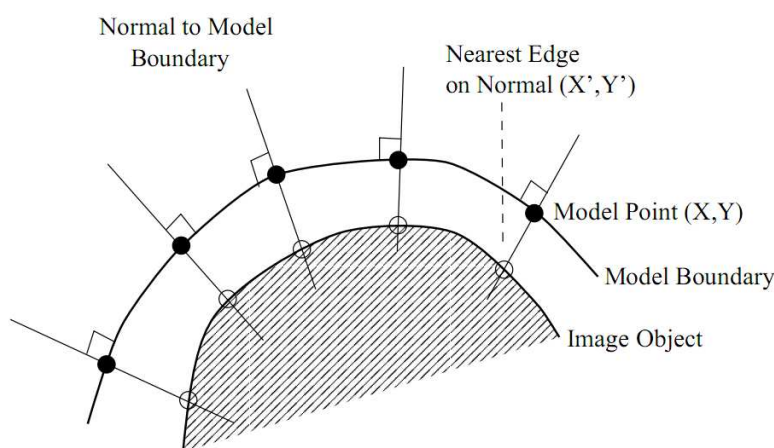


Figura 4-3

A la pràctica mirem al llarg de les línies normals als contorns de la síntesi del model en cada punt de control (Figura 4-3). Si esperem que el contorn de la síntesi del model correspongui amb un contorn, podem situar simplement el contorn més fort al llarg del perfil. La posició d'això ens dóna una nova ubicació per al punt del model.

#### 4.2.5 Model Actiu d'aparença (Active Appearance Models)

L'algoritme de cerca del Model Actiu de Forma (*Active Shape Model*) ens permet situar punts en una nova imatge, fent ús de les restriccions dels models de forma. Una desavantatge és que només utilitza les restriccions de la forma, i no utilitza tota la informació disponible – la textura de l'objecte. Això es pot fer utilitzant el Model d'Aparença.

Descriurem a continuació un algoritme que ens permet trobar els paràmetres d'un model els quals generin una imatge sintetitzada el més propera possible a la imatge objectiu, assumint que coneixem un punt d'inici aproximat.

#### 4.2.5.1 Visió general de la cerca amb AAM

Considerarem la cerca amb AAM com un problema d'optimització on voldrem minimitzar la diferència entre una imatge i la seva síntesi del model d'aparença. Un vector diferència  $\delta\mathbf{I}$  es pot definir com:

$$\delta\mathbf{I} = \mathbf{I}_i - \mathbf{I}_m \quad (8.1)$$

on  $\mathbf{I}_i$  es el vector dels valors de nivells de gris de la imatge, i  $\mathbf{I}_m$ , és el vector de valors de nivells de gris per als paràmetres actuals del model.

Per localitzar el millor ajust entre el model i la imatge, voldrem minimitzar el mòdul del vector diferència,  $\Delta = |\delta\mathbf{I}|^2$ , variant els paràmetres del model,  $\mathbf{c}$ . Donat que els models d'aparença poden tenir diversos paràmetres, inicialment sembla que ens planteja un problema d'optimització multi dimensional difícil. Tenint un coneixement a priori de com ajustar els paràmetres del model durant la cerca a la imatge, arribem a un algoritme eficient en temps computacional. En particular, el model espacial a  $\delta\mathbf{I}$ , codifica informació sobre com s'haurien de canviar els paràmetres del model per aconseguir un millor ajust. Adoptant aquesta aproximació hi ha dues parts del problema: conèixer la relació entre  $\delta\mathbf{I}$  i l'error dels paràmetres del model  $\delta\mathbf{c}$  i utilitzar aquest coneixement en un algoritme iteratiu per minimitzar  $\Delta$ .

#### 4.2.5.2 Aprenent a Corregir els Paràmetres del Model

El model d'aparença té paràmetres,  $\mathbf{c}$ , que controlen la forma i la textura (a l'espai del model) segons

$$\begin{aligned} x &= \bar{x} + \mathbf{Q}_s \mathbf{c} \\ g &= \bar{g} + \mathbf{Q}_g \mathbf{c} \end{aligned} \quad (8.2)$$

on  $\bar{x}$  es la forma mitjana,  $\bar{g}$  es la textura mitjana a la zona de la forma mitjana i  $\mathbf{Q}_s$ ,  $\mathbf{Q}_g$  són matrius que descriuen els modes de variació derivats del conjunt d'entrenament.

Una forma a l'espai de la imatge,  $\mathbf{X}$ , es pot generar aplicant una transformació adequada als punts,  $\mathbf{x}$ :  $\mathbf{X} = S_t(\mathbf{x})$ . Típicament  $S_t$  serà una transformació de Similitud descrit per un escalat,  $s$ , una rotació en el pla de la imatge,  $\theta$ , i una translació  $(t_x, t_y)$ . Per linealitat representem l'escalat i la rotació com  $(s_x, s_y)$  on  $s_x = (s \cos \theta - 1)$ ,  $s_y = s \sin \theta$ . El vector del paràmetre de col·locació,  $\mathbf{t} = (s_x, s_y, t_x, t_y)^T$  és zero, doncs, per la transformació identitat i així  $S_{t+\delta\mathbf{t}}(x) \approx S_t(S_{\delta\mathbf{t}}(x))$ .

La textura a l'espai de la imatge es genera aplicant un escalat i un offset a les intensitats,  $\mathbf{g}_{im} = T_u(\mathbf{g}) = (u_1 + 1)\mathbf{g}_{im} + u_2\mathbf{1}$ , on  $\mathbf{u}$  és el vector de paràmetres de transformació, definit de manera que  $\mathbf{u} = \mathbf{0}$  és la transformació identitat, i així  $T_{u+\delta\mathbf{u}}(g) \approx T_u(T_{\delta\mathbf{u}}(g))$

Els paràmetres del model d'aparença,  $\mathbf{c}$ , i els paràmetres de transformació de la forma,  $\mathbf{t}$ , defineixen la posició dels punts del model a l'espai de la imatge,  $\mathbf{X}$ , que representa la forma a la zona de la imatge representada pel model. Durant l'ajust mostrejarem els píxels dins d'aquesta regió de la imatge,  $\mathbf{g}_{im}$ , i els projectarem dins de l'espai del model de textura,  $\mathbf{g}_s = T^{-1}(\mathbf{g}_{im})$ . El model de textura actual ve donat per  $g_m = \bar{g} + \mathbf{Q}_g \mathbf{c}$ . La diferència actual entre el model i la imatge (mesurat en l'espai de textura normalitzat) és



$$r(p) = g_s - g_m \quad (8.3)$$

on  $\mathbf{p}$  són els paràmetres del model,  $\mathbf{p}^T = (\mathbf{c}^T \mid \mathbf{t}^T \mid \mathbf{u}^T)$ .

Una simple mesura escalar de la diferència és la suma dels quadrats dels elements de  $\mathbf{r}$ .  $E(\mathbf{p}) = \mathbf{r}^T \mathbf{r}$ .

El polinomi de Taylor de primer ordre per a l'expressió 8.3 ens dona

$$r(p + \delta p) = r(p) + \frac{\partial r}{\partial p} \delta p \quad (8.4)$$

On l'element  $ij$ -èssim de la matriu  $\frac{\partial r}{\partial p}$  és  $\frac{dr_i}{dp_j}$

Suposem que durant l'ajust el residu és  $\mathbf{r}$ . Voldrem escollir  $\delta \mathbf{p}$  que minimitzi  $|\mathbf{r}(\mathbf{p} + \delta \mathbf{p})|^2$ . Igualant l'expressió 8.4 a zero obtenim la solució RMS,

$$\delta p = -Rr(p) \text{ on } R = \left( \frac{\partial r^T}{\partial p} \frac{\partial r}{\partial p} \right)^{-1} \frac{\partial r^T}{\partial p} \quad (8.5)$$

En un esquema d'optimització estàndard seria necessari recalculer  $\frac{\partial r}{\partial p}$  a cada iteració,

una operació costosa. No obstant això, assumirem que en ser calculat en un espai de referència normalitzat, es pot considerar aproximadament fix. Així, ho podem estimar

una vegada a partir del nostre conjunt d'entrenament. Estimarem  $\frac{\partial r}{\partial p}$  per diferenciació

numèrica, desplaçant sistemàticament cada paràmetre del seu valor òptim i calculant la mitjana per a tot el conjunt d'entrenament.

#### 4.2.5.3 Refinament del model iteratiu (Iterative Model Refinement)

Donat un mètode per predir la correcció que es necessita introduir al paràmetres del model, podrem construir un mètode iteratiu per solucionar el problema d'optimització.

Donat l'estimació actual dels paràmetres del model,  $\mathbf{c}_0$ , i la mostra de la imatge normalitzada de l'estimació actual,  $\mathbf{g}_s$ , una iteració del procediment és el següent:

Avaluar el vector de l'error  $\delta \mathbf{g}_0 = \mathbf{g}_s - \mathbf{g}_m$

Avaluar l'error actual  $E_0 = |\delta \mathbf{g}_0|^2$

Calcular la predicció del desplaçament,  $\delta \mathbf{c} = A \delta \mathbf{g}_0$

Posar  $k=1$

Definim  $\mathbf{c}_1 = \mathbf{c}_0 - k \delta \mathbf{c}$

Mostregem la imatge amb aquesta nova predicció, i calculem el nou vector d'error,  $\delta \mathbf{g}_1$

Si  $|\delta \mathbf{g}_1|^2 < E_0$ , acceptem la nova estimació,  $\mathbf{c}_1$ ,

Si no provar amb  $k=1.5$ ,  $k=0.5$ ,  $k=0.25$ , etc.

Aquest procediment es repeteix fins que no es fa cap millora a l'error, i es declara la convergència.

### 4.3 Interfície AAM

#### 4.3.1 AAM-API

Mikkel B. Stegmann ha desenvolupat una Interfície de Programació d'Aplicacions (API - Application Programming Interface) on s'implementen els conceptes, estructures de dades i operacions de la teoria de Models Actius d'Aparença (AAM).

Per a una utilització més flexible, Mikkel B. Stegmann ha implementat dos tipus d'interfícies: una interfície en mode consola MS-DOS, denominada *aamc*, i una interfície de tipus finestra multi document denominada *AAMLab*, imatges de les quals podeu veure un exemple a continuació.

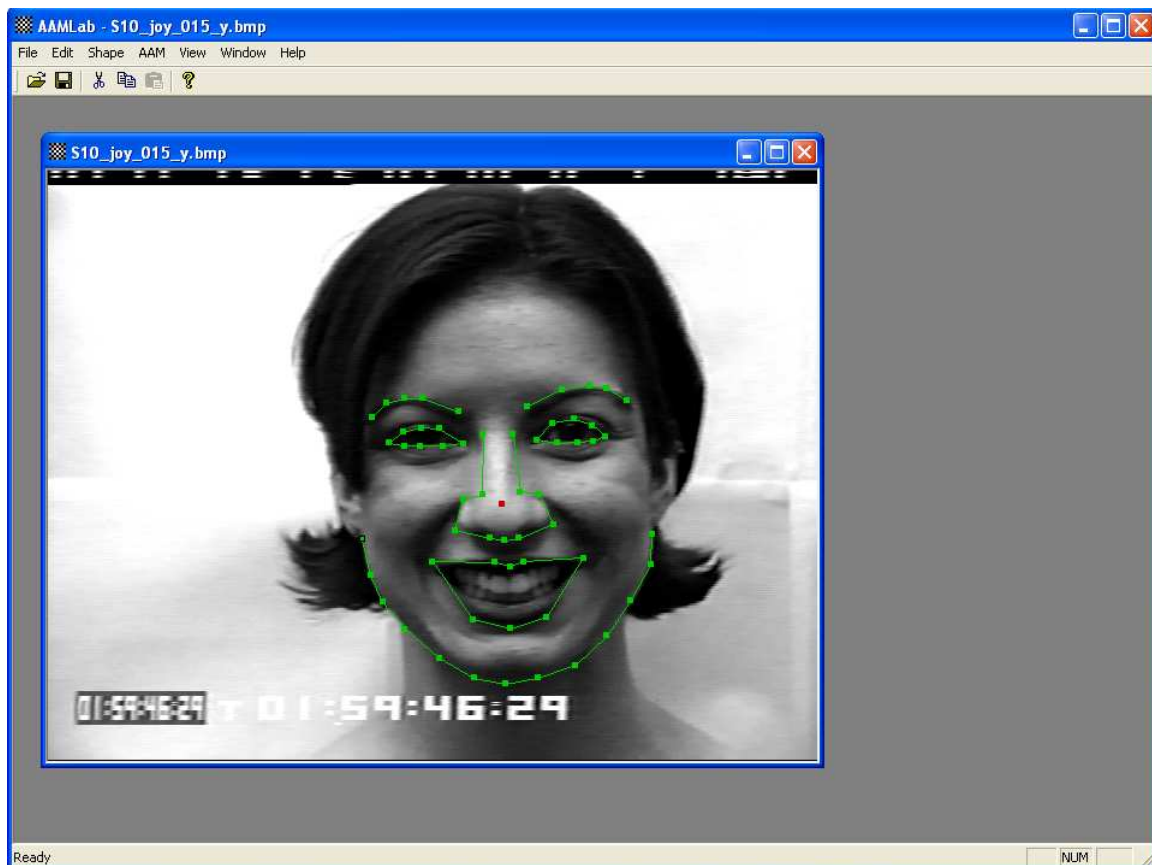


Figura 4-4  
Aplicació AAMLab

```

c:\ Selecció C:\WINDOWS\system32\cmd.exe
*****
This small AAMC demo builds an AAM and produce the
following files:

    model.txt           Model information
    model.amf           Binary model data
    modelvar.bmp        Image of the pixel variation
    registration.avi    Registration movie - i.e.
                        all training examples warped
                        to the mean shape

    shape0i.avi         i-th shape PCA mode (+/- 3 std)
    texture0i.avi       i-th texture PCA mode (+/- 3 std)
    combined0i.avi      i-th combined PCA mode (+/- 3 std)

AAM-API site: http://www.imm.dtu.dk/~aam/
*****
Presione una tecla para continuar . . .
Building Active Appearance Model.
Reading images and annotations...
Doing PCA on the shape data...
Building texture vectors...
Normalizing texture vectors...
Calculating texture variances...
Writing variance image 'modelvar.bmp'...
Doing PCA on the texture data...
Doing combined PCA...
Building gradient matrices (method==1)...
Info: Training set subsampling = 1 (3 shapes used)
Experiment 8 of 24 done (c)...
Experiment 16 of 24 done (c)...
Experiment 24 of 24 done (c)...
Experiment 16 of 48 done (pose)...
Experiment 32 of 48 done (pose)...
Experiment 48 of 48 done (pose)...
Done...
Time spent: 00:00 (0.3 secs)
Generating movie type 'all'...
Evaluating AAM...
Final COG=( 136, 110 ), E_min=      0.37, 918 configurations searched
#####
#
#   AAM Evaluation - written: Wednesday July 30 - 2008 [14:02]
#
#####

    Exp  Pt.pt.  Pt.crv.  Overlap  Iter   Maha   Error   Time
    0    2.4152  1.0034  0.9354   1    0.6868  0.3740  4.2815
    
```

Figura 4-5  
 Aplicació de consola aamc

4.3.1.1 Motor AAM-API

El motor d'AAM-API consta de diversos objectes que implementen la teoria AAM. Com a nucli trobem les següents classes:

<b>CAAMShape</b>	<b>Shape container.</b> This class act as a container for a shape. Essentially it's just a set of 2D points stored in a vector in the format xxxyyy.
<b>CAAMShapeCollection</b>	<b>Shape collection container and shape-aligner.</b> This class act as a container for a set of shapes. Secondary it can align the set of shapes to a normalised reference frame with respect to position, scale and orientation.
<b>CAAMReferenceFrame</b>	<b>The geometrical reference frame (or shape-free frame).</b> This class defined the geomtrical reference frame of an AAM. Hence, it is the spatial layout where all texture sampling takes place. The main objective of this class is to provide fast conversions from a shape-free image to a texture vector and vice versa.
<b>CAAMModel</b>	<b>The core Active Apperance Model object.</b> The core Active Apperance Model object that hold all eigenmodels, prediction matrices etc. Build by a CAAMBuilder.
<b>CAAMBuilder</b>	<b>Factory object that produces CAAMModel objects.</b> Factory object that produces CAAMModel objects. Main tasks are the estimation of parameter update matrices and verbose dumping of model information. Most other tasks are simple calls into CAAMModel, CAAMShape etc.
<b>CAAMDeformPCA</b>	<b>Performs Principal Component Analysis on a set of data vectors.</b> Performs Principal Component Analysis on a set of data vectors. The PCA basis can then be used for e.g. shape deformation.
<b>CAAMTest</b>	<b>Container for all sorts of test functions.</b> Container for all sorts of test functions. In this way a history of all mock-up test functions developed during debugging and testing are kept.
<b>CAAMEvaluationResults</b>	<b>Container to store evaluation results in.</b> Container to store evaluation results in. Further, it can print out the results along with some simple statistics.

#### 4.3.1.2 Interfície Consola aamc

L'aplicació de consola té modes amb els que podem efectuar diverses operacions. Aquests modes son els següents:

aamc b	Builds an Active Appearance Model.
aamc e	Evaluates an Active Appearance Model.
aamc loo	Leave-one-out evaluation of a training set.
aamc m	Writes Active Appearance Model mode movies.
aamc p	3-level pyramidal (multi-scale) AAM.
aamc r	Tests the regression prediction in an AAM.
aamc reg	Registration from a set of shapes.
aamc s	Active Appearance Model search.
aamc t	Performs tracking in a movie file (.avi).
aamc w	Plots an annotation into it's host image.
aamc sm	Split movie file (.avi) into frames.
aamc cm	Collects a movie file from BMP/HIPS frames.

## 5 Sistema de reconeixement d'emocions

### 5.1 Funcionament general del sistema

L'esquema general del sistema

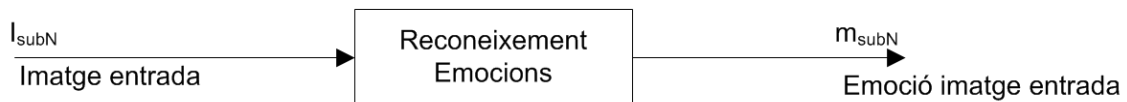


Figura 5-1

ens mostra que a partir d'una imatge d'entrada el sistema haurà de proposar l'emoció que expressa el rostre de la mateixa.

De forma més detallada

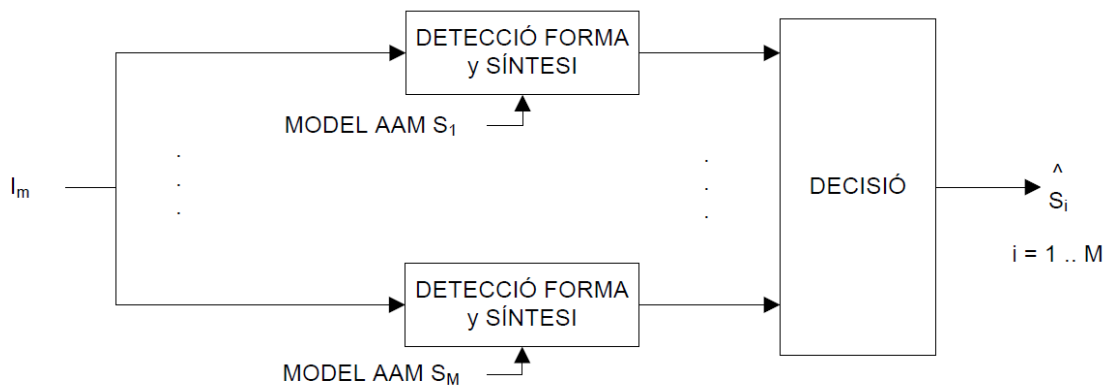


Figura 5-2

veiem que hi ha dues parts:

- detecció dels trets facials de la imatge,  $I_m$ , mitjançant la implementació de la teoria AAM per a un model concret,  $S_i$ , i síntesi de la imatge original utilitzant aquest model
- la decisió de l'emoció de la imatge d'entrada,  $I_m$ , mitjançant la comparació de les diverses formes de les imatges síntesi obtingudes,  $I_s^1$ .

Com ja hem comentat, haurem d'aconseguir primer detectar i caracteritzar el rostre de la imatge d'entrada amb  $M$  models AAM diferents. I després decidir quin dels models caracteritza millor la forma detectada.

## 5.2 Detecció de forma

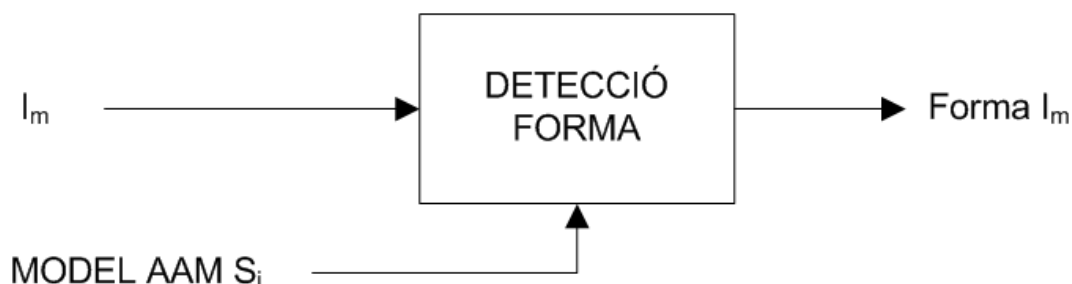


Figura 5-3

L'objectiu d'aquesta part del sistema és la detecció i caracterització de la cara humana que hi ha a una imatge,  $I_m$ , mitjançant un model AAM,  $S_i$ .

Ho aconseguirem mitjançant l'utilització de l'aplicació implementada pel Mikkel B. Stegmann, basada en la teoria AAM de detecció i caracterització de formes.

### 5.2.1 Caracterització de les imatges

El nostre sistema acceptarà un tipus d'imatge amb les següents característiques bàsiques:

- Serà una imatge en 2 dimensions (2-D), podrà ser en color o nivell de gris.
- La imatge mostrarà un rostre humà.

Per al nostre estudi i implementació les imatges utilitzades pertanyen a la base de dades d'imatges anomenada CMU [60].

Aquesta base de dades consta d'un conjunt de 363 imatges de 84 persones on es representen un total de 7 emocions diferents. En els següents quadres podem veure relacionades les dades:

Imatges	363	Imatges dels models	140	Imatges de test	223
Enuig	36	Enuig	20	Enuig	16
Fàstic	34	Fàstic	20	Fàstic	14
Por	37	Por	20	Por	17
Alegria	59	Alegria	20	Alegria	39
Neutre	84	Neutre	20	Neutre	64
Tristesa	50	Tristesa	20	Trist	30
Sorpresa	63	Sorpresa	20	Sorpresa	43

El primer quadre mostra el total de les imatges de les que disposem. En el quadre "Imatges dels models" mostrem les imatges que s'utilitzaran com a base de dades d'entrenament dels models. Les "Imatges de test" són les imatges que utilitzarem per testear el sistema.

Imatges d'exemple dels tipus d'emocions d'aquesta base de dades són les següents:



Figura 5-4

## 5.2.2 Operacions amb l'aplicació de Mikkel B. Stegmann

### 5.2.2.1 Generació dels models de cada emoció $S_i$

Els models AAM de cada emoció són entitats que es configuraran en funció de l'aplicació del sistema de Reconeixement d'Emocions.

La seva generació és un pas necessari en el procés de configuració del sistema.

Per crear cada model l'aplicació de consola *aamc* es pot executar en mode **(b)** per construir un model actiu d'aparença.

La línia de comandes que s'ha de cridar té la següent estructura:

```
aamc b <input dir> <out model> [acf file]
```

<input dir>: Directori que conté les imatges i arxius de punts de control, dades necessàries per caracteritzar el model.

<out model>: Nom de l'arxiu (amb la seva ruta) del model generat.

[acf file]: Arxiu de configuració d'AAM. Conté les dades que defineixen les característiques del model.

Un exemple de línia de comandes per cridar aquesta modalitat és:

```
aamcm b data model __conf__.acf
```

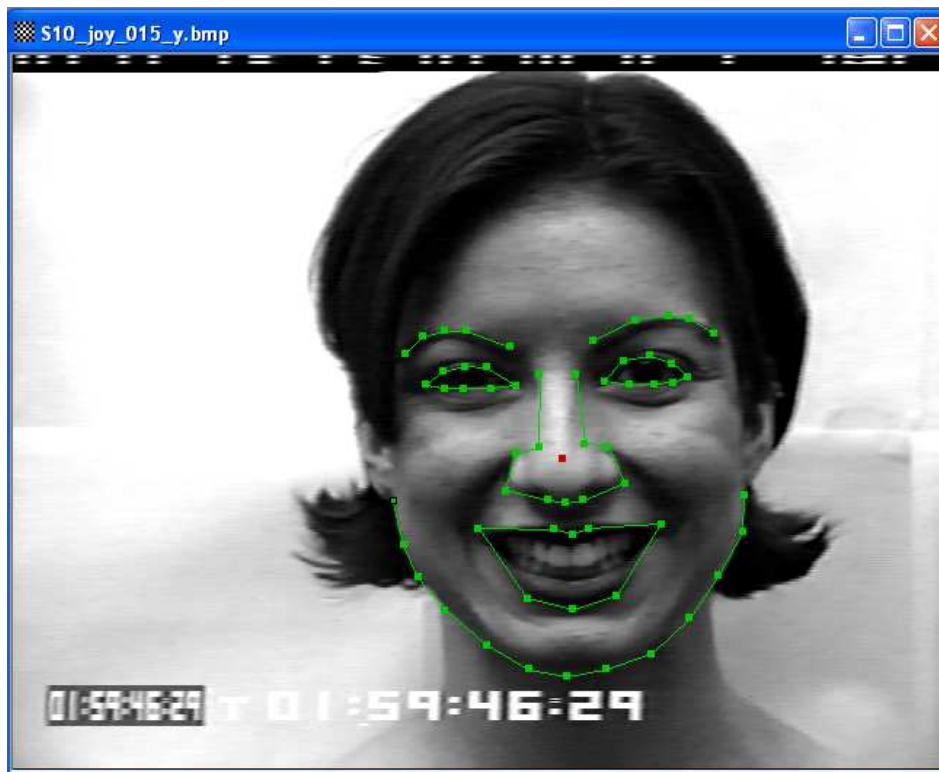


### 5.2.2.1.1 <input dir>: Directori entrada

Les dades que ha de contenir el directori d'entrada són les imatges amb les que volem construir el model i els seus arxius de punts de control (ASF - AAM Shape File)

En aquest arxiu de punts de control s'especifiquen les posicions dels punts de control i la relació entre ells dins de la imatge, de manera que es defineix així la forma i la seva disposició dins de la imatge.

Un exemple d'imatge i arxiu ASF són els següents



*Figura 5-5  
Imatge d'exemple amb els punts de control*

```
#####
#
#   AAM Shape File - written: Tuesday October 03 - 2006 [19:33]
#
#####

#
# number of model points
#
58

#
# model points
#
# format: <path#> <type> <x rel.> <y rel.> <point#> <connects from> <connects to> <user1> <user2> <user3>
#
0 4 0.40836695 0.62351555 0 0 1 0.000000 0.000000 0.000000
0 4 0.41932964 0.68477768 1 0 2 0.000000 0.000000 0.000000
0 4 0.43480340 0.72988433 2 1 3 0.000000 0.000000 0.000000
0 4 0.46358368 0.77556789 3 2 4 0.000000 0.000000 0.000000
0 4 0.50824094 0.82458514 4 3 5 0.000000 0.000000 0.000000
0 4 0.55380541 0.85909498 5 4 6 0.000000 0.000000 0.000000
0 4 0.59412801 0.86850041 6 5 7 0.000000 0.000000 0.000000
0 4 0.63656759 0.85914570 7 6 8 0.000000 0.000000 0.000000
0 4 0.68591231 0.83683002 8 7 9 0.000000 0.000000 0.000000
0 4 0.72653735 0.78674161 9 8 10 0.000000 0.000000 0.000000
0 4 0.75720763 0.72714633 10 9 11 0.000000 0.000000 0.000000
0 4 0.78316528 0.66700602 11 10 12 0.000000 0.000000 0.000000
0 4 0.78475302 0.61487687 12 11 12 0.000000 0.000000 0.000000
1 0 0.72489917 0.44973016 13 20 14 0.000000 0.000000 0.000000
1 0 0.70720768 0.43156573 14 13 15 0.000000 0.000000 0.000000
1 0 0.68369454 0.42006874 15 14 16 0.000000 0.000000 0.000000
1 0 0.65526706 0.42716086 16 15 17 0.000000 0.000000 0.000000
1 0 0.63510585 0.45575112 17 16 18 0.000000 0.000000 0.000000
1 0 0.66229838 0.46047923 18 17 19 0.000000 0.000000 0.000000
1 0 0.68860888 0.46010527 19 18 20 0.000000 0.000000 0.000000
1 0 0.70798886 0.45779192 20 19 13 0.000000 0.000000 0.000000
2 0 0.44261593 0.46100533 21 28 22 0.000000 0.000000 0.000000
2 0 0.46164316 0.44246697 22 21 23 0.000000 0.000000 0.000000
2 0 0.48447879 0.43542561 23 22 24 0.000000 0.000000 0.000000
2 0 0.50854331 0.43666777 24 23 25 0.000000 0.000000 0.000000
2 0 0.53974295 0.46160111 25 24 26 0.000000 0.000000 0.000000
2 0 0.51333165 0.46563196 26 25 27 0.000000 0.000000 0.000000
2 0 0.48324093 0.46595520 27 26 28 0.000000 0.000000 0.000000
2 0 0.46297884 0.46530876 28 27 21 0.000000 0.000000 0.000000
3 4 0.62303424 0.39851987 29 29 30 0.000000 0.000000 0.000000
3 4 0.66824603 0.36965710 30 29 31 0.000000 0.000000 0.000000
3 4 0.70342749 0.36428255 31 30 32 0.000000 0.000000 0.000000
3 4 0.72515124 0.36821201 32 31 33 0.000000 0.000000 0.000000
3 4 0.75244457 0.389555796 33 32 33 0.000000 0.000000 0.000000
4 4 0.53286290 0.40770355 34 34 35 0.000000 0.000000 0.000000
4 4 0.48689517 0.38528645 35 34 36 0.000000 0.000000 0.000000
4 4 0.46305445 0.38474146 36 35 37 0.000000 0.000000 0.000000
4 4 0.43976814 0.39372221 37 36 38 0.000000 0.000000 0.000000
4 4 0.42038810 0.41738144 38 37 38 0.000000 0.000000 0.000000
5 0 0.49939513 0.66189092 39 46 40 0.000000 0.000000 0.000000
5 0 0.58069551 0.66262001 40 39 41 0.000000 0.000000 0.000000
5 0 0.60128522 0.67003536 41 40 42 0.000000 0.000000 0.000000
5 0 0.61764115 0.66132706 42 41 43 0.000000 0.000000 0.000000
5 0 0.69601810 0.65576851 43 42 44 0.000000 0.000000 0.000000
5 0 0.64800906 0.75583810 44 43 45 0.000000 0.000000 0.000000
5 0 0.60151213 0.77522588 45 44 46 0.000000 0.000000 0.000000
5 0 0.55221772 0.75922257 46 45 39 0.000000 0.000000 0.000000
6 4 0.56494457 0.44580069 47 47 48 0.000000 0.000000 0.000000
6 4 0.56436491 0.54844302 48 47 49 0.000000 0.000000 0.000000
6 4 0.53941536 0.55630189 49 48 50 0.000000 0.000000 0.000000
6 4 0.52835178 0.60820937 50 49 51 0.000000 0.000000 0.000000
6 4 0.57442033 0.62164569 51 50 52 0.000000 0.000000 0.000000
6 4 0.59357357 0.62632304 52 51 53 0.000000 0.000000 0.000000
6 4 0.61227322 0.62234294 53 52 54 0.000000 0.000000 0.000000
6 4 0.65645164 0.59827775 54 53 55 0.000000 0.000000 0.000000
6 4 0.63868445 0.54721975 55 54 56 0.000000 0.000000 0.000000
6 4 0.61330646 0.54334098 56 55 57 0.000000 0.000000 0.000000
6 4 0.60435987 0.44569921 57 56 57 0.000000 0.000000 0.000000

#
# host image
#
S10_joy_015_y.bmp
```

Figura 5-6  
Arxiu ASF amb els punts de control

La construcció dels arxius de punts de control (ASF) de cada imatge d'entrenament del model la generarem utilitzant l'aplicació *AAMLab*, una imatge d'exemple és la Figura 5-6. En aquesta aplicació podem moure els punts amb el ratolí i trobarem opcions per desar l'arxiu de punts de control automàticament.

Per a més informació sobre els arxius de punts de control consultar Annex Configuració AAM-API.

### 5.2.2.1.2 <out model>: Model de sortida

La informació sobre el model AAM generat s'emmagatzema en un arxiu on es troben les caracterització del model (AMF - AAM Model File).

Per a més informació sobre els arxius dels models AAM consultar Annex Configuració AAM-API.

### 5.2.2.1.3 [acf file]: Arxiu de configuració del model AAM

La configuració del constructor de models Actius d'Aparença està continguda en un arxiu denominat *Arxiu de configuració del constructor dels models actius d'aparença* (Active Appearance Model Builder Configuration File) i que té l'extensió ACF

En aquest arxiu podrem definir el percentatge a partir del que hem de truncar els models tant de forma, de textura com de combinat (aparença).

Per a més informació sobre els arxius de configuració consultar Annex Configuració AAM-API.

### 5.2.2.2 Síntesi d'imatge mitjançant models AAM

El bloc de detecció de forma que podem veure a la Figura 3-6 té com a objectiu detectar la forma i les característiques de la imatge d'entrada,  $I_m$ , en base al model d'una emoció,  $S_i$  i generar una imatge sintetitzada que approximi la imatge d'entrada, utilitzant el model corresponent.

Aquest bloc el basarem, doncs, en l'aplicació de consola *aamc* en el seu mode (e).

La línia de comandes que s'ha de cridar té la següent estructura:

```
aamc e <model> <dir> [still|movie|both|none*] [pseudo|auto*]
```

<model>: El model AMF que es vol avaluar.

<dir>: Directori que conté les imatges a sintetitzar i un arxiu genèric de punts de control. Amb aquest arxiu de punts de control genèric es defineix la forma mitjana del conjunt de formes de les imatges a sintetitzar i té com a objectiu la inicialització de la optimització a la síntesi del model.

[still|movie|both|none\*]: Escriu imatges de les síntesis inicial i final del model (still) i/o vídeos amb la optimització completa (movie).

[pseudo|auto\*]: Mètode d'inicialització de la forma de referència. El mode auto implementa una inicialització automàtica. El mode pseudo implementa un desplaçament sistemàtic de la posició de referència.

Un exemple de línia de comandes per cridar aquesta modalitat és:

```
aamcm e model.amf unseen both auto
```

### 5.2.2.2.1 <dir>: Directori d'imatges d'entrada

Les dades d'entrada són les imatges que es volen avaluar pel model i, tanmateix, han d'estar acompanyades d'un arxiu genèric de punts de control

El motor AAM-API necessita aquest arxiu genèric de punts de control per inicialitzar el pas inicial de l'optimització de la forma del model. Definirà el centre de gravetat de la iteració inicial de l'optimització com el centre de gravetat de la forma definida a l'arxiu.

### 5.2.2.2.2 Dades de sortida

L'avaluació del model efectua quatre experiments per cada imatge d'entrada. La diferència entre experiment és la posició inicial del centre de gravetat de la forma, cada experiment introdueix una desviació en aquesta posició inicial.

Per a cada experiment el sistema AAM-API genera les següents dades:

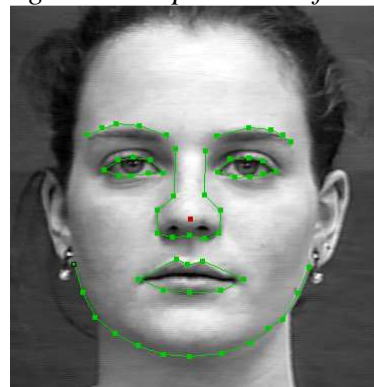
- imatge de test amb la superposició dels punts de control inicials de l'experiment.
- imatge de test amb la superposició dels punts de control finals de l'experiment.
- Arxiu de punts de control amb la forma final de l'experiment.
- Valor dels paràmetres  $\mathbf{b}_s$ ,  $\mathbf{b}_g$  i  $\mathbf{c}$  finals de l'experiment.
- Arxiu de vídeo on es mostra la síntesi del model AAM per a cada iteració de la optimització.
- Arxiu de resultats (càlcul d'errors)

Posarem a continuació uns exemples per una imatge d'entrada amb l'emoció neutre i avaluada amb els models AAM neutre i sorpresa.

*Imatge d'entrada emoció neutre*



*Imatge amb els punts de referència*

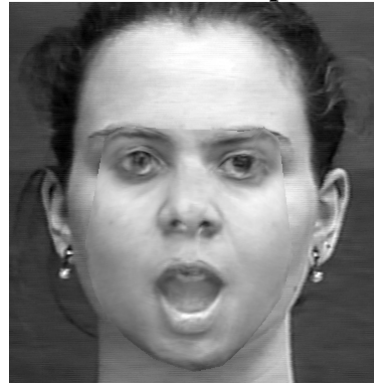




**Model AAM Neutre**



**Model AAM sorpresa**



*Fotograma final de l'arxiu de vídeo*

```
#####
#
# AAM Evaluation - written: Sunday March 14 - 2010 [13:05]
#
#####
Exp Pt.pt. Pt.crv. Overlap Iter Maha Error Time
0 7.2135 2.7235 0.7805 1 4.5986 0.2095 28.0524
```

SUMMARY

	Mean	Std.err	Median	Min	Max
Pt.pt.	7.21	0.00	7.21	7.21	7.21
Pt.crv.	2.72	0.00	2.72	2.72	2.72
Overlap	0.78	0.00	0.78	0.78	0.78
Iterations	1.00	0.00	1.00	1.00	1.00
Mahlanobis	4.60	0.00	4.60	4.60	4.60
Error	0.21	0.00	0.21	0.21	0.21
Time	28.05	0.00	28.05	28.05	28.05

```
#####
#
# AAM Evaluation - written: Sunday March 14 - 2010 [13:08]
#
#####
Exp Pt.pt. Pt.crv. Overlap Iter Maha Error Time
0 12.4709 5.5560 0.4298 1 12.4040 0.9758 29.4286
```

SUMMARY

	Mean	Std.err	Median	Min	Max
Pt.pt.	12.47	0.00	12.47	12.47	12.47
Pt.crv.	5.56	0.00	5.56	5.56	5.56
Overlap	0.43	0.00	0.43	0.43	0.43
Iterations	1.00	0.00	1.00	1.00	1.00
Mahlanobis	12.40	0.00	12.40	12.40	12.40
Error	0.98	0.00	0.98	0.98	0.98
Time	29.43	0.00	29.43	29.43	29.43

*Arxiu de resultats*

### 5.3 Decisió

L'esquema del bloc de decisió mostra dues parts diferenciades. Una primera on es calcularan els paràmetres que quantifiquin l'error de les imatges síntesi resultants del pas de les deteccions de forma, i una segona on es decidirà la imatge de síntesi de menor error i per tant el model AAM més proper a la imatge d'entrada.

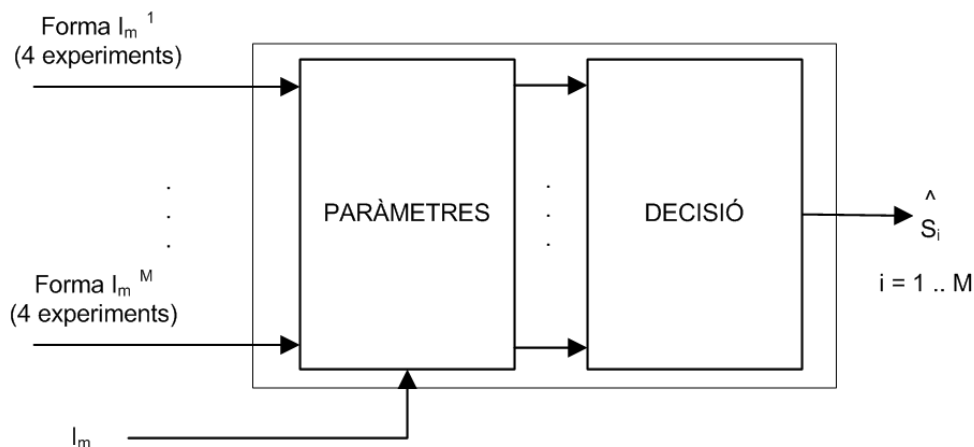
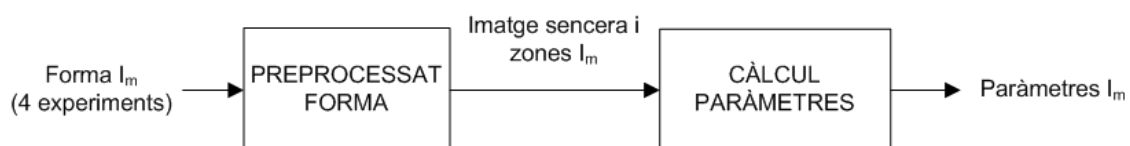


Figura 5-7

#### 5.3.1 Paràmetres

L'objectiu del bloc de paràmetres és donar un valor que quantifiqui la similitud entre la imatge d'entrada i la imatge de síntesi de cada model AAM. O, dit d'una altra manera, volem un valor que quantifiqui l'error en la síntesi de model.



Les dades de les que disposem en aquest moment són les dades de sortida en sintetitzar un model AAM sobre una imatge d'entrada, esmentades al punt 5.2.2.2 Síntesi d'imatge mitjançant models AAM.

De totes les dades que ens proporciona AAM-API, les dades de les que disposarem per al càlcul dels paràmetres seran les següents:

- imatge d'entrada
- Arxiu de vídeo on es mostra la síntesi del model AAM per a cada iteració de la optimització per a cada experiment
- Arxiu de punts de control amb la forma final per a cada experiment.

### 5.3.1.1 Pre processat de les dades

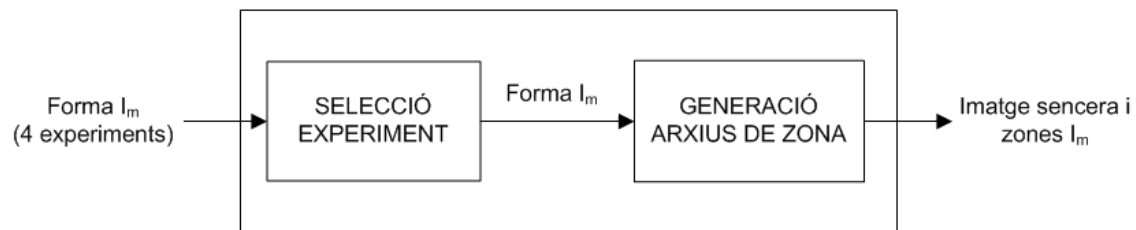


Figura 5-8

El nostre sistema de càlcul de paràmetres necessitarà un processat de les dades per aconseguir les dades desitjades següents:

- imatge d'entrada
- imatge sencera amb la síntesi del model d'un sol experiment.
- imatges de zones significatives de la síntesi del model d'un sol experiment.

#### 5.3.1.1.1 Imatge sencera de la síntesi

La imatge sencera amb la síntesi del model l'extraiem del darrer fotograma de l'arxiu de vídeo on es mostra la síntesi del model AAM per a cada iteració de la optimització.

Aquesta acció la duem a terme obrint l'arxiu .AVI mitjançant les classes de gestió d'arxius d'aquest tipus que el mateix Mikkel B. Stegmann ens proporciona (CAVIFile). Una vegada obert seleccionarem l'últim fotograma, que ens el retornarà en format BMP i, d'aquesta manera, el podrem emmagatzemar en memòria.

Farem aquesta acció per a cada un dels experiments que genera AAM-API.

#### 5.3.1.1.2 Selecció d'un experiment

Una vegada tenim tots els arxius abans esmentats hem d'escollir l'experiment que s'ha ajustat millor, és a dir, amb el que millor s'ha sintetitzat el model respecte la imatge de test original.

Calcularem l'error promig de la imatge resta  $E_m(I_r) = \frac{1}{N \cdot M} \sum_{\substack{x=1..N \\ y=1..M}} |I_r(x, y)|$  on

$I_r = |Im_s(x, y) - Im_o(x, y)|$  entre cada experiment,  $Im_s$ , i la original,  $Im_o$ .

A partir dels resultats obtinguts escollirem l'experiment que hagi resultat tenir el menor error promig.



### 5.3.1.1.3 Imatges zona de la síntesi

Com s'ha enunciat al punt 2 Estat de l'art, els experiments de Johansson suggereixen que les propietats visuals del rostre es clarifiquen si ens fixem en els moviments dels trets facials: celles, ulls i boca.

És per això que en aquest punt varem decidir tenir en compte també zones específiques i generar, per tant, imatges de zones concretes.

Generarem, doncs, dues imatges noves: imatge de la zona de la boca i imatge de la zona dels ulls i celles.

Les dades de les que disposem en aquest punt són:

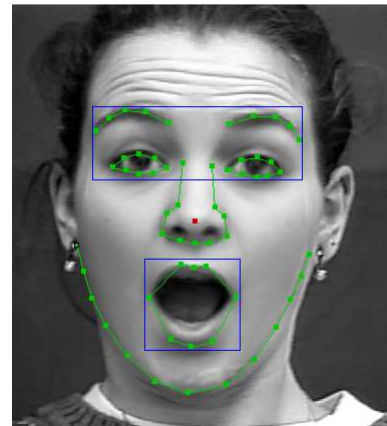
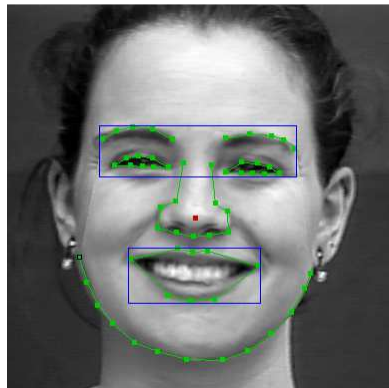
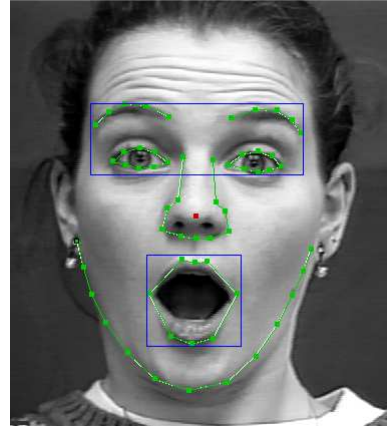
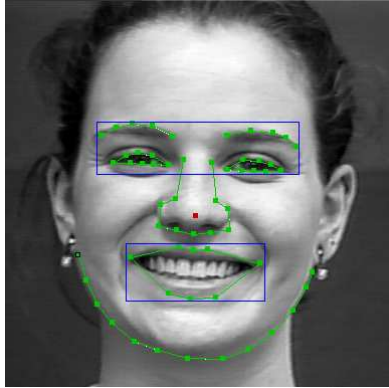
- imatge d'entrada sencera
- imatge sencera amb la síntesi del model d'un sol experiment.
- arxiu de punts de control amb la forma final per a cada experiment.

A partir de l'arxiu de punt de control de la forma final podrem saber les posicions respecte la imatge sencera de síntesi de la boca, dels ulls i de les celles.

Coneixem prèviament la configuració dels arxius de punts de control, ja que els hem utilitzat a la generació del models. Per a una informació més concreta podeu consultar l'Annex Configuració AAM-API. Sabem, doncs, que els punts que definiran la forma de la boca són els que van del número 39 al 46 i els dels ulls i les celles del 13 al 38.

Trobem uns rectangles on es contenen les zones marcades pels punts de control.

*Imatge original amb emoció alegria i síntesi del model AAM alegria.*      *Imatge original amb emoció sorpresa i síntesi del model AAM sorpresa.*



Taula 5-1

Ara, apliquem uns marges d'ampliació dels rectangles de les zones:

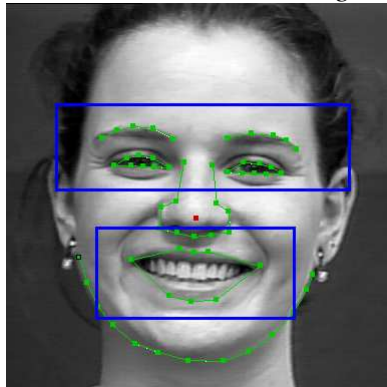
- Zona de la boca:
  - o 80% en l'alçada
  - o 50% en amplada
- Zona dels ulls i celles
  - o 50% en l'alçada
  - o 40% en amplada.

Introduïm aquests marges degut al fet de la variabilitat dels trets facials (boca, ulls i celles) a les diverses emocions. La boca pot mostrar des d'una emoció neutra o de tristesa fins a l'emoció de sorpresa.

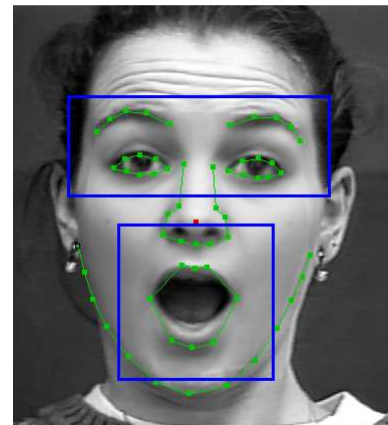
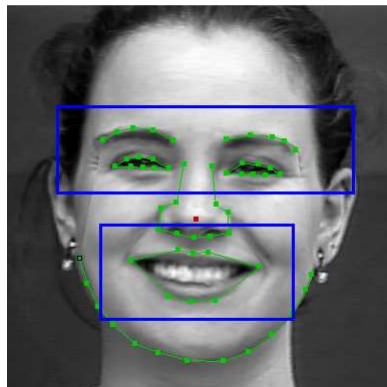
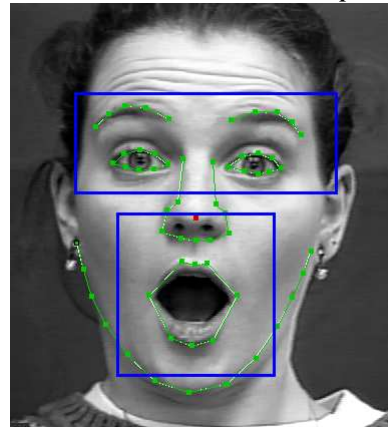
Una altra raó és que el nostre sistema compararà imatges que tindran emocions diferents (una per cada model), així si volem comparar per zones haurem d'intentar que les zones continguin la màxima informació.

Així, per a les mateixes imatges anteriors els rectangles de zona resultaran de la següent manera:

*Imatge original amb emoció alegria i síntesi del model AAM alegria.*



*Imatge original amb emoció sorpresa i síntesi del model AAM sorpresa.*



Taula 5-2

Una vegada tenim els rectangles definits generarem unes imatges de les zones tant per la imatge original com per la sintetitzada. Uns exemple d'ambdues el podem trobar a continuació.

*Imatge original amb emoció alegria i síntesi del model AAM alegria.*

*Imatge original amb emoció sorpresa i síntesi del model AAM sorpresa.*



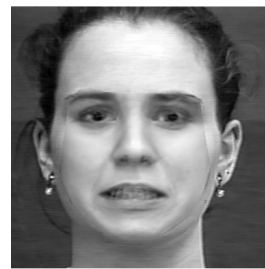
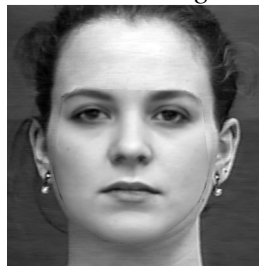
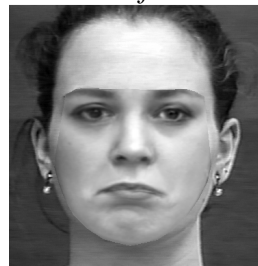
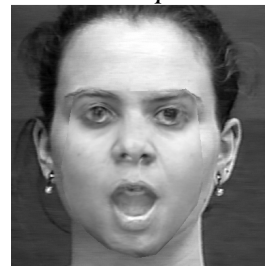
Figura 5-9

5.3.1.2 Proposta paràmetres decisió

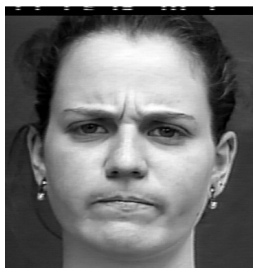
Les dades de les que disposem en aquest punt són:

- imatge d'entrada
- imatge sencera amb la síntesi del model d'un sol experiment per cada un dels models sobre els que volem decidir.
- imatges de zones significatives de la síntesi del model d'un sol experiment per cada un dels models sobre els que volem decidir.

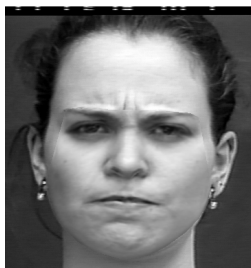
Per tant, ja hem obtingut les dades que necessitem per dur a terme el càlcul dels paràmetres per poder efectuar la decisió i proposar l'emoció de la imatge d'entrada,  $I_m$ .

*Emoció neutre**Imatge original**Imatge síntesi del model enuig**Imatge síntesi del model fàstic**Imatge síntesi del model por**Imatge síntesi del model alegria**Imatge síntesi del model neutre**Imatge síntesi del model tristesa**Imatge síntesi del model sorpresa*

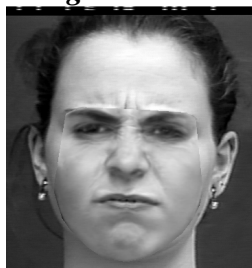
**Emoció enuig**



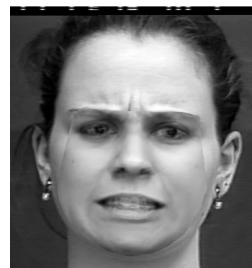
*Imatge original*



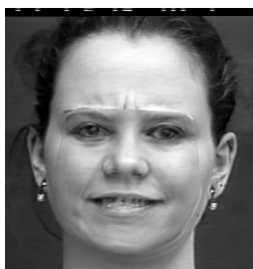
*Imatge síntesi del model enuig*



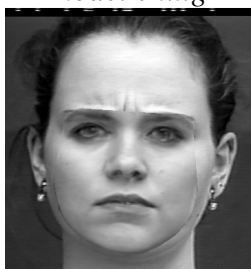
*Imatge síntesi del model fàstic*



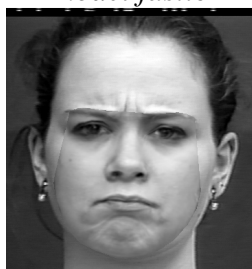
*Imatge síntesi del model por*



*Imatge síntesi del model alegria*



*Imatge síntesi del model neutre*

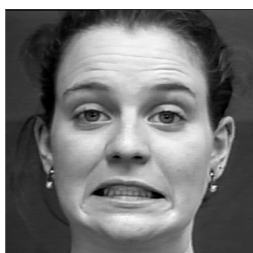


*Imatge síntesi del model tristesa*



*Imatge síntesi del model sorpresa*

**Emoció por**



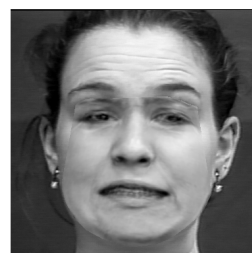
*Imatge original*



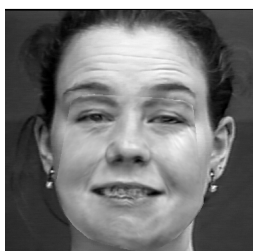
*Imatge síntesi del model enuig*



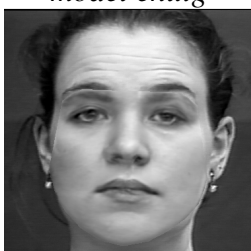
*Imatge síntesi del model fàstic*



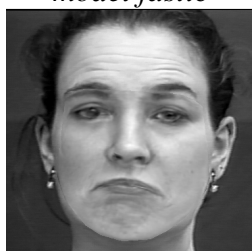
*Imatge síntesi del model por*



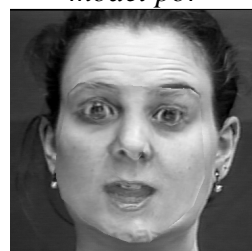
*Imatge síntesi del model alegria*



*Imatge síntesi del model neutre*

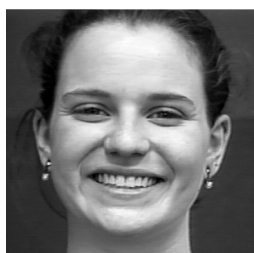


*Imatge síntesi del model tristesa*



*Imatge síntesi del model sorpresa*

**Emoció alegria**



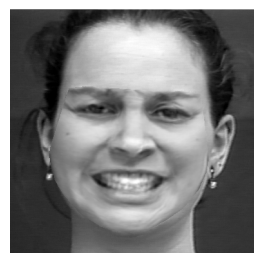
*Imatge original*



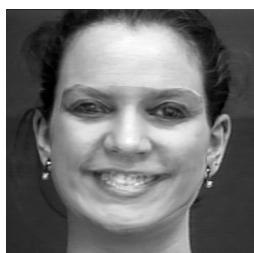
*Imatge síntesi del model enuig*



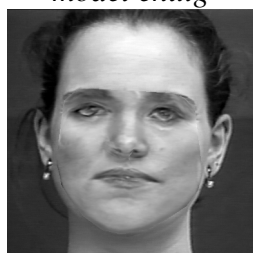
*Imatge síntesi del model fàstic*



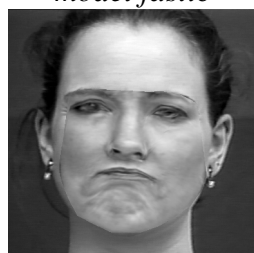
*Imatge síntesi del model por*



*Imatge síntesi del model alegria*



*Imatge síntesi del model neutre*

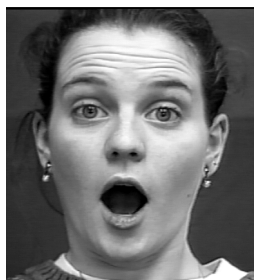


*Imatge síntesi del model tristesa*



*Imatge síntesi del model sorpresa*

**Emoció sorpresa**



*Imatge original*



*Imatge síntesi del model enuig*



*Imatge síntesi del model fàstic*



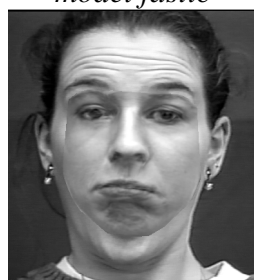
*Imatge síntesi del model por*



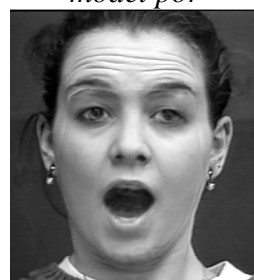
*Imatge síntesi del model alegria*



*Imatge síntesi del model neutre*



*Imatge síntesi del model tristesa*



*Imatge síntesi del model sorpresa*

**Figura 5-10**














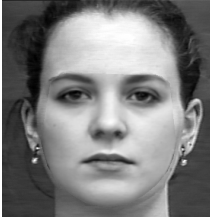


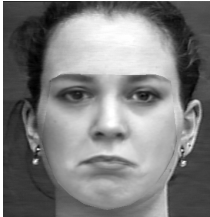


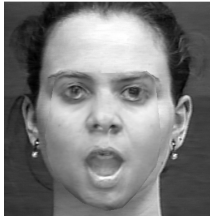

*Dades amb les que treballarem*

Podem observar a les imatges anteriors que la imatge sintetitzada de la que disposem és clarament la superposició de la síntesi del model sobre la imatge original. Podem concloure de forma preliminar, doncs, que la informació sobre l'error de la síntesi la trobarem en aquesta zona de la síntesi superposada.

Generarem, en aquest punt, la imatge resta absoluta entre la imatge original i la imatge sintetitzada. Podem veure uns exemples a continuació:

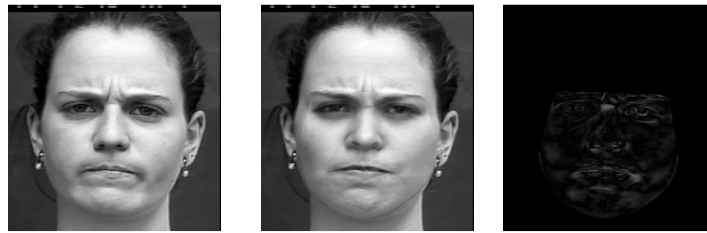


*Emoció neutre*

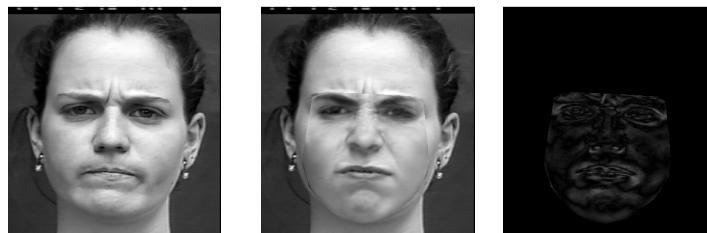
Model d'enuig			
Model de fàstic			
Model de por			
Model d'alegria			
Model neutre			
Model de tristesa			
Model de sorpresa			

*Emoció enuig*

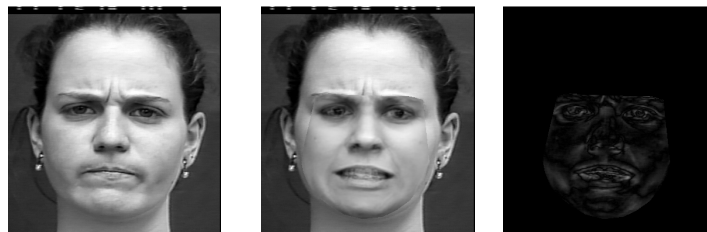
Model d'enuig



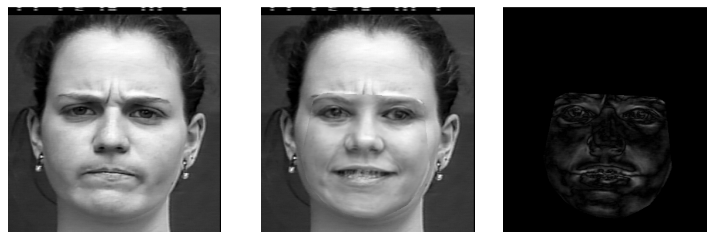
Model de fàstic



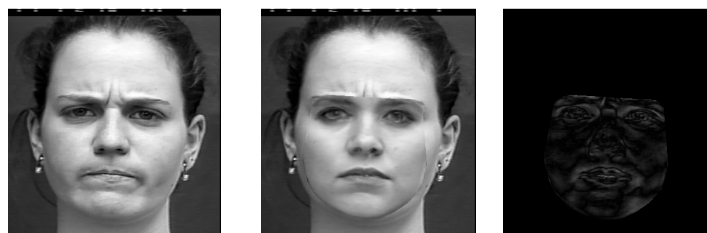
Model de por



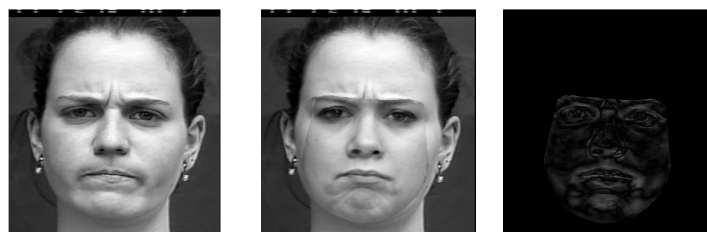
Model d'alegria



Model neutre



Model de tristesa



Model de sorpresa

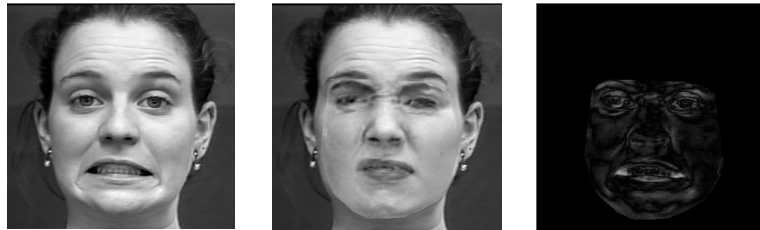


*Emoció por*

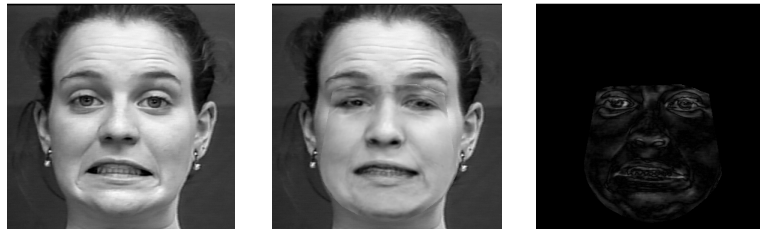
Model d'enuig



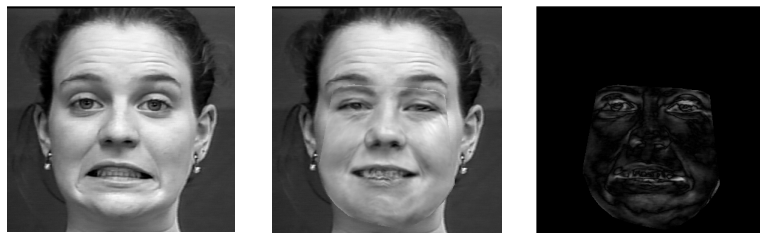
Model de fàstic



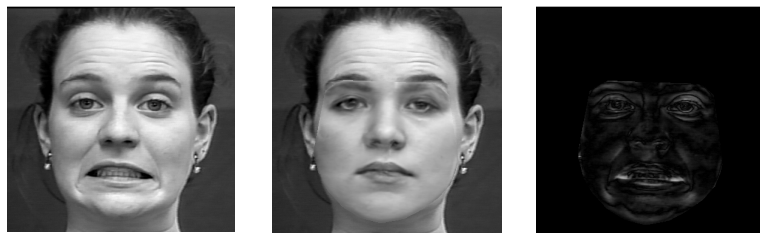
Model de por



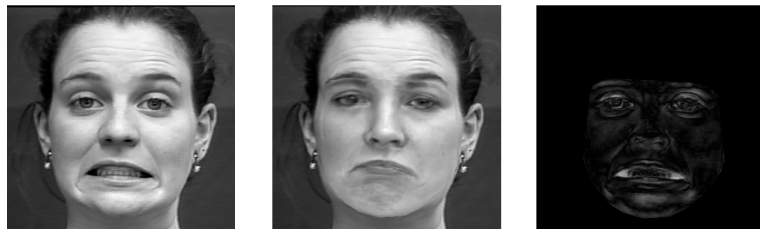
Model d'alegria



Model neutre





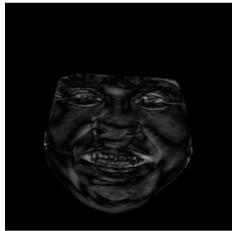




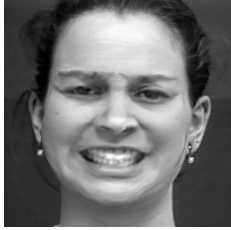



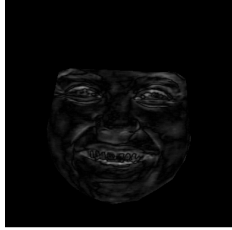

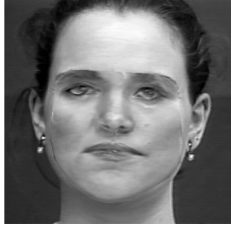


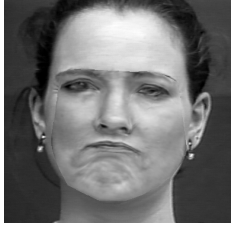




Model de tristesa



Model de sorpresa



*Emoció alegria*

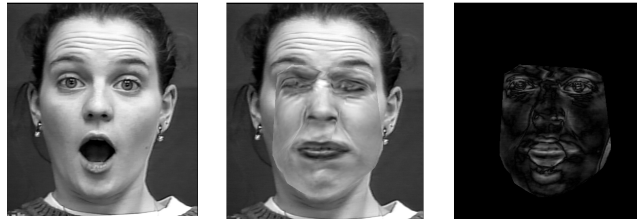
Model d'enuig			
Model de fàstic			
Model de por			
Model d'alegria			
Model neutre			
Model de tristesa			
Model de sorpresa			

*Emoció sorpresa*

Model d'enuig



Model de fàstic



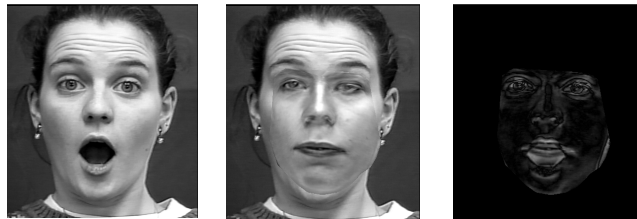
Model de por



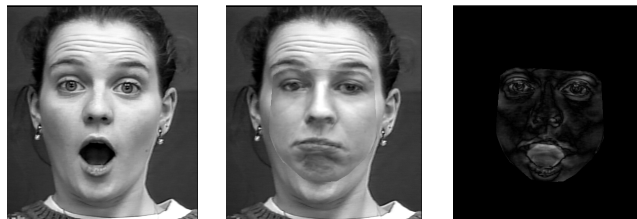
Model d'alegria



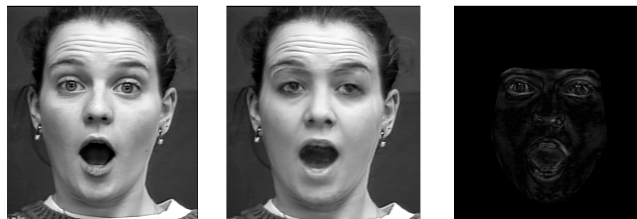
Model neutre



Model de tristesa



Model de sorpresa



Podem veure com els píxels de les zones més diferents entre ambdues imatges tenen un valor de nivell de gris proper a 255 (blanc).

Proposem, doncs, les següents operacions per generar paràmetres de decisió basats en la imatge Resta resultant de restar la imatge sintetitzada i la imatge original:

$$I_r = |\text{Im}_s(x, y) - \text{Im}_o(x, y)| \text{ on } \text{Im}_s \text{ és la imatge síntesi i } \text{Im}_o \text{ és la original.}$$

- Error promig de la imatge Resta:

$$E_m(I_r) = \frac{1}{N \cdot M} \sum_{\substack{x=1..N \\ y=1..M}} |I_r(x, y)|$$

- Error quadràtic promig de la imatge Resta:

$$E_{qm}(I_r) = \frac{1}{N \cdot M} \sqrt{\sum_{\substack{x=1..N \\ y=1..M}} |I_r(x, y)|^2}$$

Proposem també operacions no basades en la resta sinó en la correlació entre la imatge sintetitzada i la imatge original:

- Correlació màxima entre imatges

$$r_{\max} = R_{so}(0) - R_{oo}(0)$$

$$R_{oo}(0) = \sum_{\substack{x=1..N \\ y=1..M}} \text{Im}_o(x, y) \cdot \text{Im}_o(x, y)$$

$$R_{so}(0) = \sum_{\substack{x=1..N \\ y=1..M}} \text{Im}_s(x, y) \cdot \text{Im}_o(x, y)$$

- Correlació màxima normalitzada entre imatges

$$r_{\max n} = R_{son}(0) - R_{oon}(0)$$

$$R_{oon}(0) = \frac{\sum_{\substack{x=1..N \\ y=1..M}} \text{Im}_o(x, y) \cdot \text{Im}_o(x, y)}{\sqrt{\sum_{\substack{x=1..N \\ y=1..M}} \text{Im}_o(x, y)^2 \cdot \sum_{\substack{x=1..N \\ y=1..M}} \text{Im}_o(x, y)^2}}$$

$$R_{son}(0) = \frac{\sum_{\substack{x=1..N \\ y=1..M}} \text{Im}_s(x, y) \cdot \text{Im}_o(x, y)}{\sqrt{\sum_{\substack{x=1..N \\ y=1..M}} \text{Im}_s(x, y)^2 \cdot \sum_{\substack{x=1..N \\ y=1..M}} \text{Im}_o(x, y)^2}}$$

### 5.3.1.3 Estudi Paràmetres Decisió

L'objectiu de l'estudi dels paràmetres de decisió és poder determinar quin d'ells és el que ens dona més informació sobre l'error que hi ha en la imatge de síntesi del model.

Per a acomplir aquest objectiu durem a terme un seguit de proves i seguidament analitzarem els seus resultats.

L'anàlisi dels paràmetres consistirà de diverses etapes:

- Paràmetres quantificadors de l'error
- Distància de Mahalanobis
- Probabilitat d'error

#### 5.3.1.3.1 Paràmetres quantificadors de l'error

La primera etapa tindrà com a objectiu verificar la hipòtesi de que el paràmetre quantifica l'error de la imatge de síntesi. Així, hem definit dos tipus de successos:

- Succés correcte: serà la prova en que la imatge d'entrada i el model de la prova tenen la mateixa emoció.
- Succés incorrecte: en contraposició a l'anterior seran les proves en que la imatge d'entrada i el model no tenen la mateixa emoció.

En base a aquests dos successos calcularem el promig de cada succés

$$m_p = \frac{1}{N} \sum_{i=1}^N p_i,$$

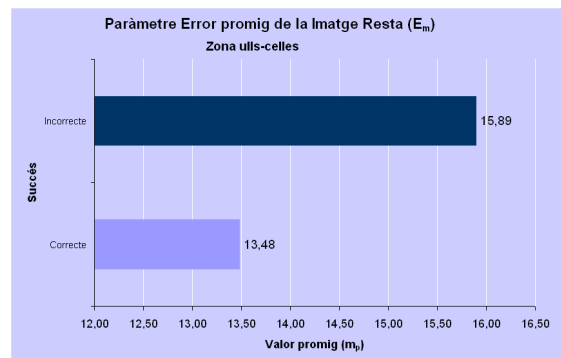
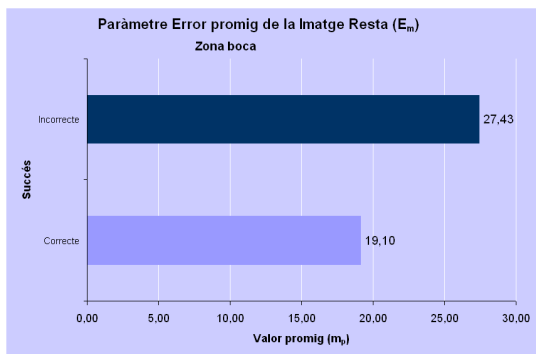
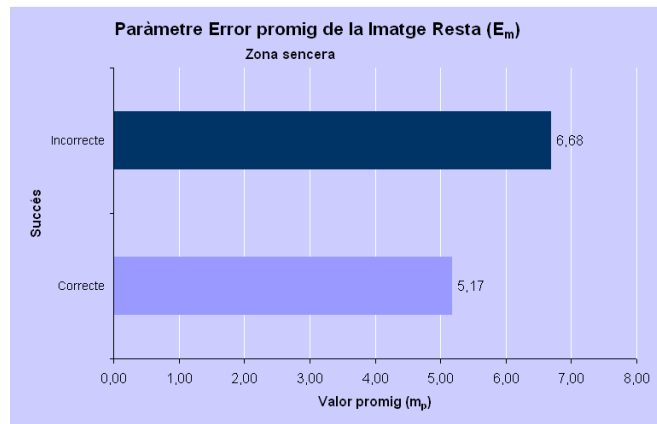
on N seran el total de proves de cada succés i  $p_i$  serà el valor del paràmetre per a la prova número  $i$ ; i la seva desviació típica

$$\sigma_p = \sqrt{V(p)} = \sqrt{\frac{1}{N} \sum_{i=1}^N |p_i - m_p|^2}$$

per a cada paràmetre.

Els resultats d'aquests càlculs els trobem a continuació:

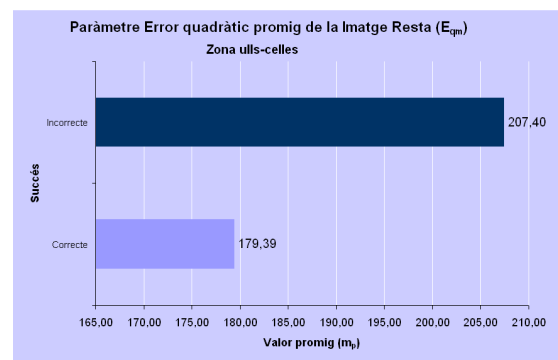
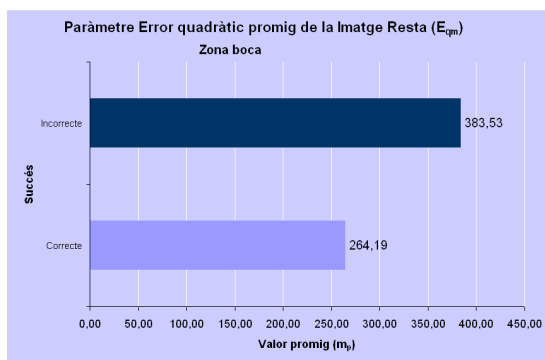
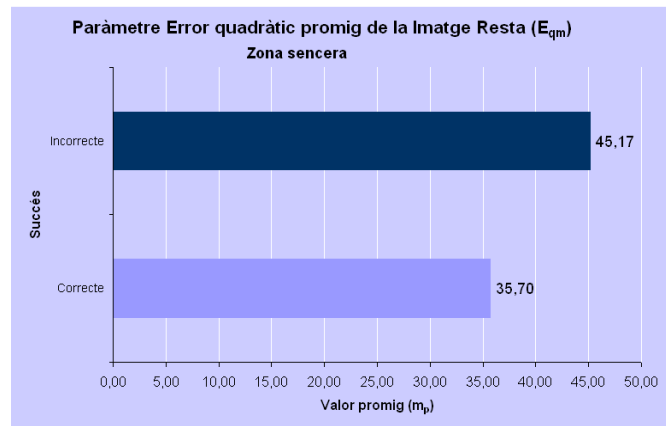
5.3.1.3.1.1 Paràmetre error promig de la imatge Rest a  $E_m$



Desviació típica	Zones		
	Sencera	Boca	Ulls - Celles
Succés correcte	1,38	5,03	3,61
Succés incorrecte	2,00	8,59	4,13

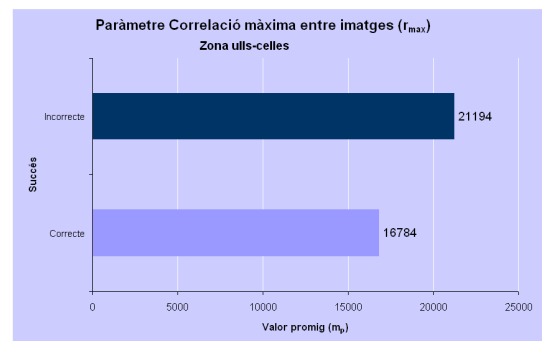
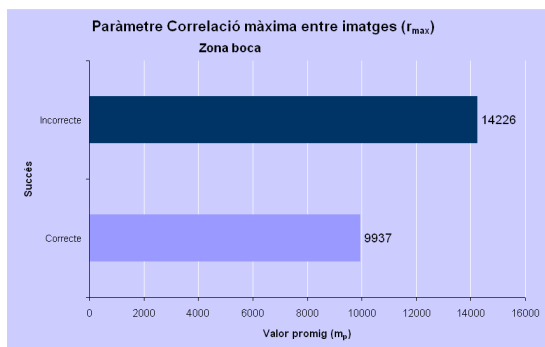
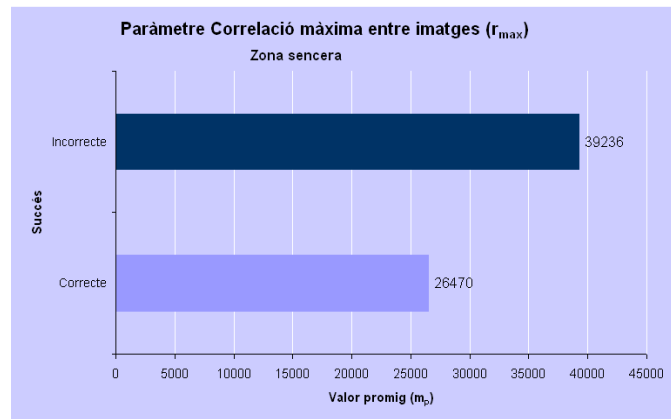


5.3.1.3.1.2 Paràmetre error quadràtic promig de la imatge Resta  $E_{qm}$



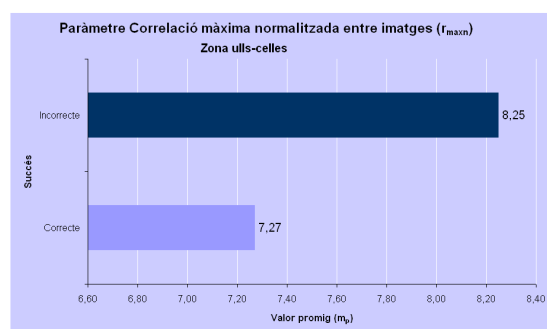
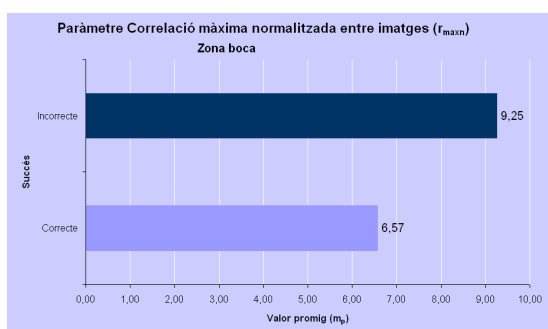
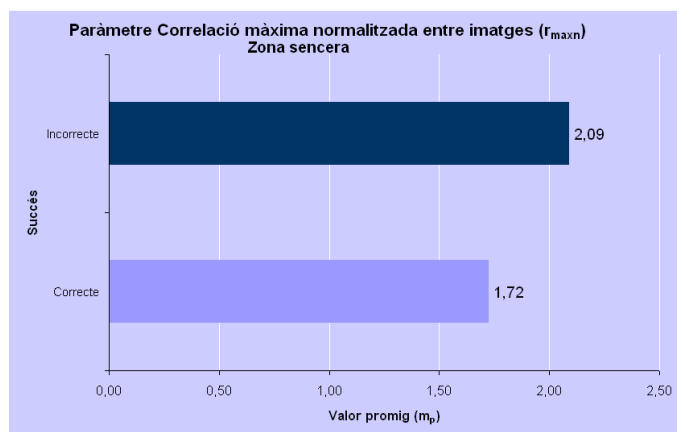
Desviació típica	Zones		
	Sencera	Boca	Ulls - Celles
Succés correcte	9,07	87,20	52,54
Succés incorrecte	11,93	141,92	55,22

5.3.1.3.1.3 Paràmetre correlació màxima entre imatges  $r_{max}$



Desviació típica	Zones		
	Sencera	Boca	Ulls - Celles
Succés correcte	32831,39	17393,91	22552,35
Succés incorrecte	50852,13	25665,39	28667,60

5.3.1.3.1.4 Paràmetre correlació màxima normalitzada entre imatges  $r_{maxn}$



Desviació típica	Zones		
	Sencera	Boca	Ulls - Celles
Succés correcte	2,79	12,25	11,05
Succés incorrecte	3,30	20,25	12,95

Després dels resultats obtinguts podem fer les següents observacions:

- En tots els paràmetres el promig en el succés correcte té com a resultat un valor menor que el succés incorrecte.
- El promig de cada paràmetre en el succés correcte és menor independentment de la zona on es calculi. Així, tant per la imatge sencera com per les imatges de la zona de la boca, ulls i celles, tots els paràmetres tenen com a promig en el succés correcte un valor menor que al succés incorrecte.

En aquest moment podem donar com a validada la hipòtesi de que el paràmetre quantifica l'error de la imatge de síntesi. No només això sinó que podem dir que tots els paràmetres quantifiquen l'error de les imatges de síntesi.

### 5.3.1.3.2 Distància de Mahalanobis

La distància de Mahalanobis és utilitzada per determinar la similitud entre dues variables aleatòries. L'equació utilitzada per calcular aquesta distància serà:

$$d_m(p_{ok}, p_{nok}) = \frac{(m_{pok} - m_{pnok})^2}{V(p)}$$

on  $m_{pok}$  serà el valor promig del paràmetre per al succés correctes,  $m_{pnok}$  serà el valor promig per al succés incorrectes i  $V(p)$  serà la variància del paràmetre.

#### 5.3.1.3.2.1 Imatge sencera

Els resultats obtinguts són.

Distància Mahalanobis	Paràmetre			
	$E_m$	$E_{qm}$	$r_{max}$	$r_{maxn}$
	0,57	0,62	0,07	0,01

En aquest cas els millors paràmetres són :

- Error quadràtic promig de la imatge Resta,  $E_{qm}$ .
- Error promig de la imatge Resta,  $E_m$ .

#### 5.3.1.3.2.2 Imatge de zona: Boca

Fem el mateix càlcul amb la zona boca i obtenim:

Distància Mahalanobis	Paràmetre			
	$E_m$	$E_{qm}$	$r_{max}$	$r_{maxn}$
	0,92	0,71	0,03	0,02

En aquest cas els millors paràmetres són :

- Error promig de la imatge Resta,  $E_m$ .
- Error quadràtic promig de la imatge Resta,  $E_{qm}$ .

#### 5.3.1.3.2.3 Imatge de zona: Ulls i celles

A la zona d'ulls i celles obtenim els següents valors:

Distància Mahalanobis	Paràmetre			
	$E_m$	$E_{qm}$	$r_{max}$	$r_{maxn}$
	0,34	0,25	0,02	0,01

En aquest cas els millors paràmetres són :

- Error promig de la imatge Resta,  $E_m$ .
- Error quadràtic promig de la imatge Resta,  $E_{qm}$ .

### 5.3.1.3.3 Probabilitat d'error

A l'hora d'estudiar quin és el millor paràmetre, una de les dades més importants que haurem de tenir en compte és la probabilitat d'error en decidir basant-nos en un paràmetre concret.

La decisió per un paràmetre té similitud amb el criteri MAP de decisió, que diu:

*Decideix que s'ha transmès el símbol  $s_j$ , si havent rebut  $y(t)$  la probabilitat d'haver transmès el símbol  $s_j$  és màxima.*

En el nostre cas, doncs, prendrem el criteri de decisió següent:

Decideix que la imatge d'entrada té l'emoció  $S_i$ , quan la imatge sintetitzada del model AAM  $S_i$  és la de major similitud o, el que és el mateix, la de mínim error.

De forma més concreta, la decisió per calcular la probabilitat d'error escollirà la síntesi del model AAM que tingui el menor valor del paràmetre.

Una vegada implementat ho apliquem a les proves fetes i utilitzades anteriorment obtenint els resultats següents segons la zona.

El càlcul de la probabilitat d'error s'ha fet de la següent forma:

Per a una imatge d'entrada d'un tipus d'emoció  $S_i$ , decidim segons una zona concreta. Si l'emoció estimada (la de mínim valor del paràmetre) és  $S_j$ , amb  $j \neq i$ , comptabilitza com a error, si l'emoció estimada és  $S_i$ , comptabilitza com a no erroni.

Finalment fem l'operació:

$$P(e) = \# \text{ estimacions amb error} / \# \text{ Total proves}$$

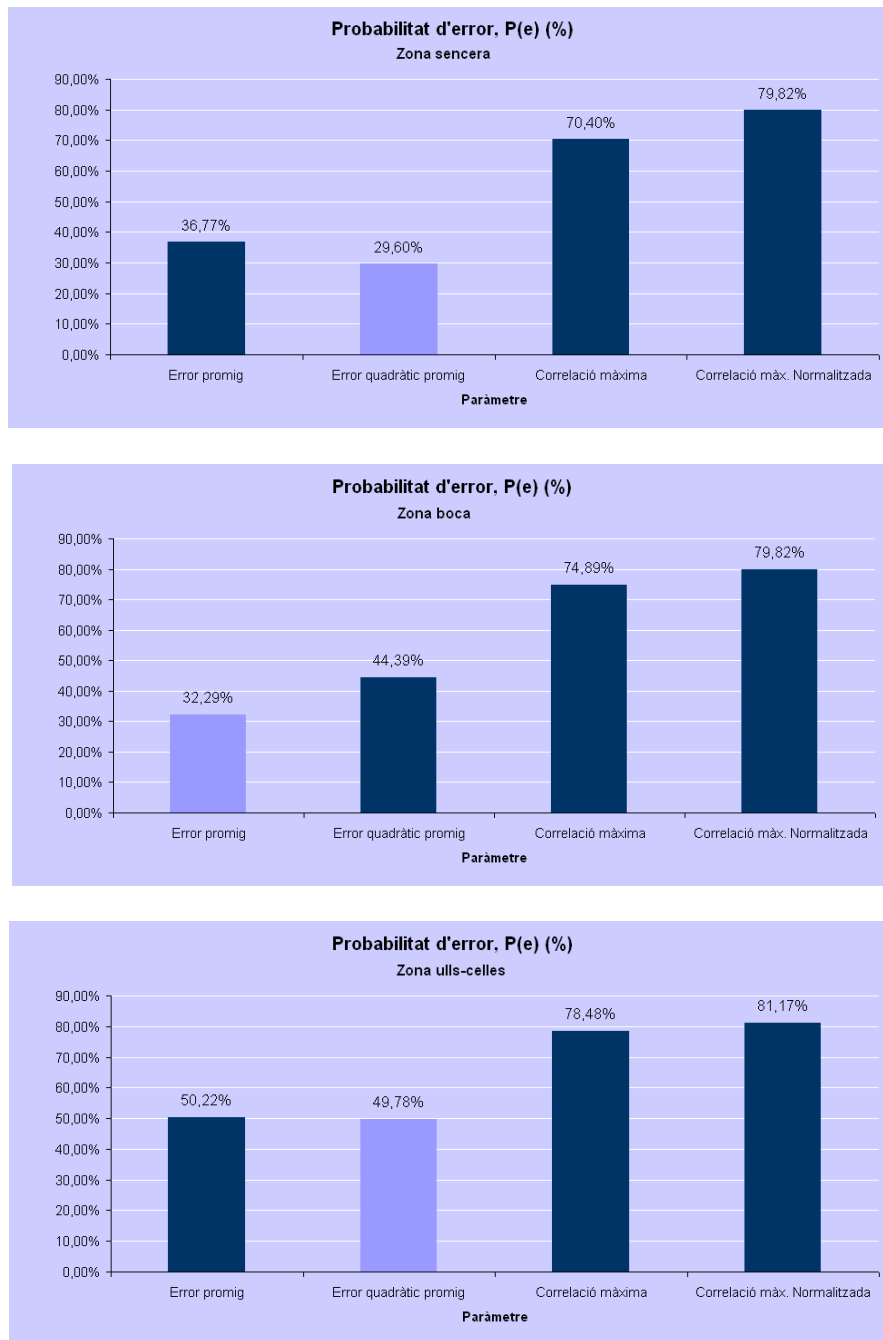


Figura 5-11

Podem veure que per a cada zona tenim un paràmetre que té una probabilitat d'error mínima.

- Imatge sencera: Error quadràtic promig de la imatge Resta,  $E_{qm}$ .
- Imatge de la zona de la boca: Error promig de la imatge Resta,  $E_m$ .
- Imatge de la zona de ulls i celles: Error quadràtic promig de la imatge Resta,  $E_{qm}$ .

### 5.3.1.4 Paràmetre de Decisió Òptim

Una vegada verificada la hipòtesi de que tots els paràmetres estudiats quantificaven l'error que hi havia a les imatges de síntesi dels models, el següent pas és definir un criteri per destriar quin serà el paràmetre òptim que utilitzarem al bloc de decisió.

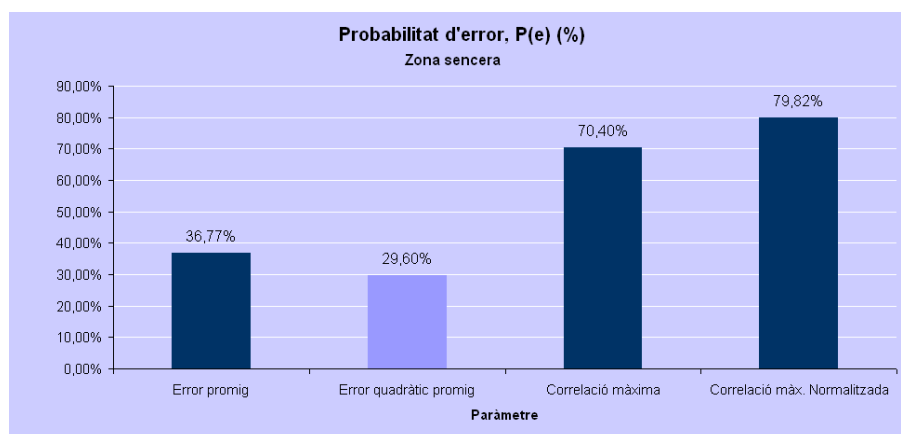
El paràmetre de decisió òptim haurà de complir el criteri següent:

- Haurà de ser el paràmetre que tingui **major distància Mahalanobis** entre els promigs dels valors de les proves dels **successos correcte i incorrecte**.
- I haurà de ser el paràmetre que tingui **menor probabilitat d'error** en una decisió per valor mínim.

De manera preliminar, escollirem un paràmetre òptim per a cada tipus d'imatge: Sencera, Zona de la boca, Zona dels ulls i celles.

#### 5.3.1.4.1 Imatge sencera

Si recuperem les dades trobades a punts anteriors sobre la probabilitat d'error i la distància de Mahalanobis entre successos, tindrem



Distància Mahalanobis	Paràmetre			
	$E_m$	$E_{qm}$	$r_{max}$	$r_{maxn}$
	0,57	0,62	0,07	0,01

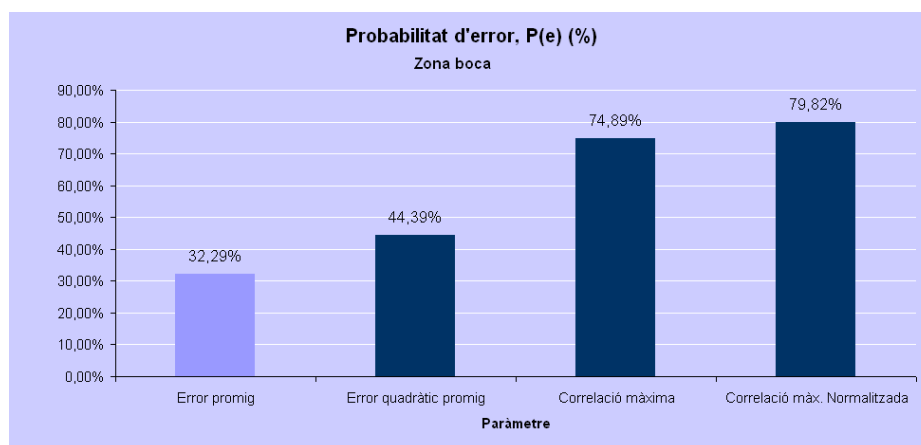
Podem veure per a les dues magnituds que podem descartar els paràmetres basats en la correlació de les imatges original i sintetitzada, i que ens haurem de fixar en els basats en la imatge resta.

Veiem una relació entre les dues magnituds: la probabilitat d'error i la distància de Mahalanobis. Cosa que és lògica ja que la probabilitat d'error es calcula en base a cada prova, amb el que la variació dels valors dels paràmetres li afectarà més, i és la desviació típica el que quantifica aquesta variació.

Així, el paràmetre que tindrem com a paràmetre de decisió per a la **imatge sencera** serà l' **Error quadràtic promig de la imatge Resta,  $E_{qm}$** .

### 5.3.1.4.2 Imatge de zona: Boca

Si recuperem les dades trobades a punts anteriors sobre la probabilitat d'error i la distància de Mahalanobis entre successos, tindrem



Distància Mahalanobis	Paràmetre			
	$E_m$	$E_{qm}$	$\Gamma_{max}$	$\Gamma_{maxn}$
	0,92	0,71	0,03	0,02

De la mateixa manera que a la imatge sencera, a la imatge de la zona de la boca podem veure per a les dues magnituds que podem descartar els paràmetres basats en la correlació de les imatges original i sintetitzada, i que ens haurem de fixar en els basats en la imatge resta.

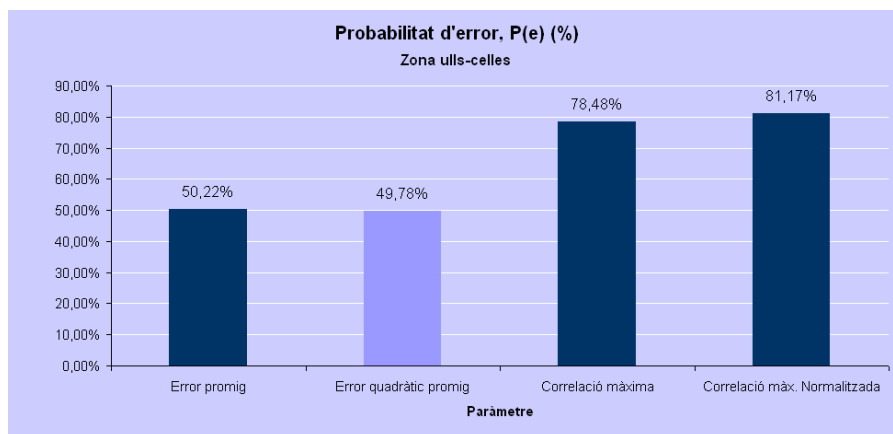
Tanmateix, també veiem una relació entre les dues magnituds. Veiem que el paràmetre que té menor probabilitat d'error també és el que té menor distància Mahalanobis.

Així, el paràmetre que tindrem com a paràmetre de decisió per a la **imatge de la zona de la boca** serà l'**Error promig de la imatge Resta,  $E_m$** .



### 5.3.1.4.3 Imatge de zona: Ulls i celles

Si recuperem les dades trobades a punts anteriors sobre la probabilitat d'error i la distància de Mahalanobis entre successos, tindrem



Distància Mahalanobis	Paràmetre			
	$E_m$	$E_{qm}$	$r_{max}$	$r_{maxn}$
	0,34	0,25	0,02	0,01

De la mateixa manera que a les dues zones anteriors, a la imatge de la zona dels ulls i celles podem veure per a les dues magnituds que podem descartar els paràmetres basats en la correlació de les imatges original i sintetitzada, i que ens haurem de fixar de nou en els basats en la imatge resta.

En aquest cas les dades de probabilitat d'error i distància de Mahalanobis són diferents. En aquest cas escollirem el paràmetre que té la major distància de Mahalanobis, ja que els valors de la probabilitat d'error són molt similars. Així, el paràmetre que tindrem com a paràmetre de decisió per a la **imatge de la zona dels ulls i les celles** serà l'**Error promig de la imatge Resta,  $E_m$** .

**En resum**, tindrem els següents paràmetres de decisió per a cada tipus de zona de la imatge:

- Imatge sencera
  - Error quadràtic promig de la imatge Resta,  $E_{qm}$ .
- Imatge de la zona de la boca
  - l'Error promig de la imatge Resta,  $E_m$ .
- Imatge de la zona dels ulls i celles
  - Error promig de la imatge Resta,  $E_m$ .

### 5.3.2 Mètode de Decisió



Figura 5-12

L'objectiu del bloc de decisió és proposar l'emoció que té la imatge d'entrada.

Així, l'objectiu és trobar la imatge de síntesi del model AAM  $S_i$  que menor error tingui. O, dit d'una altra manera, major similitud mostri amb la imatge d'entrada.

Les dades de les que disposem en aquest punt són el valor del paràmetres d'Error promig,  $E_m$ , i d'Error quadràtic promig,  $E_{qm}$ , segons el tipus de zona de la imatge per a cada model AAM del sistema. Recordar, que cada model AAM correspon a una emoció,  $S_i$ , sobre la que el sistema haurà de decidir.

De forma més detallada el diagrama de blocs és:

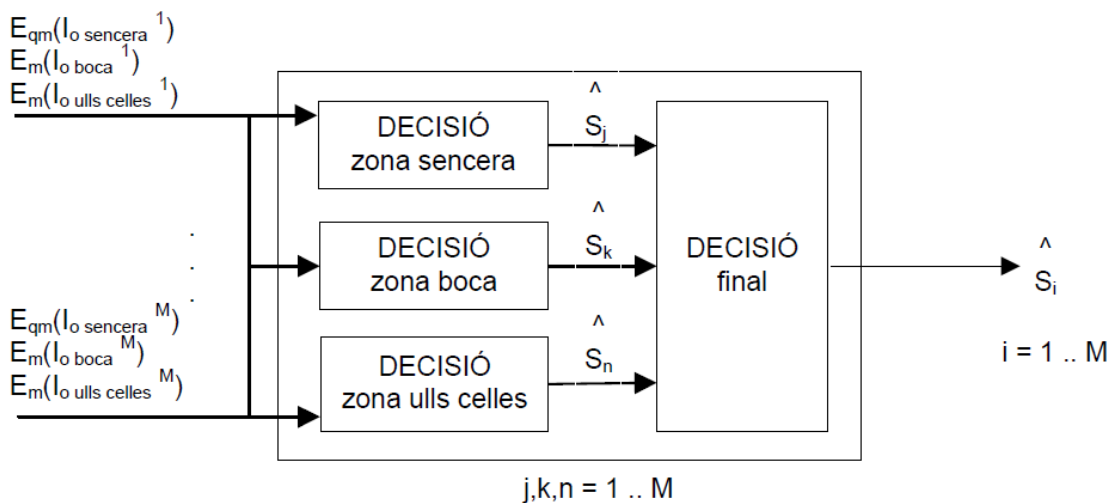


Figura 5-13

Com podem veure a la Figura 5-13 les dades en les que ens hem de basar per decidir són:

Per a cada model AAM d'emoció,  $S_i$ , tenim

- Valor del paràmetre error quadràtic promig de la imatge resta,  $E_{qm}$ , de la imatge sencera.
- Valor del paràmetre error promig de la imatge resta,  $E_m$ , de la imatge de la zona de la boca.
- Valor del paràmetre error promig de la imatge resta,  $E_m$ , de la imatge de la zona dels ulls i les celles.

Passem a dur a terme ara un estudi d'aquests valors segons el model i la zona.

### 5.3.2.1 Anàlisi dels paràmetres

Hem pogut veure al punt 5.3.1.4 Paràmetre de Decisió Òptim que l'error comés a la decisió per a una zona concreta té una relació molt estreta amb la desviació típica dels valors del paràmetre en aquella zona.

En aquest punt ens proposem extreure unes dades analítiques sobre les diferències entre les zones i les emocions. Però, per a fer això, ens trobem que ens cal una dada més concreta que, creiem, podrem trobar en la probabilitat d'error de cada zona i de cada emoció:

$$P(e)_{Z_j, S_i} = \# \text{ estimacions amb error }_{Z_j, S_i} / \# \text{ Total proves }_{Z_j, S_i}$$

on  $Z_j$  és la zona  $j$  i  $S_i$  és l'emoció  $i$ .

D'aquesta manera, farem el càlcul d'aquesta magnitud i obtenim les dades que aquí mostrem

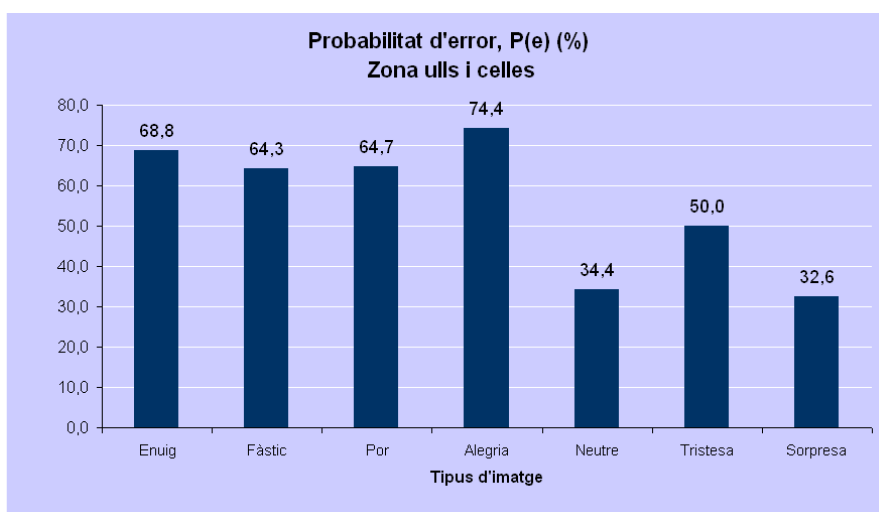
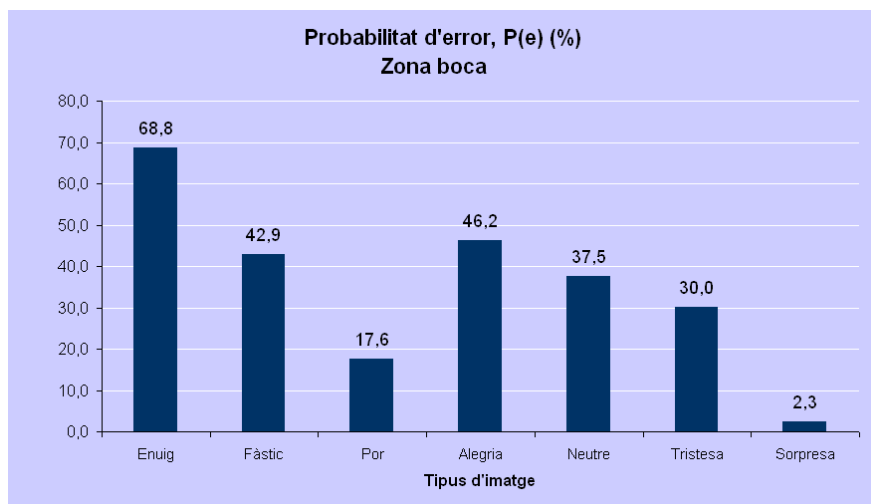
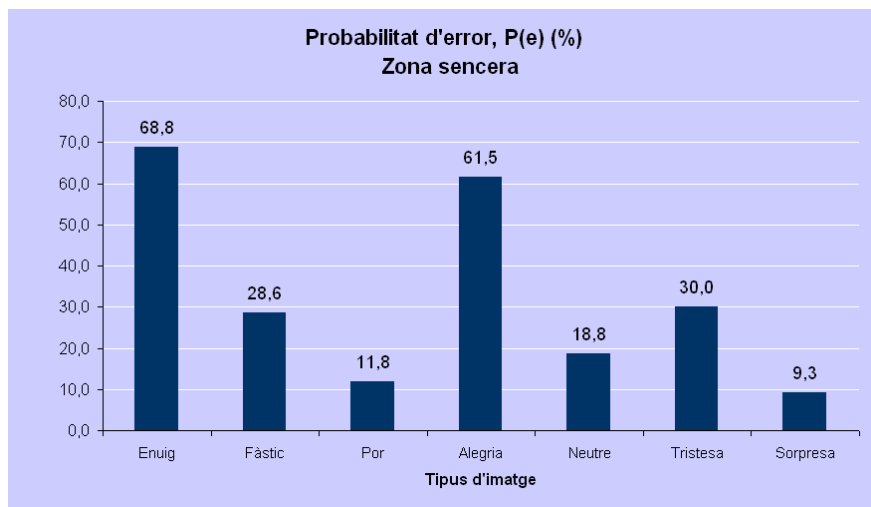


Figura 5-14

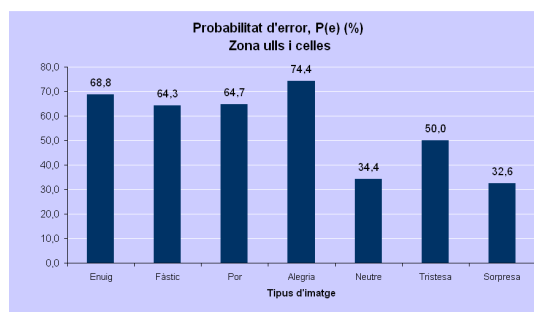
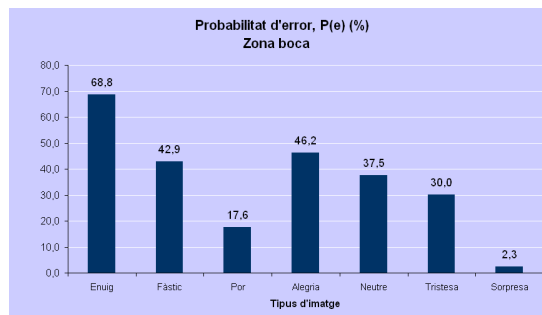
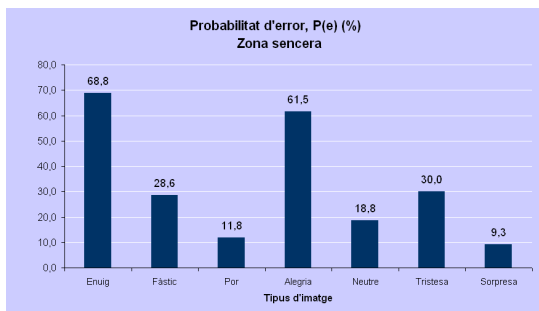
Observant les dades obtingudes, podem notar que la probabilitat d'error varia en funció de la zona i de l'emoció dintre de la zona. Amb l'objectiu d'extreure aquestes dues variacions distingirem entre dues que denominarem variació intra-zonal i variació inter-zonal.

- La variació intra-zonal la definirem com la variació de la probabilitat d'error en una zona,  $i$
- la variació inter-zonal serà la variació entre les tres zones de la probabilitat d'error d'una emoció.

Després de la constatació d'aquestes variacions el més lògic serà voler quantificar-les en forma de coeficients. Així, en els següents punts trobarem els coeficients de les dues variacions.

### 5.3.2.1.1 Coeficients intra-zonals

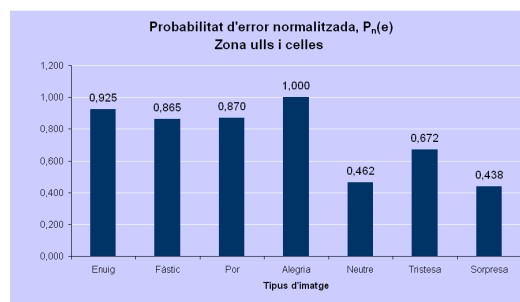
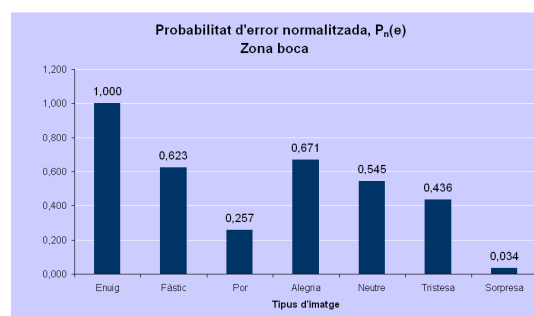
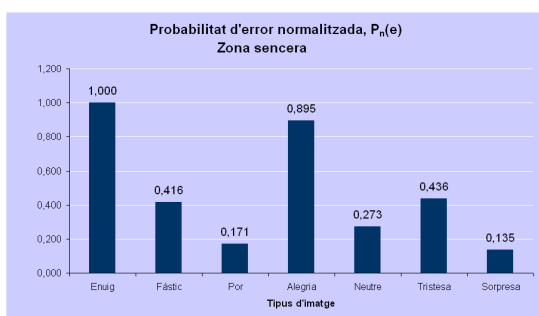
Com em dit abans la variació intra-zonal és aquella que es caracteritza per la variació de la probabilitat d'error dins de la mateixa zona entre les diferents emocions. Així, les gràfiques que la mostren són les següents:



Per poder extraure els coeficients haurem de normalitzar cada zona per la seva probabilitat d'error màxima:

- Zona sencera: la probabilitat d'error màxima és 68,8.
- Zona boca: la probabilitat d'error màxima és 68,8.
- Zona ulls i celles: la probabilitat d'error màxima és 74,4.

Obtenint les dades que es mostren a continuació:



Així, els coeficients intra-zonals seran els següents:

	<b>Enuig</b>	<b>Fàstic</b>	<b>Por</b>	<b>Alegria</b>	<b>Neutre</b>	<b>Tristesa</b>	<b>Sorpresa</b>
<b>Zona sencera</b>	1,000	0,416	0,171	0,895	0,273	0,436	0,135
<b>Zona boca</b>	1,000	0,623	0,257	0,671	0,545	0,436	0,034
<b>Zona ulls-celles</b>	0,925	0,865	0,870	1,000	0,462	0,672	0,438

*Figura 5-15*

### 5.3.2.1.2 Coeficients inter-zonals

En aquest punt farem el càlcul dels coeficients de la variació inter-zonal. Recordarem, doncs, la seva definició. La variació inter-zonal és aquella que es caracteritza per la variació de la probabilitat d'error d'una mateixa emoció entre les diferents zones. Així, per poder visualitzar-ho millor disposarem les dades de la Figura 5-14 per emoció.

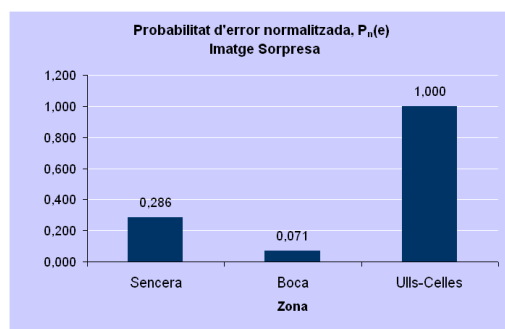
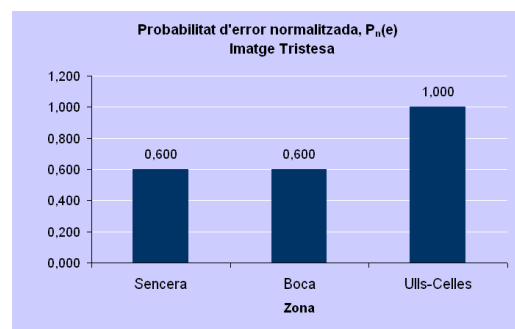
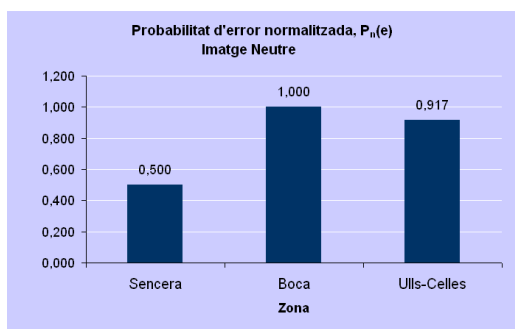
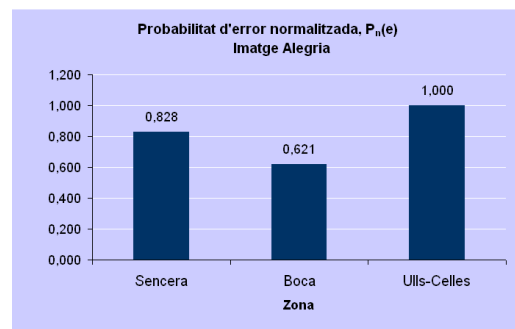
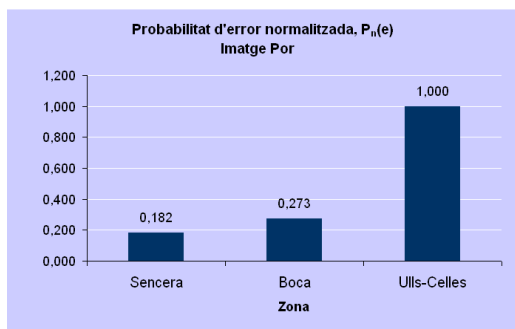
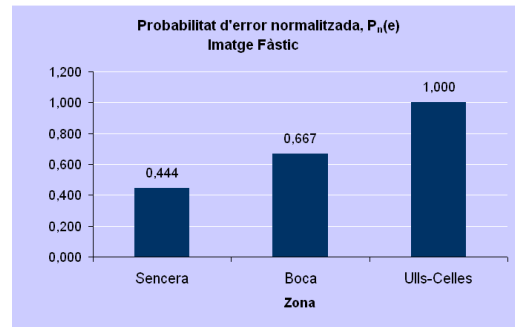
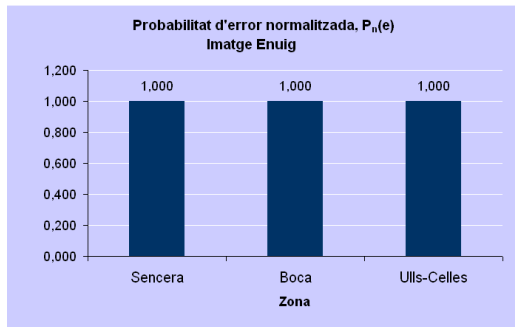




De la mateixa manera que en el cas dels coeficients intra-zonals, per poder extraure'ls haurem de normalitzar cada emoció per la seva probabilitat d'error màxima:

- Emoció Enuig: la probabilitat d'error màxima és 68,8.
- Emoció Fàstic: la probabilitat d'error màxima és 64,3.
- Emoció Por: la probabilitat d'error màxima és 64,7.
- Emoció Alegria: la probabilitat d'error màxima és 74,4.
- Emoció Neutre: la probabilitat d'error màxima és 37,5.
- Emoció Tristesia: la probabilitat d'error màxima és 50,0.
- Emoció Sorpresa: la probabilitat d'error màxima és 32,6.

Obtenint les dades que es mostren a continuació:



Així, els coeficients inter-zonals seran els següents:

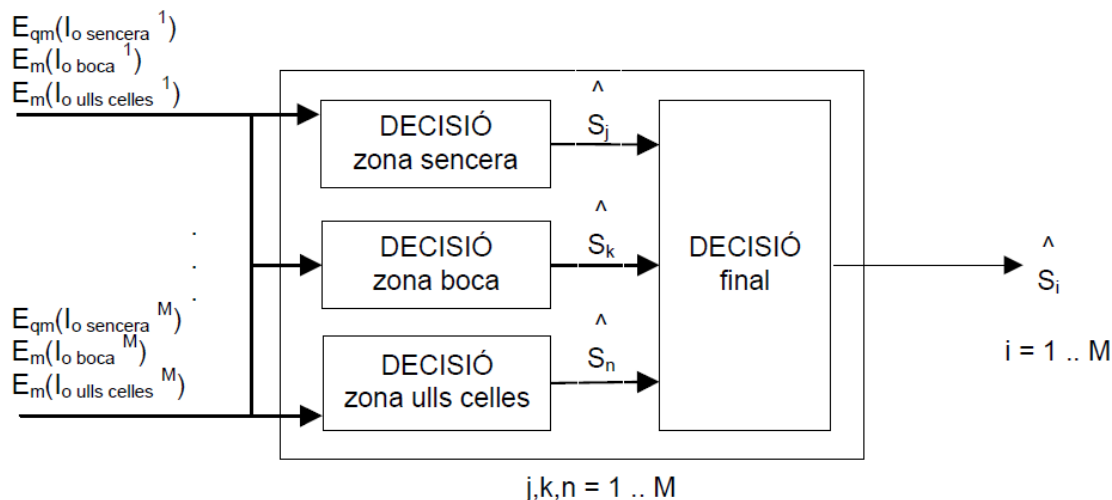
	Sencera	Boca	Ulls-Celles
<b>Enuig</b>	1,000	1,000	1,000
<b>Fàstic</b>	0,444	0,667	1,000
<b>Por</b>	0,182	0,273	1,000
<b>Alegria</b>	0,828	0,621	1,000
<b>Neutre</b>	0,500	1,000	0,917
<b>Tristesa</b>	0,600	0,600	1,000
<b>Sorpresa</b>	0,286	0,071	1,000

*Figura 5-16*

5.3.2.2 Proposta criteris de decisió

Després d'obtenir aquestes dades a partir de l'anàlisi dels resultats de les proves, definirem a continuació els possibles criteris de decisió.

Recordem ara l'esquema del a decisió mostrat anteriorment a la Figura 5-13:



Proposarem diversos criteris tant pel bloc de decisió per zona com pel bloc de decisió final de forma que la seva combinació definirà els criteris de decisió proposats.

5.3.2.2.1 Decisió per zona

Començarem, doncs, per la decisió per zona.

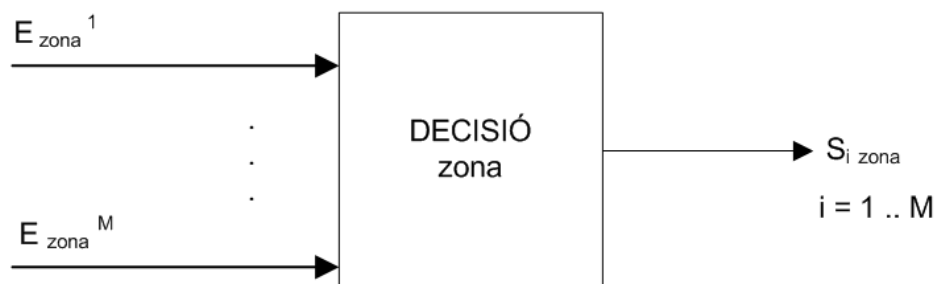
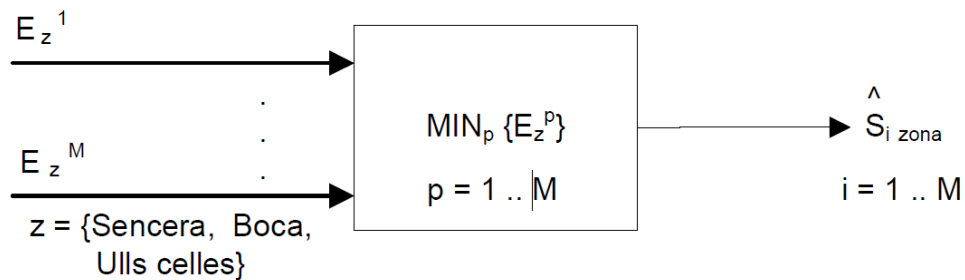


Figura 5-17

Com mostrem a la Figura 5-17 cada bloc de decisió de zona té com a dades d'entrada el paràmetre de la zona corresponent per a un model AAM de decisió. Notar que segons la zona el paràmetre serà error promig,  $E_m$ , o error quadràtic promig,  $E_{qm}$ .

Proposarem un criteris de decisió de zona:

- **Criteri de zona MÍNIM**



Aquest criteri de decisió per zona té una certa semblança amb el criteri MAP de decisió, que diu:

*Decideix que s'ha transmès el símbol  $s_j$ , si havent rebut  $y(t)$  la probabilitat d'haver transmès el símbol  $s_j$  és màxima.*

En el nostre cas, doncs, prendrem el criteri de decisió següent:

*Decideix que la imatge d'entrada té l'emoció  $S_i$ , quan la imatge sintetitzada del model AAM  $S_i$  és la de major similitud o, el que és el mateix, la de mínim error.*

De forma més concreta, la decisió dins del bloc per zona escollirà la síntesis del model AAM que tingui el menor valor del paràmetre, per a la imatge zona corresponent.

### 5.3.2.2.2 Decisió final

Després de la decisió per zona passarem al bloc de decisió final (Figura 5-18).

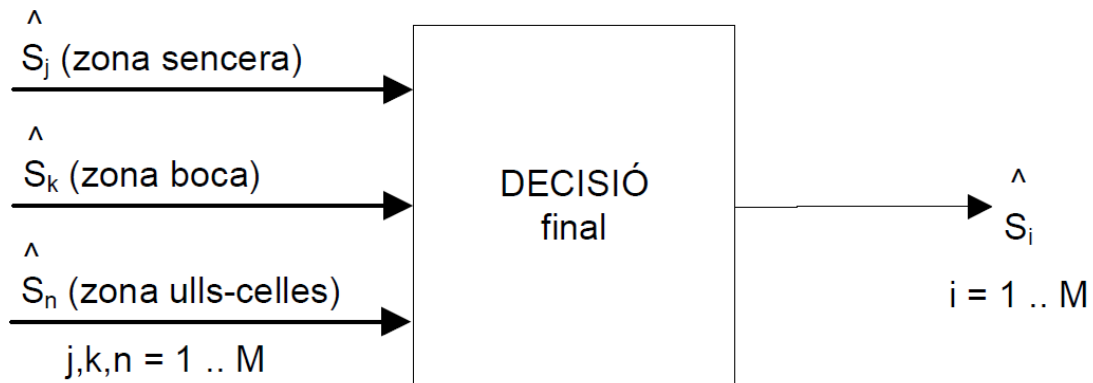


Figura 5-18

En aquest punt haurem de proposar l'emoció de la imatge d'entrada tenint com a referència les emocions que s'han decidit per a cada tipus de zona de la imatge.

Proposarem diversos criteris de decisió final:

- **Criteri final EMOCIÓ ZONA sencera, boca, ulls celles.**

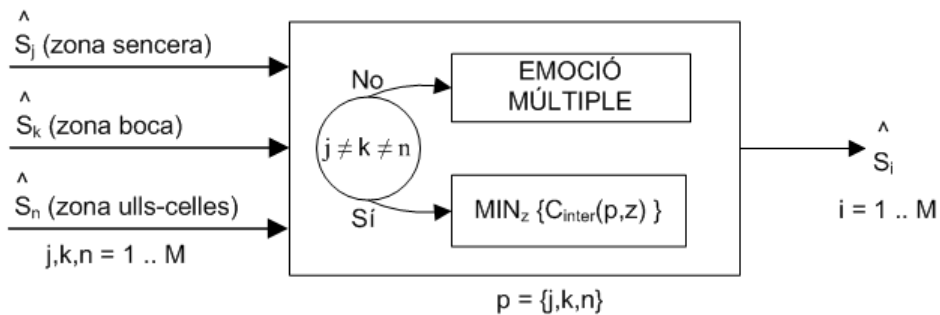
El criteri final EMOCIÓ ZONA és el criteri que es basa en la decisió més trivial, escollirem l'emoció  $S_i$  d'una zona concreta.

- **Criteri final EMOCIÓ MÚLTIPLE sencera, boca, ulls celles.**

El criteri final EMOCIÓ MÚLTIPLE seguirà la següent regla:

*Si més de dues zones tenen la mateixa emoció estimada escollirem aquesta. En el cas de que cap de les tres zones estimi la mateixa escollirem l'emoció d'una zona concreta.*

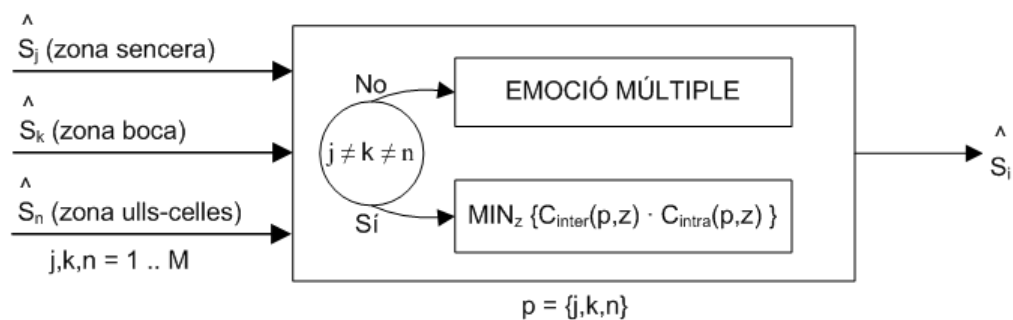
- **Criteri final EMOCIÓ MÚLTIPLE aplicant coeficients inter-zonals.**



Aquest criteri final és idèntic al criteri EMOCIÓ MÚLTIPLE excepte en el següent punt:

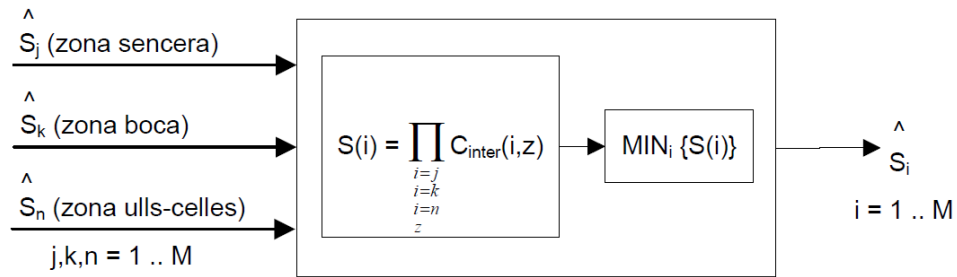
*En el cas de que cap de les tres zones estimi la mateixa introduïrem els coeficients inter-zonals i aplicarem criteri MÍNIM. Escollirem l'emoció amb menor valor de coeficients.*

- **Criteri final EMOCIÓ MÚLTIPLE aplicant coeficients intra i inter-zonals.**



Aquest criteri és idèntic al criteri EMOCIÓ MÚLTIPLE inter-zonals però aplicant també els coeficients intra-zonals.

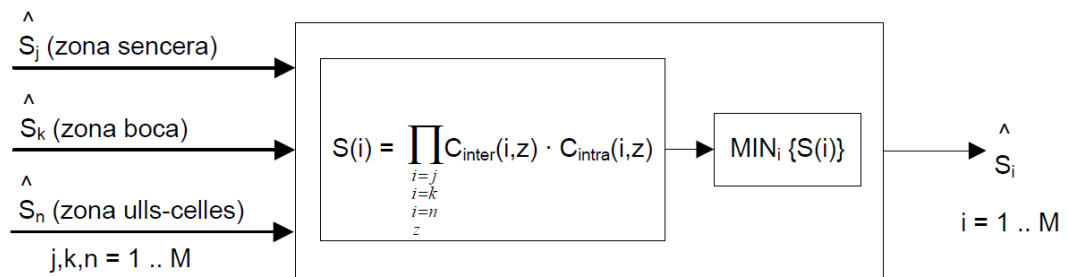
- **Criteri final MÍNIM aplicant coeficients inter-zonals**



Aquest criteri de decisió aplica el concepte de MAP als valors dels coeficients inter-zonals:

*Decideix que la imatge d'entrada té l'emoció  $S_i$ , quan el valor dels coeficients aplicats segons el model AAM  $S_i$  i la zona és el menor.*

- **Criteri final MÍNIM aplicant coeficients intra i inter-zonals**



Aquest criteri segueix la mateixa regla que el criteri final MAP però aplicant tant els coeficients intra-zonals com els inter-zonals.



## 5.3.2.2.3 Decisió completa

Una vegada definits aquest criteri de zona i els criteris finals definirem criteris complets combinant els anteriors:

Criteri complet		Criteri zona	Criteri final	
EMOCIÓ ZONA	Sencera	MÍNIM	EMOCIÓ ZONA	Sencera
	Boca			Boca
	Ulls celles			Ulls celles
EMOCIÓ MÚLTIPLE	Sencera	MÍNIM	EMOCIÓ MÚLTIPLE	Sencera
	Boca			Boca
	Ulls celles			Ulls celles
EMOCIÓ MÚLTIPLE Inter		MÍNIM	EMOCIÓ MÚLTIPLE Inter	
EMOCIÓ MÚLTIPLE Intra Inter		MÍNIM	EMOCIÓ MÚLTIPLE Intra Inter	
MÍNIM Inter		MÍNIM	MÍNIM Inter	
MÍNIM Intra Inter		MÍNIM	MÍNIM Intra Inter	

Taula 5-3

### 5.3.2.3 Estudi dels criteris de decisió

En aquest punt farem l'estudi dels diversos criteris complets definits a la Taula 5-3.

També direm que tots els criteris els avaluarem per a diversos grups de models de decisió. Els models de decisió són aquells models de les emocions sobre les que decidirem. Si una emoció no està en els models de decisió mai la podrem decidir.

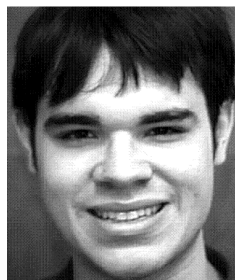
Definim, ara doncs, els diversos grups de models de decisió.

#### *Grup de decisió bàsic.*

El grup de decisió bàsic contindrà els models corresponents als tipus d'imatges



Neutre



Alegria



Sorpresa



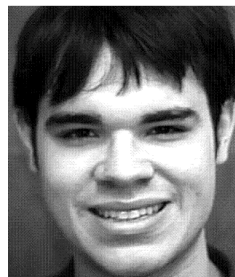
Tristesa

#### *Grup de decisió estès.*

El grup de decisió estès contindrà els models corresponents als tipus d'imatges



Neutre



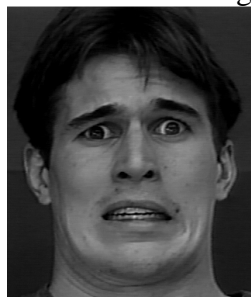
Alegria



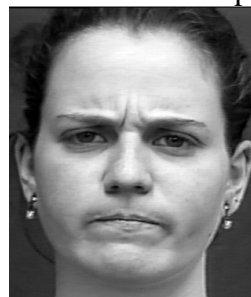
Sorpresa



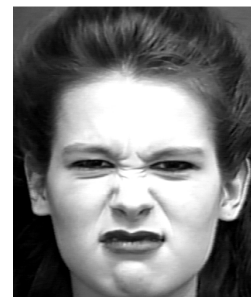
Tristesa



Por



Enuig



Fàstic

Una vegada implementat ho apliquem a les proves fetes i utilitzades anteriorment obtenint els resultats següents:

#### Taxa d'error

	Grup de decisió bàsic	Grup de decisió estès
<b>Criteri EMOCIÓ ZONA Sencera</b>	11,36%	29,60%
<b>Criteri EMOCIÓ ZONA Boca</b>	8,52%	32,29%
<b>Criteri EMOCIÓ ZONA Ulls Celles</b>	38,07%	49,78%
<b>Criteri EMOCIÓ MÚLTIPLE Sencera</b>	9,09%	27,35%
<b>Criteri EMOCIÓ MÚLTIPLE Boca</b>	6,82%	25,56%
<b>Criteri EMOCIÓ MÚLTIPLE Ulls Celles</b>	9,66%	30,49%
<b>Criteri EMOCIÓ MÚLTIPLE Inter</b>	7,95%	26,91%
<b>Criteri EMOCIÓ MÚLTIPLE Intra Inter</b>	7,95%	26,91%
<b>Criteri MÍNIM Inter</b>	9,66%	31,84%
<b>Criteri MÍNIM Intra Inter</b>	8,52%	31,39%

Taula 5-4

Detallem a continuació els resultats obtinguts en funció de les emocions d'entrada i dels models AAM dels grups de decisió.

## 5.3.2.3.1 Criteris EMOCIÓ ZONA

Criteri EMOCIÓ ZONA Sencera				
	Alegria	Neutre	Tristesa	Sorpresa
Alegria	87,18%	7,69%	5,13%	0,00%
Neutre	0,00%	92,19%	7,81%	0,00%
Tristesa	0,00%	16,67%	80,00%	3,33%
Sorpresa	4,65%	0,00%	4,65%	90,70%

Criteri EMOCIÓ ZONA Boca				
	Alegria	Neutre	Tristesa	Sorpresa
Alegria	97,44%	0,00%	0,00%	2,56%
Neutre	0,00%	87,50%	12,50%	0,00%
Tristesa	0,00%	13,33%	80,00%	6,67%
Sorpresa	0,00%	0,00%	0,00%	100,00%

Criteri EMOCIÓ ZONA Ulls celles				
	Alegria	Neutre	Tristesa	Sorpresa
Alegria	35,90%	48,72%	10,26%	5,13%
Neutre	0,00%	76,56%	9,38%	14,06%
Tristesa	0,00%	26,67%	56,67%	16,67%
Sorpresa	4,65%	13,95%	13,95%	67,44%

Taula 5-5 Grup de decisió bàsic

Criteri EMOCIÓ ZONA Sencera							
	Enuig	Fàstic	Por	Alegria	Neutre	Tristesa	Sorpresa
Enuig	31,25%	6,25%	6,25%	0,00%	50,00%	6,25%	0,00%
Fàstic	7,14%	71,43%	14,29%	0,00%	7,14%	0,00%	0,00%
Por	0,00%	5,88%	88,24%	0,00%	0,00%	5,88%	0,00%
Alegria	0,00%	0,00%	56,41%	38,46%	0,00%	5,13%	0,00%
Neutre	1,56%	1,56%	14,06%	0,00%	81,25%	1,56%	0,00%
Tristesa	3,33%	0,00%	13,33%	0,00%	13,33%	70,00%	0,00%
Sorpresa	0,00%	0,00%	2,33%	2,33%	0,00%	4,65%	90,70%

Criteri EMOCIÓ ZONA Boca							
	Enuig	Fàstic	Por	Alegria	Neutre	Tristesa	Sorpresa
Enuig	31,25%	43,75%	0,00%	0,00%	25,00%	0,00%	0,00%
Fàstic	7,14%	57,14%	14,29%	0,00%	21,43%	0,00%	0,00%
Por	0,00%	5,88%	82,35%	5,88%	5,88%	0,00%	0,00%
Alegria	0,00%	2,56%	43,59%	53,85%	0,00%	0,00%	0,00%
Neutre	4,69%	18,75%	6,25%	0,00%	62,50%	7,81%	0,00%
Tristesa	6,67%	6,67%	16,67%	0,00%	0,00%	70,00%	0,00%
Sorpresa	0,00%	0,00%	2,33%	0,00%	0,00%	0,00%	97,67%

Criteri EMOCIÓ ZONA Ulls celles							
	Enuig	Fàstic	Por	Alegria	Neutre	Tristesa	Sorpresa
Enuig	31,25%	12,50%	0,00%	0,00%	37,50%	18,75%	0,00%
Fàstic	21,43%	35,71%	14,29%	0,00%	21,43%	7,14%	0,00%
Por	5,88%	0,00%	35,29%	5,88%	35,29%	5,88%	11,76%
Alegria	17,95%	2,56%	20,51%	25,64%	25,64%	2,56%	5,13%
Neutre	6,25%	1,56%	7,81%	0,00%	65,63%	6,25%	12,50%
Tristesa	3,33%	3,33%	10,00%	0,00%	23,33%	50,00%	10,00%
Sorpresa	2,33%	0,00%	6,98%	0,00%	11,63%	11,63%	67,44%

Taula 5-6 Grup de decisió estès

## 5.3.2.3.2 Criteris EMOCIÓ MÚLTIPLE

Criteri EMOCIÓ MÚLTIPLE Sencera				
	Alegria	Neutre	Tristesa	Sorpresa
Alegria	87,18%	7,69%	5,13%	0,00%
Neutre	0,00%	95,31%	4,69%	0,00%
Tristesa	0,00%	16,67%	80,00%	3,33%
Sorpresa	2,33%	0,00%	2,33%	95,35%

Criteri EMOCIÓ MÚLTIPLE Boca				
	Alegria	Neutre	Tristesa	Sorpresa
Alegria	89,74%	7,69%	0,00%	2,56%
Neutre	0,00%	95,31%	4,69%	0,00%
Tristesa	0,00%	13,33%	83,33%	3,33%
Sorpresa	0,00%	0,00%	0,00%	100,00%

Criteri EMOCIÓ MÚLTIPLE Ulls celles				
	Alegria	Neutre	Tristesa	Sorpresa
Alegria	87,18%	12,82%	0,00%	0,00%
Neutre	0,00%	93,75%	3,13%	3,13%
Tristesa	0,00%	13,33%	80,00%	6,67%
Sorpresa	0,00%	4,65%	0,00%	95,35%

Taula 5-7 Grup de decisió bàsic

Criteri EMOCIÓ MÚLTIPLE Sencera							
	Enuig	Fàstic	Por	Alegria	Neutre	Tristesa	Sorpresa
Enuig	31,25%	12,50%	6,25%	0,00%	43,75%	6,25%	0,00%
Fàstic	7,14%	71,43%	14,29%	0,00%	7,14%	0,00%	0,00%
Por	0,00%	5,88%	82,35%	0,00%	5,88%	5,88%	0,00%
Alegria	0,00%	0,00%	51,28%	43,59%	0,00%	5,13%	0,00%
Neutre	1,56%	1,56%	10,94%	0,00%	84,38%	1,56%	0,00%
Tristesa	3,33%	0,00%	13,33%	0,00%	13,33%	70,00%	0,00%
Sorpresa	0,00%	0,00%	0,00%	2,33%	0,00%	2,33%	95,35%

Criteri EMOCIÓ MÚLTIPLE Boca							
	Enuig	Fàstic	Por	Alegria	Neutre	Tristesa	Sorpresa
Enuig	37,50%	18,75%	0,00%	0,00%	43,75%	0,00%	0,00%
Fàstic	7,14%	57,14%	14,29%	0,00%	21,43%	0,00%	0,00%
Por	0,00%	5,88%	88,24%	0,00%	5,88%	0,00%	0,00%
Alegria	0,00%	2,56%	41,03%	56,41%	0,00%	0,00%	0,00%
Neutre	4,69%	7,81%	4,69%	0,00%	79,69%	3,13%	0,00%
Tristesa	3,33%	0,00%	13,33%	0,00%	10,00%	73,33%	0,00%
Sorpresa	0,00%	0,00%	2,33%	0,00%	0,00%	0,00%	97,67%

Criteri EMOCIÓ MÚLTIPLE Ulls celles							
	Enuig	Fàstic	Por	Alegria	Neutre	Tristesa	Sorpresa
Enuig	31,25%	12,50%	0,00%	0,00%	43,75%	12,50%	0,00%
Fàstic	7,14%	57,14%	28,57%	0,00%	7,14%	0,00%	0,00%
Por	0,00%	5,88%	82,35%	0,00%	5,88%	0,00%	5,88%
Alegria	10,26%	0,00%	35,90%	41,03%	12,82%	0,00%	0,00%
Neutre	3,13%	3,13%	6,25%	0,00%	78,13%	3,13%	6,25%
Tristesa	3,33%	0,00%	13,33%	0,00%	10,00%	70,00%	3,33%
Sorpresa	0,00%	0,00%	0,00%	0,00%	4,65%	0,00%	95,35%

Taula 5-8 Grup de decisió estès

## 5.3.2.3.3 Criteris EMOCIÓ MÚLTIPLE aplicant coeficients

Criteri EMOCIÓ MÚLTIPLE Inter				
	Alegria	Neutre	Tristesa	Sorpresa
Alegria	87,18%	7,69%	2,56%	2,56%
Neutre	0,00%	95,31%	4,69%	0,00%
Tristesa	0,00%	16,67%	80,00%	3,33%
Sorpresa	0,00%	0,00%	0,00%	100,00%

Criteri EMOCIÓ MÚLTIPLE Intra Inter				
	Alegria	Neutre	Tristesa	Sorpresa
Alegria	87,18%	7,69%	2,56%	2,56%
Neutre	0,00%	95,31%	4,69%	0,00%
Tristesa	0,00%	16,67%	80,00%	3,33%
Sorpresa	0,00%	0,00%	0,00%	100,00%

Taula 5-9 Grup de decisió bàsic

Criteri EMOCIÓ MÚLTIPLE Inter							
	Enuig	Fàstic	Por	Alegria	Neutre	Tristesa	Sorpresa
Enuig	31,25%	12,50%	6,25%	0,00%	43,75%	6,25%	0,00%
Fàstic	7,14%	71,43%	14,29%	0,00%	7,14%	0,00%	0,00%
Por	0,00%	5,88%	88,24%	0,00%	5,88%	0,00%	0,00%
Alegria	0,00%	0,00%	56,41%	41,03%	0,00%	2,56%	0,00%
Neutre	1,56%	1,56%	10,94%	0,00%	84,38%	1,56%	0,00%
Tristesa	3,33%	0,00%	13,33%	0,00%	13,33%	70,00%	0,00%
Sorpresa	0,00%	0,00%	2,33%	0,00%	0,00%	0,00%	97,67%

Criteri EMOCIÓ MÚLTIPLE Intra Inter							
	Enuig	Fàstic	Por	Alegria	Neutre	Tristesa	Sorpresa
Enuig	31,25%	12,50%	6,25%	0,00%	43,75%	6,25%	0,00%
Fàstic	7,14%	71,43%	14,29%	0,00%	7,14%	0,00%	0,00%
Por	0,00%	5,88%	88,24%	0,00%	5,88%	0,00%	0,00%
Alegria	0,00%	0,00%	56,41%	41,03%	0,00%	2,56%	0,00%
Neutre	1,56%	1,56%	10,94%	0,00%	84,38%	1,56%	0,00%
Tristesa	3,33%	0,00%	13,33%	0,00%	13,33%	70,00%	0,00%
Sorpresa	0,00%	0,00%	2,33%	0,00%	0,00%	0,00%	97,67%

Taula 5-10 Grup de decisió estès

## 5.3.2.3.4 Criteris del tipus MÍNIM

Criteri MÍNIM Inter				
	Alegria	Neutre	Tristesa	Sorpresa
Alegria	87,18%	7,69%	2,56%	2,56%
Neutre	0,00%	92,19%	7,81%	0,00%
Tristesa	0,00%	16,67%	76,67%	6,67%
Sorpresa	0,00%	0,00%	0,00%	100,00%

Criteri MÍNIM Intra Inter				
	Alegria	Neutre	Tristesa	Sorpresa
Alegria	87,18%	7,69%	2,56%	2,56%
Neutre	0,00%	95,31%	4,69%	0,00%
Tristesa	0,00%	16,67%	76,67%	6,67%
Sorpresa	0,00%	0,00%	0,00%	100,00%

Taula 5-11 Grup de decisió bàsic

Criteri MÍNIM Inter							
	Enuig	Fàstic	Por	Alegria	Neutre	Tristesa	Sorpresa
Enuig	12,50%	25,00%	6,25%	0,00%	50,00%	6,25%	0,00%
Fàstic	7,14%	71,43%	14,29%	0,00%	7,14%	0,00%	0,00%
Por	0,00%	5,88%	94,12%	0,00%	0,00%	0,00%	0,00%
Alegria	0,00%	0,00%	69,23%	28,21%	0,00%	2,56%	0,00%
Neutre	0,00%	3,13%	15,63%	0,00%	79,69%	1,56%	0,00%
Tristesa	3,33%	0,00%	16,67%	0,00%	13,33%	66,67%	0,00%
Sorpresa	0,00%	0,00%	2,33%	0,00%	0,00%	0,00%	97,67%

Criteri MÍNIM Intra Inter							
	Enuig	Fàstic	Por	Alegria	Neutre	Tristesa	Sorpresa
Enuig	12,50%	25,00%	6,25%	0,00%	50,00%	6,25%	0,00%
Fàstic	7,14%	71,43%	14,29%	0,00%	7,14%	0,00%	0,00%
Por	0,00%	5,88%	94,12%	0,00%	0,00%	0,00%	0,00%
Alegria	0,00%	0,00%	69,23%	28,21%	0,00%	2,56%	0,00%
Neutre	0,00%	3,13%	14,06%	0,00%	81,25%	1,56%	0,00%
Tristesa	3,33%	0,00%	16,67%	0,00%	13,33%	66,67%	0,00%
Sorpresa	0,00%	0,00%	2,33%	0,00%	0,00%	0,00%	97,67%

Taula 5-12 Grup de decisió estès

5.3.2.4 Criteri de decisió final

En aquest moment hem de seleccionar en base a les dades anteriors el criteri que escollirem com a òptim.

**Taxa d'error**

	Grup de decisió bàsic	Grup de decisió estès
Criteri EMOCIÓ ZONA Sencera	11,36%	29,60%
Criteri EMOCIÓ ZONA Boca	<b>8,52%</b>	32,29%
Criteri EMOCIÓ MÚLTIPLE Sencera	9,09%	<b>27,35%</b>
Criteri EMOCIÓ MÚLTIPLE Boca	<b>6,82%</b>	<b>25,56%</b>
Criteri EMOCIÓ MÚLTIPLE Ulls Celles	9,66%	30,49%
Criteri EMOCIÓ MÚLTIPLE Inter	<b>7,95%</b>	<b>26,91%</b>
Criteri EMOCIÓ MÚLTIPLE Intra Inter	7,95%	26,91%
Criteri MÍNIM Inter	9,66%	31,84%
Criteri MÍNIM Intra Inter	<b>8,52%</b>	31,39%

Taula 5-13

En aquesta taula podem veure marcats en negreta els tres valors de la taxa d'error mínims per a ambdós grups de decisió.

Grup de decisió bàsic		Grup de decisió estès	
Criteri	Taxa d'error	Criteri	Taxa d'error
Criteri EMOCIÓ ZONA boca	8,52%	Criteri EMOCIÓ MÚLTIPLE Sencera	27,35%
<b>Criteri EMOCIÓ MÚLTIPLE Boca</b>	<b>6,82%</b>	<b>Criteri EMOCIÓ MÚLTIPLE Boca</b>	<b>25,56%</b>
<b>Criteri EMOCIÓ MÚLTIPLE Inter</b>	<b>7,95%</b>	<b>Criteri EMOCIÓ MÚLTIPLE Inter</b>	<b>26,91%</b>
Criteri MÍNIM Intra Inter	8,52%		

Veiem, doncs, que per als dos grups de decisió hi ha dos criteris que tenen l'error mínim: Criteri EMOCIÓ MÚLTIPLE Boca i Criteri EMOCIÓ MÚLTIPLE Inter. Donat que el que busquem és el mínim error a la decisió escollirem com a criteri de decisió el **Criteri EMOCIÓ MÚLTIPLE Boca**.



## 5.4 Implementació

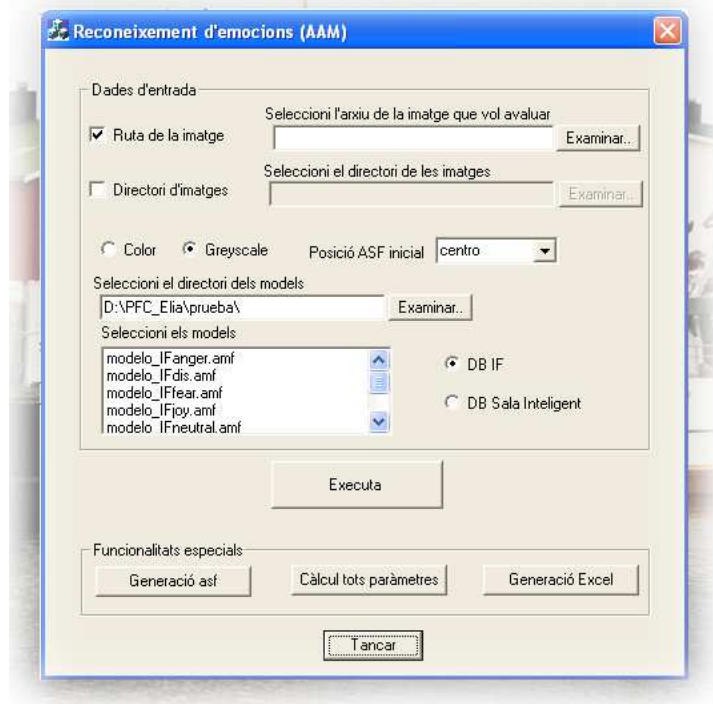
El desenvolupament d'aquest projecte s'ha dut a terme mitjançant la implementació de dues aplicacions: la primera, *Reconeixement d'emocions (AAM)*, basada en el llenguatge Visual C++ 6.0 i la segona, *ENG PFC: Estudi Reconeixement d'emocions*, implementada en llenguatge Visual Basic .NET.

Les operacions principals que duu a terme l'aplicació *Reconeixement d'emocions (AAM)* són la síntesi de diversos models AAM respecte a una o varies imatges d'entrada, el càlcul dels paràmetres de cada síntesi i la decisió òptima.

L'aplicació *ENG PFC: Estudi Reconeixement d'emocions* es va construir amb l'objectiu d'analitzar els paràmetres i els mètodes de decisió.

### 5.4.1 Reconeixement d'emocions (AAM)

Mostrem a continuació una figura de la interfase d'aquesta aplicació:



Com es pot veure, l'aplicació té dues parts:

- L'execució del **reconeixement d'emocions**.
- Les **funcionalitats especials**: Generació dels arxius de control de forma ASF (*Generació asf*), el càlcul de paràmetres (*Càlcul tots paràmetres*) i la generació de l'arxiu de resum dels resultats en format Excel (*Generació Excel*).

#### 5.4.1.1 Execució del reconeixement d'emocions

Amb aquesta interfase executarem tot el procés del sistema de reconeixement d'emocions basat en els models AAM.

Les dades d'entrada que tindrem són:

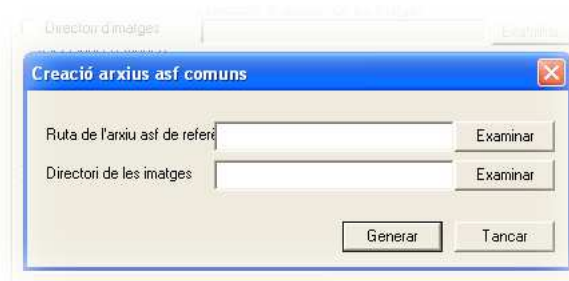
- Imatges de test. Podrem escollir entre fer un test o fer tants tests com imatges d'un directori.
- Tipus de color: color o nivells de gris.
- Models sobre els que farem la decisió dels tests.

Una vegada iniciem el test del sistema els passos que seguim són els següents:

- Execució de la síntesi AAM de cada model escollit. Aquest conjunt d'accions les hem encapsulat a la classe *CGestorAAM*.
- Càlcul dels paràmetres de decisió. També encapsulat a la classe *CParametros*. El càlcul dels paràmetres l'hem implementat mitjançant les operacions contingudes a la llibreria *OpenCV* de processat d'imatge.
- Decisió de l'emoció estimada de la imatge d'entrada. Encapsulat a la classe *CDecision*.

#### 5.4.1.2 Generació asf

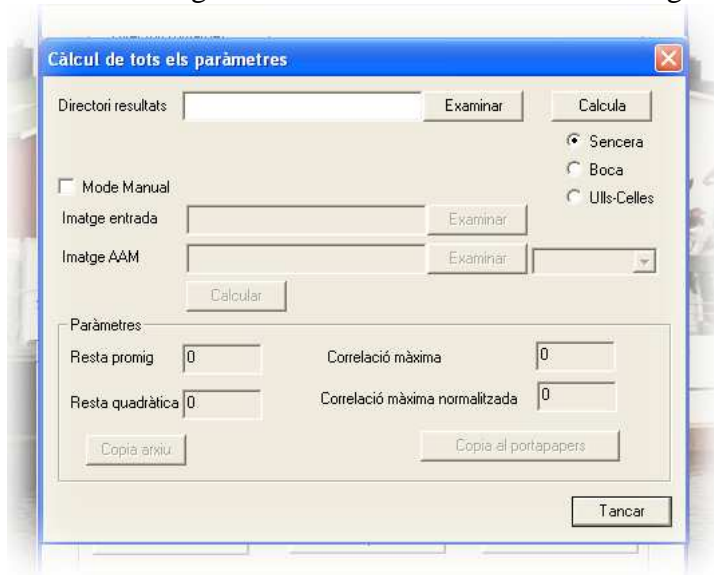
Aquesta utilitat ens serà útil per crear els arxius de punts de controls de referència de cada imatge de test d'entrada com vàrem explicar al punt 5.2.2.2 Síntesi d'imatge mitjançant models AAM.



Amb aquesta utilitat anirem creant arxius .asf com el de la Figura 5-6 (Arxiu ASF amb els punts de control) amb el nom de la imatge corresponent i amb el nom de l'arxiu igual que el nom de la imatge de test.

#### 5.4.1.3 Càlcul de paràmetres

Amb aquesta utilitat podrem fer el càlcul dels paràmetres de forma manual. Escollirem el directori on estan els arxius generats com a resultats de cada imatge.



Podrem calcular els paràmetres per les diverses zones per separat. Podrem fer-ho per una sola imatge d'entrada i per una sola imatge de síntesi d'un model. I podrem veure els resultats al mateix temps.

#### 5.4.1.4 Generació Excel



Amb aquesta utilitat generarem l'arxiu amb els resultats dels tests executats. Escollirem el directori on estan els arxius generats com a resultats de cada imatge.

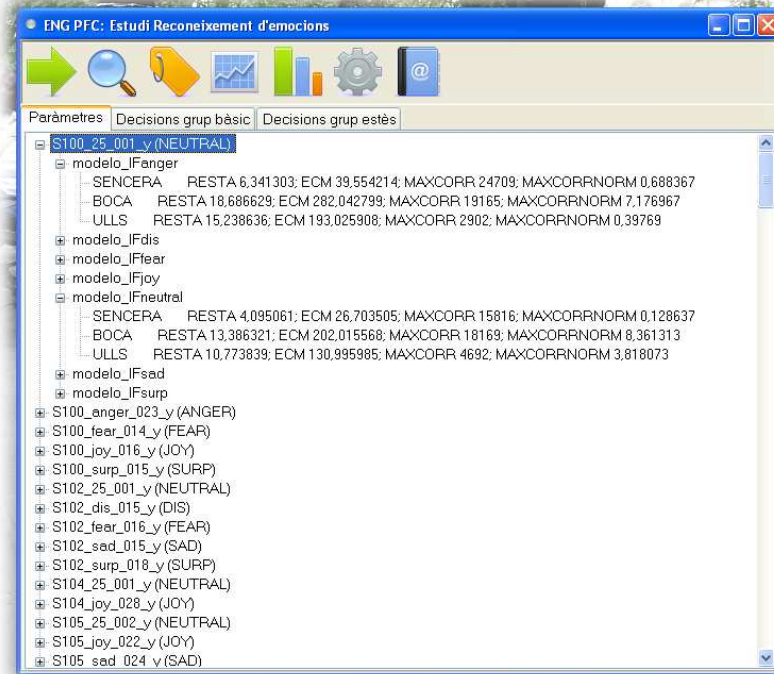
### 5.4.2 ENG PFC: Estudi Reconeixement d'emocions

L'objectiu d'aquesta aplicació és l'estudi dels diversos paràmetres i la recerca i estudi del mètode de decisió òptim.

Tenim tres pestanyes on es mostren diversos tipus de dades:

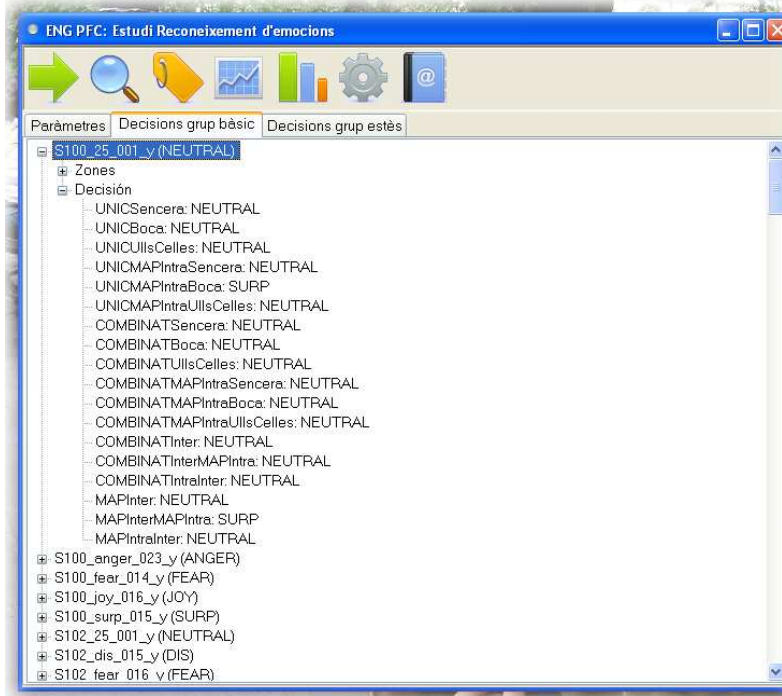
- Paràmetres

Les dades que es mostren són els resultats dels càlculs dels paràmetres de decisió per cada prova amb tots els models.

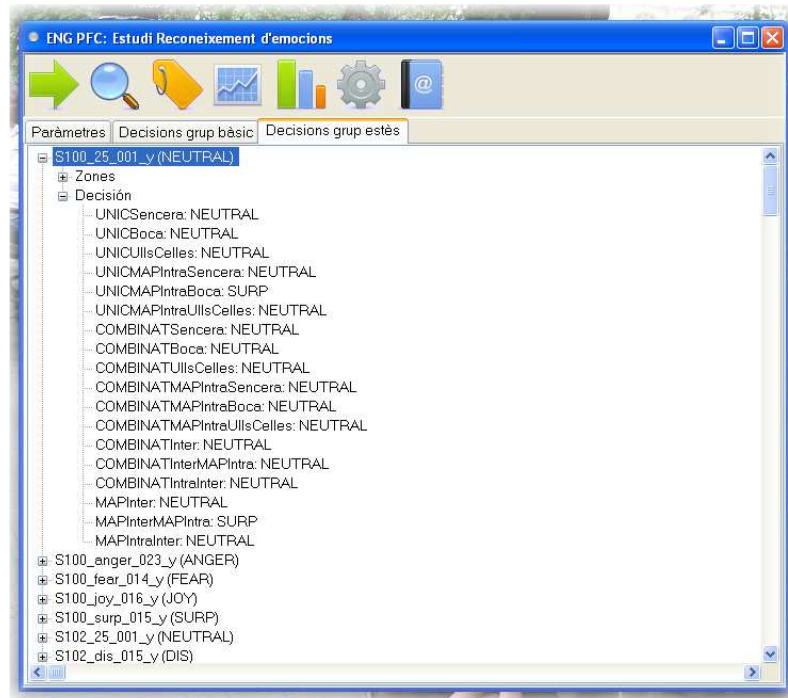


- Decisions grup bàsic

Les dades que es mostren són els resultats de les decisions per zona i les decisions proposades al llarg de l'estudi. Totes elles per el grup de decisió bàsic.



- Decisions grup estès  
Les dades que es mostren són els resultats de les decisions per zona i les decisions proposades al llarg de l'estudi. Totes elles per el grup de decisió estès.



Accions implementades:



- 1) Importar dades des de Excel
- 2) Inicialitzar els tipus d'imatges i models
- 3) Configuració de l'estudi de dades
- 4) Generar els estudis de dades
- 5) Decisió per zones
- 6) Criteris de decisió final
- 7) Resum de les decisions

#### 5.4.2.1 Importar dades des de Excel

Aquesta utilitat la farem servir per importar les dades provinents de l'aplicació *Reconeixement d'emocions (AAM)*. Des de l'aplicació *Reconeixement d'emocions (AAM)* poder exportar les dades a Excel en una estructura coneguda per ambdues aplicacions. La informació sobre l'estructura d'aquest arxiu s'ha explicat al punt anterior 5.4.1 Reconeixement d'emocions (AAM).

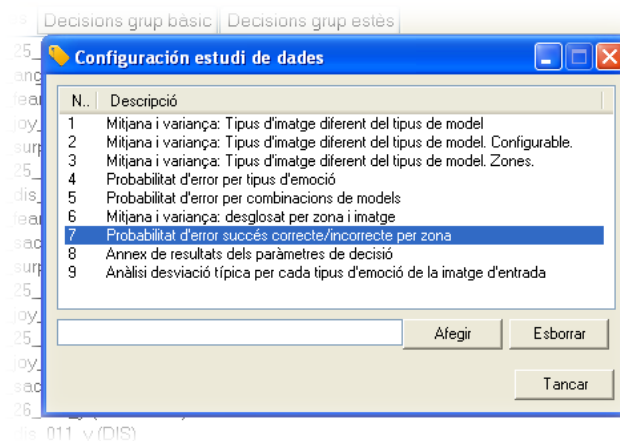
#### 5.4.2.2 Inicialitzar els tipus d'imatges i models

Una vegada importades les dades de l'arxiu Excel amb l'acció anterior haurem de definir quin tipus d'emoció conté la imatge original de cada prova per poder, després, dur a terme els pertinents càlculs de les probabilitats d'error.

#### 5.4.2.3 Configuració de l'estudi de dades

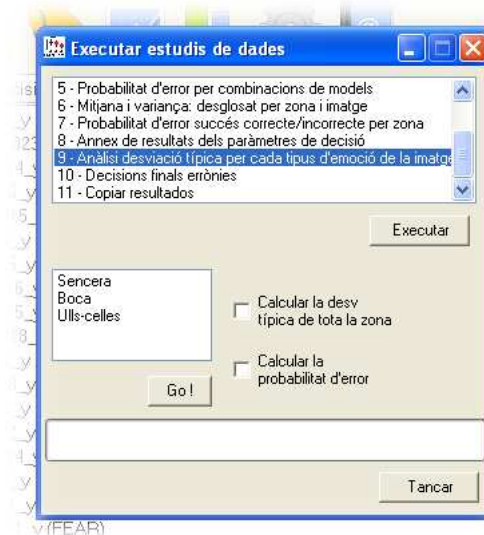
Amb aquesta utilitat configurarem els tipus d'estudis de dades per a poder analitzar els diferents paràmetres. Els resultats d'aquests estudis de dades els poden trobar al punt

##### 5.3.1.3 *Estudi Paràmetres Decisió.*



#### 5.4.2.4 Generar els estudis de dades

A partir dels tipus configurats l'executarem i es generaran els resultats en un arxiu Excel.

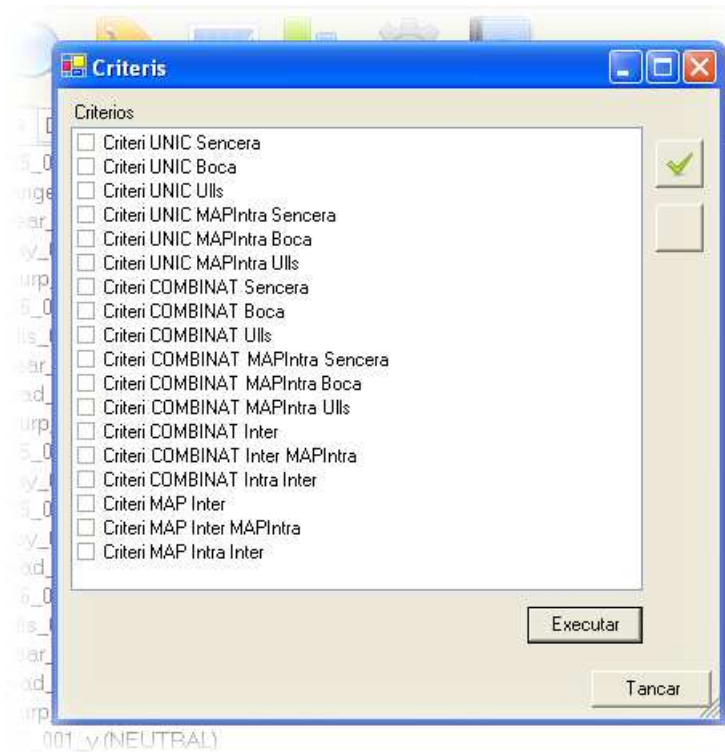


#### 5.4.2.5 Decisió per zones

Segons el que s'ha exposat al punt 5.3.2.2.1 Decisió per zona executarem la decisió de les zones. Una vegada executada la desarem en un arxiu XML i mostrarem els resultats a les pestanyes de decisions de la pantalla principal.

#### 5.4.2.6 Criteris de decisió final

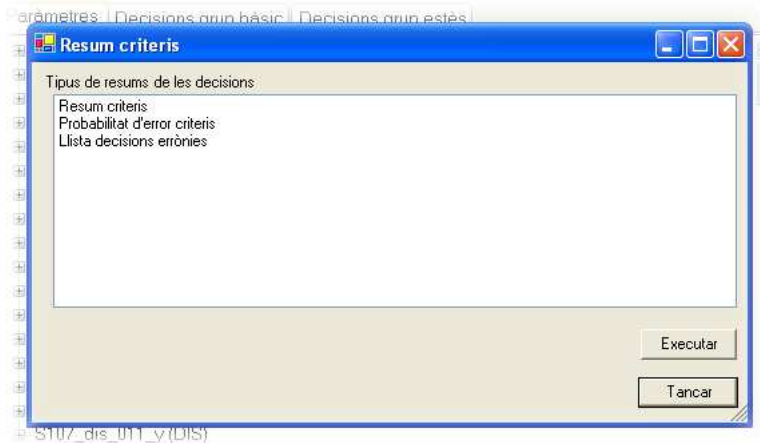
Amb aquesta acció executarem els diversos criteris definits al punt 5.3.2.2 Proposta criteris de decisió.



Els resultats obtinguts els desarem en un arxiu XML i actualitzarem les dades de les pestanyes de Decisions de la pantalla principal.

5.4.2.7 Resum de les decisions

És una utilitat per extreure resums diversos amb els resultats dels diversos criteris abans executats.



En el tipus de *resum criteris* extraurem el detall del número dels tests amb una emoció estimada segons l'emoció estimada i l'emoció d'entrada. Un exemple és:

Criteri UNIC Boca					
	Alegria	Neutral	Tristesa	Sorpresa	Total
Alegria	38	0	0	1	39
Neutral	0	56	8	0	64
Tristesa	0	4	24	2	30
Sorpresa	0	0	0	43	43

En el tipus de *probabilitat d'error criteris* extraurem el detall de la probabilitat d'error dels tests segons l'emoció estimada i l'emoció d'entrada. Un exemple és:

Criteri UNIC Boca				
	Alegria	Neutral	Tristesa	Sorpresa
Alegria	97,44%	0,00%	0,00%	2,56%
Neutral	0,00%	87,50%	12,50%	0,00%
Tristesa	0,00%	13,33%	80,00%	6,67%
Sorpresa	0,00%	0,00%	0,00%	100,00%



## 6 Resultats

Els resultats obtinguts basats en las imatges de la **base de dades CMU** són les següents:

	Neutre	Alegria	Sorpresa	Tristesa	Fàstic	Enuig	Por
Neutre	79,69%	0,00%	0,00%	3,13%	7,81%	4,69%	4,69%
Alegria	0,00%	56,41%	0,00%	0,00%	2,56%	0,00%	41,03%
Sorpresa	0,00%	0,00%	97,67%	0,00%	0,00%	0,00%	2,33%
Tristesa	10,00%	0,00%	0,00%	73,33%	0,00%	3,33%	13,33%
Fàstic	21,43%	0,00%	0,00%	0,00%	57,14%	7,14%	14,29%
Enuig	43,75%	0,00%	0,00%	0,00%	18,75%	37,50%	0,00%
Por	5,88%	0,00%	0,00%	0,00%	5,88%	0,00%	88,24%

Probabilitat d'error 25,56%

Una mostra de les emocions en imatges és el següent:



Analitzant els resultats en detall podem enunciar que per al sistema de reconeixement:

- l'emoció d'**alegria** es confon amb l'emoció de por quasi la meitat dels casos. En canvi l'emoció de por no es confon amb la d'alegria.
- Tanmateix, l'emoció d'**enuig** es confon amb l'emoció neutre però l'emoció neutre només es decideix com a enuig un 5% aproximadament. L'enuig també es confon amb l'emoció de fàstic encara que en menor grau que amb la neutre. Observant la mostra de les emocions anterior podem veure que existeix una certa similitud que podria explicar aquesta confusió.
- De la mateixa manera, l'emoció de **fàstic** es confon amb l'emoció neutre però l'emoció neutre només es decideix com a fàstic un 8% aproximadament. El fàstic també es confon amb l'emoció de por encara que en menor grau que amb la neutre.

Per tal de poder definir les causes d'aquestes confusions mostrarem els resultats per a nous grups de decisió:

- **Test informació de cada emoció.** Els grups de decisió estaran formats per l'emoció neutre i cadascuna de les altres.
- **Test confusió.** Estaran formats per combinacions de les emocions que es confonen i que s'han enunciat anteriorment.

Mostrem, doncs, a continuació els grups de decisió per al **test informació** de cada emoció.

	Neutre	Alegria
Neutre	100,00%	0,00%
Alegria	12,82%	87,18%

Probabilitat d'error 4,85%

	Neutre	Enuig
Neutre	98,44%	1,56%
Enuig	62,50%	37,50%

Probabilitat d'error 13,75%

	Neutre	Por
Neutre	93,75%	6,25%
Por	5,88%	94,12%

Probabilitat d'error 6,17%

	Neutre	Sorpresa
Neutre	100,00%	0,00%
Sorpresa	2,33%	97,67%

Probabilitat d'error 0,93%

	Neutre	Fàstic
Neutre	93,75%	6,25%
Fàstic	28,57%	71,43%

Probabilitat d'error 10,26%

	Neutre	Tristesa
Neutre	93,75%	6,25%
Tristesa	16,67%	83,33%

Probabilitat d'error 9,57%

Podem veure que els test que donen una probabilitat d'error més elevada són els que decideixen amb **enuig** i amb **fàstic**, seguit de **tristesa**. Es vol notar en aquest punt que aquestes tres són les emocions amb més similitud amb la neutre. Així es coherent que al sistema li costi decidir correctament en el 100% dels casos les emocions que més s'assemblen.

Mostrem, doncs, a continuació els grups de decisió per al **test confusió**:

	<b>Alegria</b>	<b>Por</b>
<b>Alegria</b>	<b>48,72%</b>	51,28%
<b>Por</b>	0,00%	<b>100,00%</b>

**Probabilitat d'error 35,71%**

	<b>Neutre</b>	<b>Fàstic</b>	<b>Enuig</b>
<b>Neutre</b>	<b>89,06%</b>	7,81%	3,13%
<b>Fàstic</b>	28,57%	<b>64,29%</b>	7,14%
<b>Enuig</b>	56,25%	12,50%	<b>31,25%</b>

**Probabilitat d'error 24,47%**

	<b>Neutre</b>	<b>Fàstic</b>	<b>Por</b>
<b>Neutre</b>	<b>89,06%</b>	4,69%	6,25%
<b>Fàstic</b>	21,43%	<b>64,29%</b>	14,29%
<b>Por</b>	5,88%	5,88%	<b>88,24%</b>

**Probabilitat d'error 14,74%**

Mostrem a continuació un conjunt d'exemples on es va decidir correctament i on es va decidir erròniament.

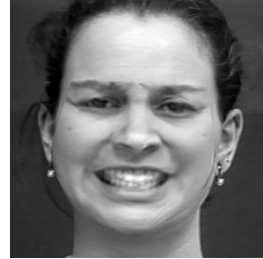
- **Imatges d'alegria**

**Imatge original**

**Imatge de síntesi  
model d'alegria**

**Imatge de síntesi  
model de por**

Decisió correcta



Decisions errònies



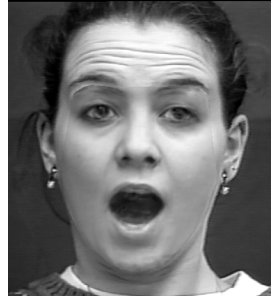
- **Imatges de sorpresa**

**Imatge original**

**Imatge de síntesi  
model de sorpresa**

**Imatge de síntesi  
model de por**

Decisions correctes



Decisió errònia



- **Imatges de tristesa**

**Imatge original**

**Imatge de síntesi  
model de tristesa**

**Imatge de síntesi  
model neutre**

**Imatge de síntesi  
model de por**

Decisió correcta



Decisions errònies (Neutre, Por)



- Imatges de fàstic

Imatge original	Imatge síntesi model fàstic	Imatge síntesi model neutre	Imatge síntesi model enuig	Imatge síntesi model de por
-----------------	-----------------------------	-----------------------------	----------------------------	-----------------------------

Decisió correcta



Decisions errònies (enuig, por, neutre)



- Imatges d'enuig

Imatge original	Imatge de síntesi model d'enuig	Imatge de síntesi model neutre	Imatge de síntesi model de fàstic
-----------------	---------------------------------	--------------------------------	-----------------------------------

Decisions correctes


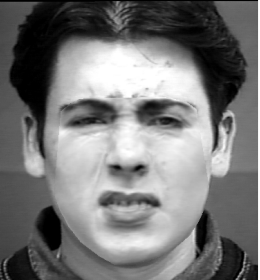



Decisions errònies (Fàstic, Neutre)





- Imatges de por

Imatge original	Imatge de síntesi model de por	Imatge de síntesi model neutre	Imatge de síntesi model de fàstic
Decisió correcta			
			
			
Decisions errònies (Neutre, Fàstic)			
			
			

Si observem els resultats podem veure que una de les causes d'una decisió errònia és que la síntesi i ajust del model a la imatge original no sigui correcta. Quantificarem a partir d'un anàlisi visual les proves que han tingut aquest efecte.

	Tests erronis	Tests amb síntesi errònia	Tests sistema Recon. Emoc. Erronis	Total tests	% tests amb síntesi errònia
Neutre	13	4	9	64	6,25%
Alegria	17	13	4	39	33,33%
Sorpresa	1	0	1	43	0,00%
Tristesa	8	5	3	30	16,67%
Fàstig	6	3	3	14	21,43%
Enuig	10	5	5	16	31,25%
Por	2	2	0	17	11,76%
	<b>57</b>	<b>32</b>	<b>25</b>	<b>223</b>	<b>14,35% Total</b>

Tornarem a calcular els encerts i la probabilitat d'error restant les senyals errònies tant del total de tests com de les decisions. Obtenim el quadre de la Figura 6-2, disposarem també per poder comparar millor el quadre inicial (Figura 6-1).

	Neutre	Alegria	Sorpresa	Tristesa	Fàstic	Enuig	Por
Neutre	<b>79,69%</b>	0,00%	0,00%	3,13%	7,81%	4,69%	4,69%
Alegria	0,00%	<b>56,41%</b>	0,00%	0,00%	2,56%	0,00%	41,03%
Sorpresa	0,00%	0,00%	<b>97,67%</b>	0,00%	0,00%	0,00%	2,33%
Tristesa	10,00%	0,00%	0,00%	<b>73,33%</b>	0,00%	3,33%	13,33%
Fàstic	21,43%	0,00%	0,00%	0,00%	<b>57,14%</b>	7,14%	14,29%
Enuig	43,75%	0,00%	0,00%	0,00%	18,75%	<b>37,50%</b>	0,00%
Por	5,88%	0,00%	0,00%	0,00%	5,88%	0,00%	<b>88,24%</b>

Probabilitat d'error 25,56%

Figura 6-1

	Neutre	Alegria	Sorpresa	Tristesa	Fàstig	Enuig	Por
Neutre	<b>85,00%</b>	0,00%	0,00%	1,67%	6,67%	5,00%	1,67%
Alegria	0,00%	<b>84,62%</b>	0,00%	0,00%	0,00%	0,00%	15,38%
Sorpresa	0,00%	0,00%	<b>97,67%</b>	0,00%	0,00%	0,00%	2,33%
Tristesa	8,00%	0,00%	0,00%	<b>88,00%</b>	0,00%	0,00%	4,00%
Fàstig	18,18%	0,00%	0,00%	0,00%	<b>72,73%</b>	0,00%	9,09%
Enuig	36,36%	0,00%	0,00%	0,00%	9,09%	<b>54,55%</b>	0,00%
Por	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	<b>100,00%</b>

Probabilitat d'error 13,09%

Figura 6-2

Podríem dir que aquests són els resultats propis del sistema de reconeixement d'emocions implementat.

Analitzant de nou aquests resultats podem enunciar que

- l'emoció d'**alegria** ja no es confon en tant alt grau amb l'emoció de por. En canvi l'emoció de por segueix sense confondre's amb la d'alegria.
- Tanmateix, l'emoció d'**enuig** es segueix confonent amb l'emoció neutre però en menor grau l'emoció neutre només es decideix com a enuig un 5%. L'enuig també es confon amb l'emoció de fàstic en un 9%.

Observant la mostra de les emocions anterior podem veure que existeix una certa similitud que podria explicar aquesta confusió.

- De la mateixa manera, l'emoció de **fàstic** es confon amb l'emoció neutre però l'emoció neutre només es decideix com a fàstic un 7% aproximadament. El fàstic també es confon amb l'emoció de por encara que en menor grau que amb la neutre.

La taxa final d'error ha baixat d'un 25,56 % a un 13,09 %.

En aquest punt mostrarem els resultats del sistema en utilitzar la base de dades d'imatges generades a la UPC.

Aquesta base de dades consta d'un conjunt de 137 imatges de 16 persones on es representen un total de 5 emocions diferents. En els següents quadres podem veure relacionades les dades:

Imatges	133	Imatges dels models	80	Imatges de test	53
Enuig	33	Enuig	20	Enuig	13
Alegria	30	Alegria	20	Alegria	10
Neutre	30	Neutre	20	Neutre	10
Tristesa	9	Tristesa	0	Trist	9
Sorpresa	31	Sorpresa	20	Sorpresa	11

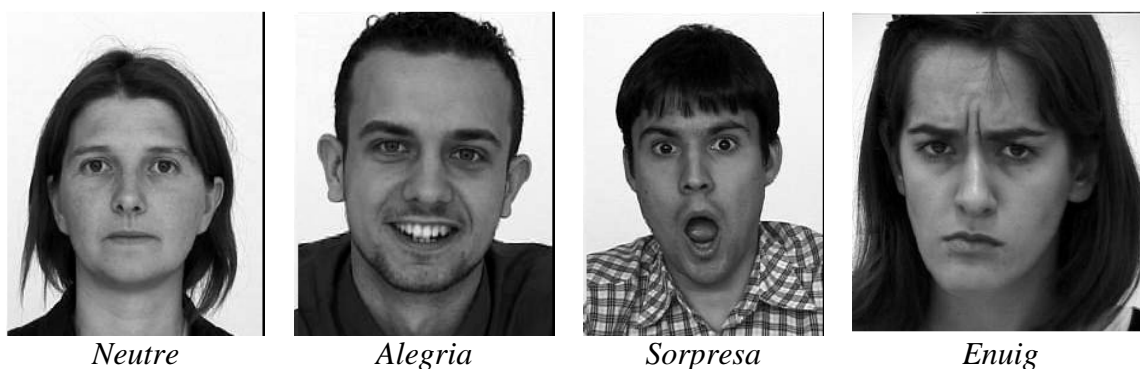
El primer quadre mostra el total de les imatges disponibles. En el quadre Imatges dels models mostrem les imatges que s'utilitzaran com a base de dades d'entrenament dels models. Les Imatges de test són les imatges que utilitzarem per testejar el sistema.

Els resultats obtinguts decidint amb els models de totes les emocions disponibles són els següents:

	Enuig	Neutre	Alegria	Sorpresa
Enuig	69,23%	23,08%	0,00%	7,69%
Neutre	20,00%	70,00%	10,00%	0,00%
Alegria	0,00%	0,00%	100,00%	0,00%
Sorpresa	9,09%	18,18%	9,09%	63,64%

Probabilitat d'error 25,00%

Una mostra de les emocions en imatges és el següent:



Analitzant els resultats en detall podem enunciar que per al sistema de reconeixement:

- Tornem a veure que l'emoció d'**enuig** es confon amb l'emoció neutre però en aquest cas l'emoció neutre també es confon amb la d'enuig.
- Tanmateix l'emoció de **sorpresa** es confon amb l'emoció neutre però en aquest cas l'emoció neutre no es confon amb la de sorpresa.

Per tal de poder definir les causes d'aquestes confusions mostrarem els resultats per a nous grups de decisió:

- **Test confusió.** Estaran formats per combinacions de les emocions que es confonen i que s'han enunciat anteriorment.
- Així com el **test informació** per a l'emoció d'alegria.

Mostrem, doncs, a continuació els grups de decisió per al **test confusió** i també el **test informació**.

	Neutre	Enuig
Neutre	80,00%	20,00%
Enuig	23,08%	76,92%

Probabilitat d'error 21,74%

	Neutre	Alegria
Neutre	90,00%	10,00%
Alegria	0,00%	100,00%

Probabilitat d'error 5,00%

	Neutre	Sorpresa
Neutre	90,00%	10,00%
Sorpresa	36,36%	63,64%

Probabilitat d'error 23,81%

Podem veure que els test que donen una probabilitat d'error més elevada són els que decideixen amb **sorpresa** i amb **enuig**.

Mostrem a continuació un conjunt d'exemples on es va decidir correctament i on es va decidir erròniament.

- **Imatges d'enuig**

<b>Imatge original</b>	<b>Imatge de síntesi model d'enuig</b>	<b>Imatge de síntesi model neutre</b>	<b>Imatge de síntesi model de sorpresa</b>
------------------------	--	---------------------------------------	--

Decisions correctes



Decisions errònies (Sorpresa, neutre)



- **Imatges neutre**

Imatge original	Imatge de síntesi model neutre	Imatge de síntesi model d'enuig	Imatge de síntesi model d'alegria
-----------------	--------------------------------	---------------------------------	-----------------------------------

Decisions correctes



Decisions errònies (Enuig, alegria)



- **Imatges d'alegria**

**Imatge original**

**Imatge de síntesi  
model d'alegria**

Decisió correcta

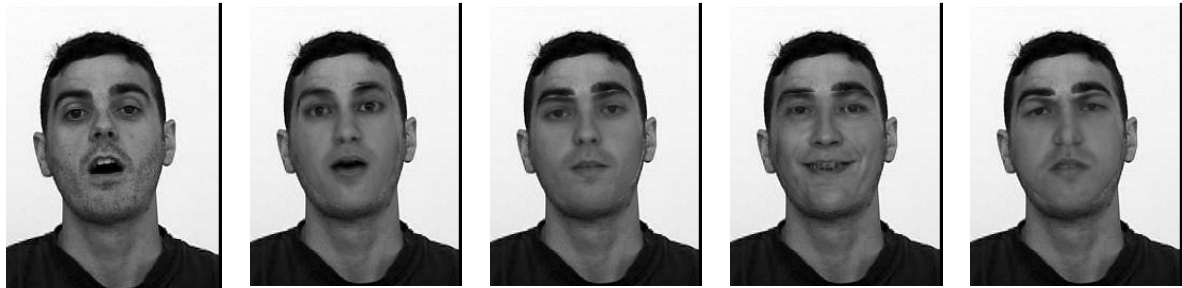




- Imatges de sorpresa

Imatge original	Imatge síntesi model sorpresa	Imatge síntesi model neutre	Imatge síntesi model alegria	Imatge síntesi model enuig
-----------------	-------------------------------	-----------------------------	------------------------------	----------------------------

Decisió correcta



Decisions errònies (alegria, neutre, enuig)



Quantificarem a partir d'un anàlisi visual de les proves les decisions errònies degudes a una síntesi i ajust del model a la imatge original no correcta i les extraurem del càlcul:

	Tests erronis	Tests amb síntesi errònia	Tests sistema Recon. Emoc. Erronis	Total tests	% tests amb síntesi errònia
<b>Enuig</b>	4	1	3	13	7,69%
<b>Neutre</b>	3	1	2	10	10,00%
<b>Alegria</b>	0	0	0	10	0,00%
<b>Sorpresa</b>	4	2	2	11	18,18%
	<b>11</b>	<b>4</b>	<b>7</b>	<b>44</b>	<b>9,09% Total</b>

Tornarem a calcular els encerts i la probabilitat d'error restant les senyals errònies tant del total de tests com de les decisions. Obtenim el quadre de la Figura 6-4, disposarem també per poder comparar millor el quadre inicial (Figura 6-3).

	<b>Enuig</b>	<b>Neutre</b>	<b>Alegria</b>	<b>Sorpresa</b>
<b>Enuig</b>	<b>69,23%</b>	23,08%	0,00%	7,69%
<b>Neutre</b>	20,00%	<b>70,00%</b>	10,00%	0,00%
<b>Alegria</b>	0,00%	0,00%	<b>100,00%</b>	0,00%
<b>Sorpresa</b>	9,09%	18,18%	9,09%	<b>63,64%</b>

**Probabilitat d'error 25,00%**

*Figura 6-3*

	<b>Enuig</b>	<b>Neutre</b>	<b>Alegria</b>	<b>Sorpresa</b>
<b>Enuig</b>	<b>75,00%</b>	25,00%	0,00%	0,00%
<b>Neutre</b>	22,22%	<b>77,78%</b>	0,00%	0,00%
<b>Alegria</b>	0,00%	0,00%	<b>100,00%</b>	0,00%
<b>Sorpresa</b>	0,00%	11,11%	11,11%	<b>77,78%</b>

**Probabilitat d'error 17,50%**

*Figura 6-4*

Podríem dir que aquests són els resultats propis del sistema de reconeixement d'emocions implementat.

Analitzant de nou aquests resultats podem enunciar que les confusions es mantenen encara que millorem la taxa d'error per a totes les emocions i per la total, passant d'un 25 % a un 17,50 %.

## 7 Conclusió

Les taxes d'error obtingudes per al sistema de reconeixement d'emocions han estat les següents:

Taxes d'error		
	Amb síntesis errònies	Sense síntesis errònies
CMU	25,56%	13,09%
UPC	25,00%	17,50%

Així, si les síntesis dels models s'executen correctament podem afirmar que tindrem unes taxes d'error del 17%. Però, com no podem assegurar-ho a priori, les taxes d'error més probables són del 25%.

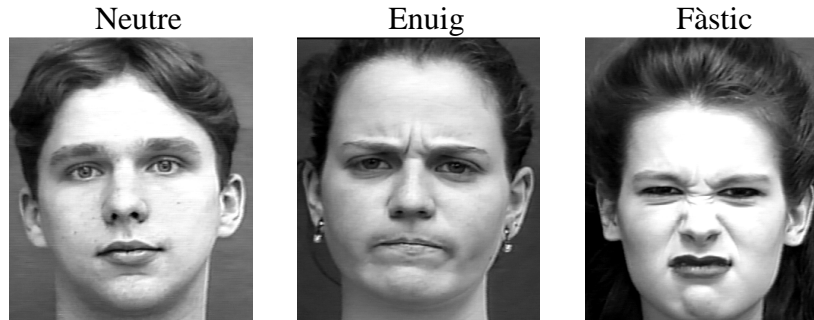
Volem fer notar, que un requisit per a que el sistema funcioni correctament és que no ha de ser possible sintetitzar una emoció utilitzant un model AAM que correspongui a una altra emoció. Així, podem dir que el sistema té com a requisit l'ortogonalitat dels models AAM amb els que es sintetitza la imatge d'entrada. És a dir, l'espai d'emoció que conté un model AAM ha de correspondre inequívocament a l'emoció que representa. D'aquesta manera, els models AAM han de ser entrenats amb un conjunt d'imatges que corresponguin a l'emoció i que les formes d'aquestes no siguin molt diverses.

El sistema de reconeixement genera decisions que confonen unes emocions amb altres de forma quantitativament rellevant. Analitzant aquest fet podem dir que és degut a diverses raons que enunciarem com:

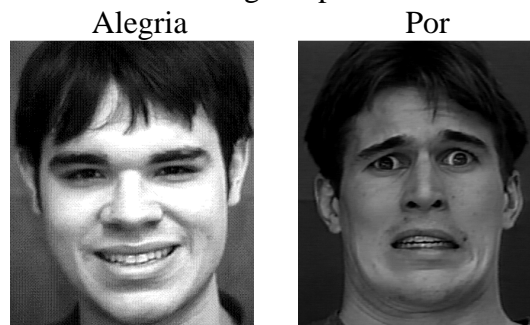
- **Emocions visualment similars**

Podem trobar aquesta causa entre les següents emocions:

Primer entre l'emoció de neutre, enuig i fàstic.

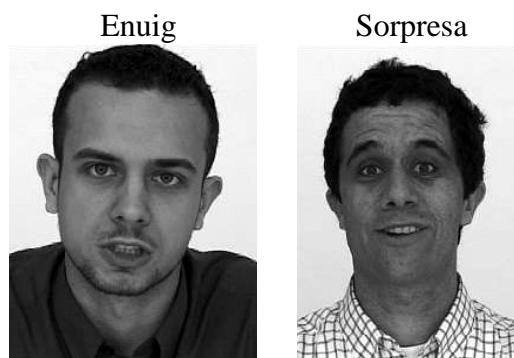


I després entre les emocions d'alegria i por.



- **Imatges amb un patró d'emoció poc definit**

Uns exemples d'aquesta causa són:



- **Error deguts al mètode de quantificació de l'error. Això és, l'error en la textura.**

Es dona el cas de que la síntesi del model sigui correcta, és a dir, la forma dels contorns és correcta per a l'emoció que representa, però la textura de la imatge no concorda a la perfecció amb la imatge d'origen. Així, el sistema detecta un error superior.

## 8 Referències

- [1] V. Bruce, "What the Human Face Tells the Human Mind: Some Challenges for the Robot-Human Interface." *Proc. Int'l Workshop Robot and Human Comm.*, pp. 44-51, 1992.
- [2] A. Takeuchi and K. Nagao, "Communicative Facial Displays as a New Conversational Modality." *Proc. ACM INTERCHI*, pp. 187-193, 1993.
- [3] F. Hara and H. Kobayashi, "State of the Art in Component Development for Interactive Communication with Humans." *Advanced Robotics*, vol. 11, no. 6, pp. 585-604, 1997.
- [4] C. Darwin, "The Expression of the Emotions in Man and Animals." *J. Murray*, London, 1872.
- [5] P. Ekman, W.V. Friesen, "Constants across cultures in the face and emotion" *J. Personality Social Psychol.* 17 (2) (1971) 124-129.
- [6] M. Suwa, N. Sugie, K. Fujimora, "A preliminary note on pattern recognition of human emotional expression" *Proceedings of the Fourth International Joint Conference on Pattern Recognition*, Kyoto, Japan, 1978, pp. 408-410.
- [7] K. Mase, A. Pentland, "Recognition of facial expression from optical flow", *IEICE Trans. E* 74 (10) (1991) 3474-3483.
- [8] L.S. Chen, T.S. Huang, T. Miyasato, and R. Nakatsu, "Multimodal Human Emotion/Expression Recognition" *Proc. Int'l Conf. Automatic Face and Gesture Recognition*, pp. 366-371, 1998.
- [9] L.C. De Silva, T. Miyasato, and R. Nakatsu, "Facial Emotion Recognition Using Multimodal Information" *Proc. Information, Comm., and Signal Processing Conf.*, pp. 397-401, 1997.
- [10] R. Nakatsu, "Toward the Creation of a New Medium for the Multimedia Era" *Proc. IEEE*, vol. 86, no. 5, pp. 825-836, 1998.
- [11] T. Johanstone, R. Banse, and K.S. Scherer, "Acoustic Profiles in Prototypical Vocal Expressions of Emotions" *Proc. Int'l Conf. Phonetic Science*, vol. 4, pp. 2-5, 1995.
- [12] V.A. Petrushin, "Emotion in Speech: Recognition and Application to Call Centers" *Proc. Conf. Artificial Neural Networks in Eng.*, 1999.
- [13] T.S. Polzin and A.H. Waibel, "Detecting Emotions in Speech" *Proc. Conf. Cooperative Multimedia Comm.*, 1998.
- [14] N. Tosa and R. Nakatsu, "Life-Like Communication Agent-Emotion Sensing Character MIC and Feeling Session Character MUSE" *Proc. Conf. Multimedia Computing and Systems*, pp. 12-19, 1996.
- [15] R.W. Picard and E. Vyzas, "Offline and Online Recognition of Emotion Expression from Physiological Data" *Emotion-Based Agent Architectures Workshop Notes, Int'l Conf. Autonomous Agents*, pp. 135-142, 1999.
- [16] V.I. Pavlovic, R. Sharma, and T.S. Huang, "Visual Interpretation of Hand Gestures for Human-Computer Interaction: Review" *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 677-695, 1997.
- [17] R.J. Holt, T.S. Huang, A.N. Netravali, and R.J. Qian, "Determining Articulated Motion from Perspective Views" *Pattern Recognition*, vol. 30, no. 9, pp. 1435-1449, 1997.
- [18] A.D. Wilson and A.F. Bobick, "Parametric Hidden Markov Models for Gesture Recognition" *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 21, no. 9, pp. 884-900, Sept. 1999.

- [19] H.K. Lee and J.H. Kim, "An HMM-Based Threshold Model Approach for Gesture Recognition" *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 21, no. 10, pp. 961-973, Oct. 1999.
- [20] G. Schwartz, P. Fair, P. Salt, M. Mandel, G. Klerman, "Facial expression and imagery in depression: an electromyographic study" *Psychosomatic Med.* 38 (337-347) (1976).
- [21] P. Ekman, "Emotions in the Human Face", *Cambridge University Press*, Cambridge, 1982.
- [22] A. Lanitis, C. Taylor, T. Cootes, "Automatic interpretation and coding of face images using flexible models", *IEEE Trans. Pattern Anal. Mach. Intell.* 19 (7) (1997) 743-756.
- [23] H. Hong, H. Neven, C. Von der Malsburg, "Online facial expression recognition based on personalized galleries, Proceedings of the Second International Conference on Automatic Face and Gesture Recognition, (FG'98)", *IEEE, Nara, Japan*, 1998, pp. 354-359.
- [24] J. Steffens, E. Elagin, H. Neven, "PersonSpotter-fast and robust system for human detection, tracking and recognition", *Proceedings of the Second International Conference on Face and Gesture Recognition, (FG'98), Nara, Japan*, 1998, pp. 516-521.
- [25] I. Essa, A. Pentland, "Coding, analysis, interpretation and recognition of facial expressions", *IEEE Trans. Pattern Anal. Mach. Intell.* 19 (7) (1997) 757-763.
- [26] A. Lanitis, C. Taylor, T. Cootes, "Automatic interpretation and coding of face images using flexible models", *IEEE Trans. Pattern Anal. Mach. Intell.* 19 (7) (1997) 743-756.
- [27] M. Black, Y. Yacoob, "Recognizing facial expressions in image sequences using local parameterized models of image motion", *Internat. J. Comput. Vision* 25 (1) (1997) 23-48.
- [28] D. Pollen, S. Ronner, "Phase relationship between adjacent simple cells in the visual cortex", *Science* 212 (1981) 1409-1411.
- [29] J. Daugman, "Complete discrete 2D Gabor transform by neural networks for image analysis and compression", *IEEE Trans. Acoustics, Speech Signal Process.* 36 (1988) 1169-1179.
- [30] W. Fellenz, J. Taylor, N. Tsapatsoulis, S. Kollias, "Comparing template-based, feature-based and supervised classification of facial expressions from static images", *Proceedings of Circuits, Systems, Communications and Computers (CSCC'99), Nigata, Japan*, 1999, pp. 5331-5336.
- [31] M. Dailey, G. Cottrell, "PCA Gabor for expression recognition", *Institution UCSD*, Number CS-629, 1999.
- [32] M. Bartlett, "Face image analysis by unsupervised learning and redundancy reduction," *Ph.D. Thesis*, University of California, San Diego, 1998.
- [33] Z. Zhang, M. Lyons, M. Schuster, S. Akamatsu, "Comparison between geometry-based and Gabor-wavelets-based facial expression recognition using multi-layer perceptron", *IEEE Proceedings of the Second International Conference on Automatic Face and Gesture Recognition (FG'98), Nara, Japan*, 1998, pp. 454-459.
- [34] M. Lyons, J. Budynek, S. Akamatsu, "Automatic classification of single facial images", *IEEE Trans. Pattern Anal. Mach. Intell.* 21 (12) (1999).
- [35] T. Cootes, G. Edwards, C. Taylor, "Active appearance models", *IEEE PAMI* 23 (6) (2001) 681-685.

- [36] G. Edwards, T. Cootes, C. Taylor, "Face recognition using active appearance models", *Proceedings of the Fifth European Conference on Computer Vision (ECCV)*, Vol. 2, University of Freiburg, Germany, 1998, pp. 581–695.
- [37] H. Kobayashi, F. Hara, "Facial interaction between animated 3D face robot and human beings", *Proceedings of the International Conference on Systems, Man and Cybernetics*, Orlando, FL, USA, 1997, pp. 3732–3737.
- [38] M. Pantic, L. Rothkrantz, "Expert system for automatic analysis of facial expression", *Image Vision Comput.* J. 18 (11) (2000) 881–905.
- [39] T. Otsuka, J. Ohya, "Spotting segments displaying facial expression from image sequences using HMM", *IEEE Proceedings of the Second International Conference on Automatic Face and Gesture Recognition (FG'98)*, Nara, Japan, 1998, pp. 442–447.
- [40] J. Cohn, A. Zlochower, J. Lien, Y. Wu, T. Kanade, "Automated face coding: a computer-vision based method of facial expression analysis", *Seventh European Conference on Facial Expression Measurement and Meaning*, Salzburg, Austria, 1997, pp. 329–333.
- [41] N. Oliver, A. Pentland, F. Berard, "LAFTER: a real-time lips and face tracker with facial expression recognition", *Proceedings of the IEEE Conference on Computer Vision (CVPR97)*, S. Juan, Puerto Rico, 1997.
- [42] M. Rosenblum, Y. Yacoob, L. Davis, "Human expression recognition from motion using a radial basis function network architecture", *IEEE Trans. Neural Networks* 7 (5) (1996) 1121–1138.
- [43] H. Kobayashi, F. Hara, "Dynamic recognition of basic facial expressions by discrete-time recurrent neural network", *Proceedings of the International Joint Conference on Neural Networks*, 1993, pp. 155–158.
- [44] C. Lisetti, D. Rumelhart, "Facial expression recognition using a neural network", *Proceedings of the 11th International Flairs Conference*, AAAI Press, New York, 1998.
- [45] M. Yoneyama, Y. Iwano, A. Ohtake, K. Shirai, "Facial expression recognition using discrete Hopfield neural networks", *Proceedings of the International Conference on Image Processing (ICIP)*, Santa Barbara, CA, USA, Vol. 3, 1997, pp. 117–120.
- [46] J. Zhao, G. Kearney, "Classifying facial emotions by backpropagation neural networks with fuzzy inputs", *Proceedings of the International Conference on Neural Information Processing*, Vol. 1, 1996, pp. 454–457.
- [47] D. Matsumoto, "Cultural similarities and differences in display rules", *Motivation Emotion* 14 (3) (1990) 195–214.
- [48] K. Karpouzis, G. Votsis, G. Moschovitis, S. Kollias, "Emotion recognition using feature extraction and 3-D models", *Proceedings of IMACS International Multiconference on Circuits and Systems Communications and Computers (CSCC'99)*, Athens, Greece, 1999, pp. 5371–5376.
- [49] Y. Tian, T. Kanade, J. Cohn, "Recognizing action units for facial expression analysis", *IEEE Trans. Pattern Anal. Mach. Intell.* 23 (2) (2001) 97–115.
- [50] T. Otsuka, J. Ohya, "Extracting facial motion parameters by tracking feature points", *Proceedings of First International Conference on Advanced Multimedia Content Processing*, Osaka, Japan, 1998, pp. 442–453.
- [51] G. Donato, S. Bartlett C. Hager, P. Ekman, J. Sejnowski, "Classifying facial actions", *IEEE Trans. Pattern Anal. Mach. Intell.* 21 (10) (1999) 974–989.

- [52] B. Fasel, J. Luetin, "Recognition of asymmetric facial action unit activities and intensities", *Proceedings of the International Conference on Pattern Recognition (ICPR 2000)*, Barcelona, Spain, 2000.
- [53] J. Lien, T. Kanade, J. Cohn, C. Li, "Automated facial expression recognition based on FACS action units", *IEEE Proceedings of the Second International Conference on Automatic Face and Gesture Recognition (FG'98)*, Nara, Japan, 1998.
- [54] J.N. Bassili, "Facial Motion in the Perception of Faces and of Emotional Expression", *J. Experimental Psychology* 4, pp. 373-379, 1978.
- [55] V. Bruce, *Recognizing Faces*. Hove, East Sussex: Lawrence Erlbaum Assoc., 1986.
- [56] T.F. Cootes, C.J. Taylor, "Statistical Models of Appearance for Computer Vision", Manchester, 2004
- [57] G.J. Edwards, C.J. Taylor, T.F. Cootes, "Interpreting Face Images using Active Appearance Models", *Int. Conf. on Face and Gesture Recognition 1998*. pp. 300-305
- [58] Mikkel B. Stegmann, D. Delgado Gomez, "A brief Introduction to Statistical Shape analysis", Denmark, 2002.
- [59] Angel Gutiérrez, Detección de estados anímicos utilizando imágenes fijas, PFC ETSETB, Noviembre 2006.
- [60] Kanade, T., Cohn, J.F., Tian. Y., Comprehensive Database for Facial Expression Analysis, Proceedings of the Fourth IEEE International Conference on Automatic Face and Gestures Recognition, Grenoble, France, 2000



## **Annex Configuració AAM-API**

### **Arxiu d'especificació de forma AAM (ASF)**

L'arxiu ASF s'estructura com un conjunt de línies separades pels caràcters CR. A qualsevol lloc de l'arxiu, es poden incloure comentaris sempre que comencin per caràcter '#'. El layout d'un arxiu ASF es com s'especifica a continuació:

- La Línia 1 conté el nombre total de punts,  $n$ , a la forma.
- La Línia 2 fins la  $n+1$  contenen la informació dels punts de la forma (una línia per punt), que és la posició, tipus, connectivitat etc. explicat més endavant.
- La Línia  $n+2$  conté la imatge host. És a dir, el nom de l'arxiu de la imatge on s'ha definit l'anotació.

El format de la definició d'un punt és el següent:

point := <path#> <type> <x-pos> <y-pos> <point#> <connects from> <connects to>

<path#> La línia de la forma a la que pertany el punt. Els punts de línies diferents de la forma no poden estar intercanviats dins de l'ordre de les línies de l'arxiu ASF.

<type> Un camp mapejat per bits que defineix el tipus de punt:

- Bit 1: Punt de bord exterior / punt interior.
- Bit 2: Punt d'anotació original / Punt d'anotació artificial.
- Bit 3: Punt de línia de forma tancada / Punt de línia de forma oberta.
- Bit 4: No forat (Non-hole) / punt de forat (Hole point).

La resta de bits han d'estar a zero.

<x-pos> La posició relativa en x del punt. S'obté dividint el la posició del píxel dintre del rang de coordenades de la imatge [0; amplada imatge -1] per l'amplada de la imatge. Per exemple, el píxel  $x=47$  (el 48è píxel) en una imatge d'amplada 256, té la posició relativa  $47/256 = 0.18359375$ .

<y-pos> La posició relativa en y del punt. S'obté dividint el la posició del píxel dintre del rang de coordenades de la imatge [0; alçada imatge -1] per l'alçada de la imatge. Per exemple, el píxel  $y=47$  (el 48è píxel) en una imatge d'alçada 256, té la posició relativa  $47/256 = 0.18359375$ .

<point#> El número de punt. El primer punt es zero.

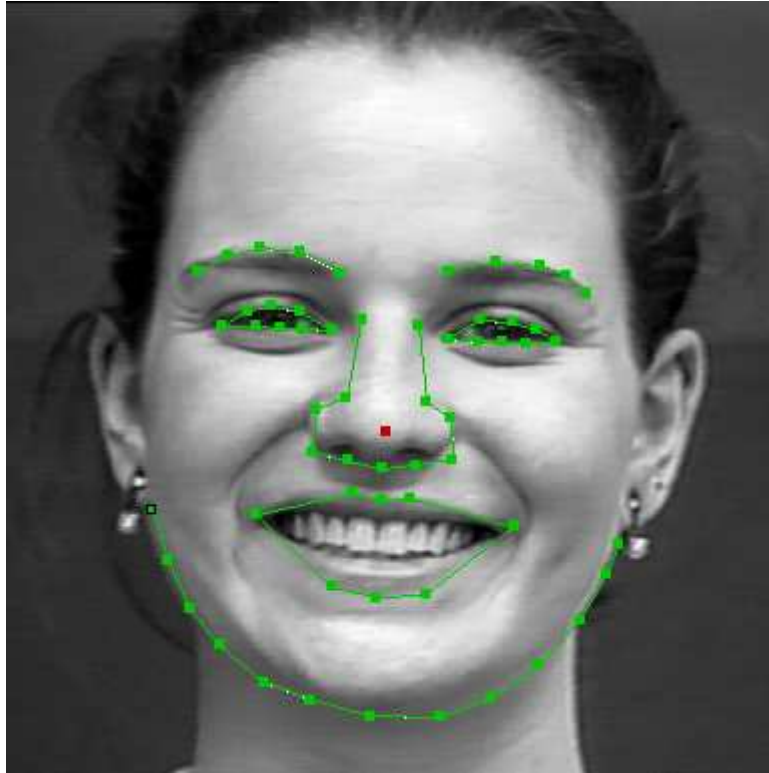
<connects from> El punt anterior a l'actual dintre de la línia de la forma a la que pertany. Si no n'hi ha posarem el valor de <point#> actual.

<connects to> El punt següent a l'actual dintre de la línia de la forma a la que pertany. Si no n'hi ha posarem el valor de <point#> actual.

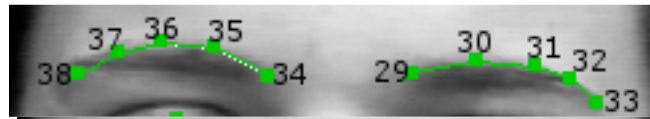
A més, apliquem les següents regles en el format:

- Els camps en una especificació d'un punt estan separats per espais o per tabuladors.
- Les línies de la forma suposen orientació en el sentit de les agulles del rellotge.
- Els punts es defineixen el quart quadrant.

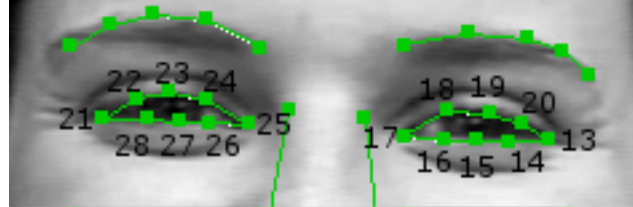
**Arxiu ASF dels punts de control d'un rostre**



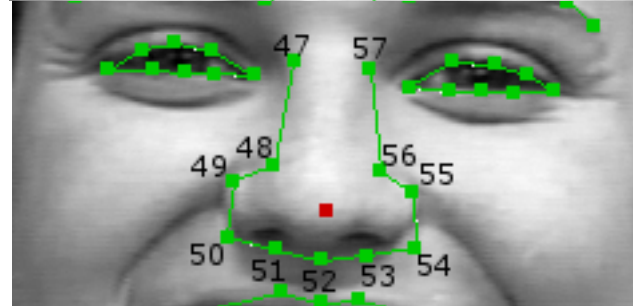
Path #4 cella esquerra  
Path #3 cella dreta



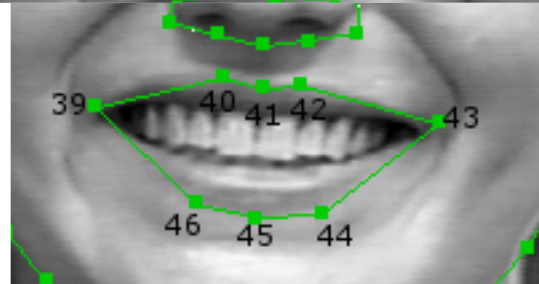
Path #1 ull dret  
Path #2 ull esquerra



Path #6 nas



Path #5 boca



Path #0 contorn

