



ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA Y SISTEMAS DE TELECOMUNICACIÓN

PROYECTO FIN DE GRADO

TÍTULO: Integración de algoritmos basados en Unidades de Acción en una herramienta de análisis de reconocimiento de emociones

AUTOR: Almudena Gil Gaitán

TITULACIÓN: Sonido e imagen

TUTOR (o Director en su caso): Martina Eckert

DEPARTAMENTO: Teoría de la Señal y Comunicaciones

VºBº

Miembros del Tribunal Calificador:

PRESIDENTE: Marisa Martín Ruiz

VOCAL: Martina Eckert

SECRETARIO: Enrique Rendón Angulo

Fecha de lectura: 1 de Julio de 2016

Calificación:

El Secretario,

*“El genio comienza las grandes
obras pero solo el trabajo las
acaba”.*

Joseph Joubert

Agradecimientos

Con este proyecto termina una etapa de esfuerzo, trabajo y sacrificio, llena de momentos buenos y malos, y que pese a esos malos momentos puedo decir que ha merecido la pena vivirla.

Quiero mostrar mi agradecimiento a todos aquellos que han compartido conmigo esta etapa, a los que me han ayudado a superar las dificultades que he ido encontrando en estos años, a los que me han dado su apoyo y han tenido tanta paciencia.

En especial quiero dar las gracias a mi familia y amigos por comprenderme, darme su cariño, su confianza y apoyo incondicional a lo largo de mi vida. Por demostrarme que todo esfuerzo tiene su recompensa y por hacerme ver y creer que se puede llegar tan lejos como uno quiera. Por todo ello, por hacer que esta vida sea mucho más fácil para mí y sobre todo porque no sería lo que soy sin vosotros, GRACIAS.

También quiero dar las gracias a mis compañeros y amigos Sara González y Diego Zapatero por su colaboración y ayuda en la realización de este proyecto. Al resto de compañeros que he tenido el placer de conocer durante mi carrera universitaria, por los buenos momentos que hemos vivido y que han hecho que esta etapa sea inolvidable.

Por último y no menos importante quiero darle las gracias a mi tutora Martina Eckert por la oportunidad de poder colaborar con ella en el centro de investigación CITSEM y por su ayuda y apoyo al realizar este Proyecto de Fin de Grado.

Gracias a todos por tanto.

Almudena Gil

Julio 2016

Resumen

La comunicación no verbal entre los seres humanos cada día cobra más importancia, y las expresiones faciales son el mejor ejemplo de ello. En los últimos años, gracias al avance en el procesado digital de la imagen se han desarrollado numerosos sistemas capaces de reconocer estas emociones de una manera automática.

En este proyecto se parte de un trabajo anterior a partir del cual se realizan mejoras y avances, centrándose en el análisis de los movimientos de los músculos faciales para poder reconocer emociones. A raíz de ello, se realiza como primer paso un estudio sobre qué acciones de músculos o unidades de acción son las más características de cada emoción. Como segundo paso se implementan las reglas necesarias para poder reconocer cada unidad de acción y con la ayuda e implementación de un clasificador de la emoción llegar al reconocimiento de la misma. Todo ello a través de una herramienta modular y ampliable para mayor comodidad.

Por último, se realizan diferentes pruebas con el fin de obtener el mayor número de conclusiones posibles para poder determinar qué mejoras serían necesarias y cuáles podrían ser las posibles líneas de trabajo a partir de ellas.

Abstract

Non-verbal communication between humans becomes more important every day, and facial expressions are the best example of it. In recent years, thanks to advances in digital image processing, a lot of systems able to recognize these emotions automatically have been developed.

This project is based on previous work in this field, from which I have developed and improved, focusing on analysis of facial muscle movement in order to recognize emotions. The first step is a study of the muscle movements or action units that are mostly appearing in the six basic emotions. The second step is the formulation of unique rules that can be used to recognize the emotions depending on the detected action units, with the help of an emotion. For this purpose, modular and extendable too has been developed.

Finally, different tests are performed in order to obtain important conclusions to determine which improvements would be needed and which lines of work could be possible.

Índice

Lista de acrónimos	I
Índice de figuras	II
Índice de tablas	III
1. Introducción	1
2. Reconocimiento facial de emociones	2
2.1. Técnicas de reconocimiento de emociones	2
2.2. Las Unidades de Acción como técnica elegida	8
2.3. Unidades de Acción	11
2.3.1. Estudio de las publicaciones	11
2.3.2. Análisis de las publicaciones para la extracción de reglas comunes	20
2.3.3. Estudio subjetivo de las AUs	29
2.3.4. Análisis final	35
3. Reconocimiento de las AUs	38
3.1. Colocación de puntos	38
3.2. Detección de AUs y CAUs con 19 puntos	46
3.3. Árbol de Decisión	52
4. Herramienta de pruebas	54
4.1. Antecedentes	54
4.2. Ampliación de la herramienta	57
5. Pruebas y resultados	63
5.1. Base de datos CohnKanade+	63
5.1.1. Método Thres	65
5.1.2. Método Canny	68
5.2. Base de datos JAFFE	69
5.2.1. Método Thres	70
5.2.2. Método Canny	71
5.3. Discusión de resultados	71
5.3.1. Pruebas de diferenciación de emociones efusivas y no efusivas	72
6. Conclusiones	74
7. Trabajo futuro	75
8. Referencias	76

Lista de acrónimos

- FACS: Facial Action Coding System (Sistema de Codificación de Acción Facial)
- AU: Action Unit (Unidad de Acción)
- CAU: Combined Action Unit (Unidad de Acción Combinada)
- PCA: Principal Component Analysis (Análisis de componentes principales)
- AAM: Active Appearance Model (Modelo de apariencia activa)
- ASM: Active Shape Model (Modelo de forma activa)
- LBP: Local Binary Patterns (Patrones binarios locales)
- NN: Neural Network (Red neuronal)
- ANN: Artificial Neural Network (Red neuronal artificial)
- ICA: Independent Components Analysis (Análisis de componentes independientes)
- LDA: Linear Discriminant Analysis (Análisis discriminante lineal)
- DAM: Direct Appearance Model (Modelo de apariencia directa)
- DBN: Dynamic Bayesian Network (Red bayesiana dinámica)
- SVM: Support Vector Machine (Máquina de soporte vectorial)
- DCT: Discrete Cosine Transform (Transformada de coseno discreta)
- CK: Cohn -Kanade (Cohn-Kanade)
- CK+: Extended Cohn-Kanade (Cohn-Kanade extendida)
- JAFFE: Japanese Female Facial Expression (Expresión facial de mujeres japonesas)
- MMI: Michel F. Valstar, Maja Panatic Imperial College London (Michel F. Valstar, Maja Panatic Colegio Imperial de Londres)
- MUG: Multimedia Understanding Group (Grupo de comprensión multimedia)
- CITSEM: Centro de Investigación en Tecnologías Software y Sistemas Multimedia para la Sostenibilidad

Índice de figuras

Figura 1 Diagrama de bloques con las etapas para el reconocimiento de emociones	2
Figura 2 Filtro Gabor en el dominio del espacio (izquierda) y de la frecuencia (derecha) [6].....	4
Figura 3 Wavelet de Haar [7].....	4
Figura 4 Imágenes proyectadas en el espacio PCA para obtener eigenfaces	5
Figura 5 Ejemplo del método LBP [9]	5
Figura 6 Red Neuronal [10].....	7
Figura 7 Árbol de decisión	7
Figura 8 Escalas de los niveles de intensidad.....	9
Figura 9 Detección de los 19 puntos característicos de la cara en [3]	11
Figura 10 Diagrama de bloques usando los métodos AAM y SVM [12]	12
Figura 11 Etapas del reconocimiento facial [13].....	13
Figura 12 Diagrama de bloques del sistema de reconocimiento de acción facial [17]	15
Figura 13 Flujo de datos para el proceso de reconocimiento usando en [27]	18
Figura 14 Cluster para el set o conjunto de AUs de la Tabla 12	28
Figura 15 Imágenes con la emoción de happy de [12], [22], [31] y [32]	29
Figura 16 Cluster para el Set o conjunto subjetivo de AUs de la Tabla 21	34
Figura 17 Cluster para el set o Conjunto de AUs decisivo de la Tabla 26.....	37
Figura 18 Proceso de extracción de características para el primer método [3]	39
Figura 19 Proceso de extracción de características para el segundo método	40
Figura 20 Proceso de extracción de características para el tercer método [33]	41
Figura 21 Detección de las regiones para los métodos de Canny, Thres y Landmarks	42
Figura 22 Colocación de puntos para los métodos de Canny, Thres y Landmarks.....	42
Figura 23 Gráfica comparativa de la detección de puntos imágenes en color	45
Figura 24 Gráfica comparativa de la detección de puntos imágenes en grises	46
Figura 25 Triángulos utilizados para detección de las AUs y CAUs.	47
Figura 26 Estructura de un Árbol de Decisión	52
Figura 27 Ejemplo de estructura simple de Árbol de Decisión	54
Figura 28 Interfaz gráfica de la herramienta del PFG Sara González	55
Figura 29 Interfaz gráfica de la herramienta del PFG Sara González, tras su ejecución.....	57
Figura 30 Interfaz gráfica de la herramienta ampliada tras ser ejecutada	60
Figura 31 Interfaz gráfica de la herramienta ampliada tras el reconocimiento de la emoción..	61
Figura 32 Interfaz gráfica de la herramienta ampliada tras pulsar distintos botones.....	62
Figura 33 Árbol de Decisión con las combinaciones de la Tabla 27 para la CK+	65
Figura 34 Árbol de Decisión con las combinaciones de la Tabla 36 para la CK+	67
Figura 35 Imágenes de la CK+ para el método de Canny.....	69
Figura 36 Imágenes de la base de datos JAFFE con la emoción de fear.....	70
Figura 37 Imágenes de la base de datos JAFFE con la emoción de anger y sadness.....	71
Figura 38 Árbol de Decisión para emociones efusivas y no efusivas	73

Índice de tablas

Tabla 1 Unidades de Acción (AUs) más importantes según el FACS.....	9
Tabla 2 Niveles de intensidad según FACS	10
Tabla 3 Correspondencia de AUs para cada emoción según el FACS	10
Tabla 4 AUs únicas y ausentes para cada emoción según [23].....	16
Tabla 5 AUs presentes en cada emoción, establecidas por las publicaciones estudiadas	21
Tabla 6 Combinaciones de AUs para cada emoción, por las publicaciones estudiadas.....	22
Tabla 7 Estadística de recomendaciones de las AUs sobre 10 publicaciones.....	23
Tabla 8 Medias aritméticas de la frecuencia de aparición de AUs para cada emoción	24
Tabla 9 Conjunto o set de AUs características según las publicaciones estudiadas	25
Tabla 10 AUs exclusivas.....	26
Tabla 11 AUs de exclusión o no presentes	26
Tabla 12 Set o conjunto de AUs que se pueden detectar con los 19 puntos de la cara	27
Tabla 13 Combinaciones de AUs para el set o conjunto de AUs de la Tabla 9	28
Tabla 14 Combinaciones de AUs para el set de AUs de la Tabla 9 teniendo en cuenta las AUs que se pueden detectar con los 19 puntos de la cara	29
Tabla 15 Set o conjunto subjetivo de AUs encontradas en el estudio imágenes	30
Tabla 16 Estadística de la presencia de AUs tras un estudio visual subjetivo	31
Tabla 17 Medias aritméticas de la frecuencia de aparición de AUs para el estudio subjetivo	31
Tabla 18 Set o conjunto subjetivo de AUs características mayormente utilizadas.....	32
Tabla 19 AUs exclusivas, análisis subjetivo.....	32
Tabla 20 AUs de exclusión o no presentes del análisis subjetivo	33
Tabla 21 Set o conjunto de AUs del análisis subjetivo para los 19 puntos de la cara	33
Tabla 22 Posibles combinaciones de AUs para el set de AUs subjetivo de la Tabla 18	34
Tabla 23 Combinaciones de AUs para el set de AUs subjetivo de la Tabla 18 para las AUs que se pueden detectar con los 19 puntos de la cara	34
Tabla 24 Comparación de AUs de publicaciones y subjetivo, para los 19 puntos	35
Tabla 25 Selección de AUs y definición de CAUs	36
Tabla 26 Set o Conjunto de AUs decisivo	37
Tabla 27 Combinaciones de AUs para el set o conjunto decisivo	38
Tabla 28 Porcentajes de acierto en la detección de puntos para imágenes en color	43
Tabla 29 Porcentajes de acierto en la detección de puntos para imágenes en grises [3]	44
Tabla 30 Triángulos y puntos implicados en la detección de cada AU y CAU	48
Tabla 31 Dibujos para cada AU y CAU	49
Tabla 32 Ejemplo sencillo de matriz de combinaciones	53
Tabla 33 Número de imágenes de la CohnKanade+	63
Tabla 34 Porcentajes de AUs y CAUs detectadas en cada emoción	64
Tabla 35 Matriz de confusión para las combinaciones de Tabla 27	65
Tabla 36 Combinaciones de AUs para el set o conjunto decisivo, tras realizar ajustes	67
Tabla 37 Matriz de confusión CK+ para las combinaciones de Tabla 36, Thres.....	68
Tabla 38 Matriz de confusión CK+ para las combinaciones de Tabla 36, Canny.....	68
Tabla 39 Matriz de confusión JAFFE para las combinaciones de Tabla 36, Thres.....	70
Tabla 40 Matriz de confusión JAFFE para las combinaciones de Tabla 36, Canny	71
Tabla 41 Matriz de confusión CK+ para emociones efusivas y no efusivas, Thres.....	73

1. Introducción

En los últimos tiempos se ha dado cada vez más importancia al reconocimiento facial de emociones, ya que las expresiones faciales son la forma que tenemos las personas para comunicarnos y para expresar nuestro estado de ánimo en cada momento. Gran parte de este impulso se debe al aumento y desarrollo de posibilidades del procesado de la imagen, permitiendo la detección de emociones de forma automática.

Actualmente existen muchas áreas en las que se puede aplicar el reconocimiento facial de emociones, como son: mejoras para cámaras fotográficas, médicas, robóticas, videojuegos, y en el sector del marketing para saber la emoción que causa un producto sobre un cliente.

Con motivo del gran número de métodos y sistemas existentes, es muy complicado elaborar un método nuevo que no ralentice al sistema de reconocimiento y que introduzca mejoras. Pero todavía es más difícil poder comparar métodos y resultados, ver cuáles son mejores, realizar pruebas con distintas bases de datos o poder ver cómo influyen las modificaciones introducidas de forma directa.

Este proyecto se centra en el estudio de los movimientos o acciones musculares (AUs) que permiten llegar a reconocer una emoción con el uso de puntos característicos de la cara que se reconocen previamente y que pueden definir estas acciones. Para ello se parte del trabajo previo realizado en el ámbito de prácticas en el centro de investigación CITSEM durante el 2014/2015 [1], sobre el reconocimiento de emociones mediante las unidades de acción.

Tras este estudio se determinan que movimientos de músculos o unidades de acción (AUs) son características para cada emoción, para poder así posteriormente establecer unas reglas de detección que permitan saber cuándo una unidad de acción está presente en una imagen y cuando no. Siendo uno de los objetivos principales de este proyecto.

Con la implementación y aplicación de estas reglas y con la ayuda de un método de clasificación de emociones se puede llegar a reconocer seis emociones básicas: *fear* (miedo), *anger* (ira), *happiness* (felicidad), *surprise* (sorpresa), *sadness* (tristeza) y *disgust* (aversión). Estas emociones son las más utilizadas en los sistemas de reconocimiento, es por ello por lo que se decide centrarse en ellas y no en otras.

Para este proyecto se utiliza como método de clasificación el Árbol de Decisión a través del cual se crea un clasificador que muestran posibles resultados. Este método se toma del trabajo realizado por Diego Zapatero tanto en el ámbito de prácticas del CITSEM [1] como en su PFG [2] y se implementa en el actual proyecto.

Con el objetivo de poder determinar e implementar las unidades de acción propias de cada emoción y el Árbol de Decisión se utiliza la herramienta que se propone en el PFG de Sara González [3], a partir de la cual se ofrece una forma intuitiva y accesible para poder analizar y comparar distintos métodos de extracción de características y de clasificación de la emoción.

Por otro lado se realiza una ampliación de esta versión inicial de la herramienta a través de la cual se pueden ver los resultados de acierto y fallo en el proceso de reconocimiento, así como la posibilidad de elegir diferentes métodos de extracción de características.

A continuación se detalla la organización del proyecto. El capítulo dos presenta el estado del arte de las técnicas de reconocimiento de emociones en el ámbito de la detección de la cara, extracción de características y clasificación de la emoción. Además, contiene una introducción y un estudio detallado de las Unidades de Acción, tanto de publicaciones existentes como de estudios subjetivos, para poder llegar a determinar que AUs son las más importantes para cada emoción. En el capítulo tres se analiza la precisión de la colocación de puntos de la cara para los métodos de extracción de características que se utilizan, y se describe el proceso de detección de las AUs para los 19 puntos de la cara que se localizan. También se explica

brevemente el método de clasificación utilizado para llevar a cabo el reconocimiento de emociones con el uso de las AUs. A continuación, en el capítulo cuatro se describe la herramienta de la que se parte y la ampliación de la misma realizada en este proyecto, y en el capítulo cinco se muestran los resultados obtenidos utilizando distintas bases de datos y combinaciones de métodos. Por último en el capítulo seis se extraen conclusiones y en el siete se proponen futuras líneas de investigación.

2. Reconocimiento facial de emociones

2.1. Técnicas de reconocimiento de emociones

Los sistemas de reconocimiento facial de emociones se componen de una serie de etapas las cuales se desarrollan a través de diferentes métodos y algoritmos, como se puede observar en la Figura 1. Estas etapas se explican a continuación:

- Obtención de la imagen a analizar (cara), ya sea por medio de una imagen estática de una base de datos o por medio de una webcam.
- Extracción de características, es decir, se selecciona el método de extracción de características adecuado para llevar a cabo el reconocimiento de la emoción. Con este método se pueden detectar por ejemplo los puntos característicos de la cara y sus características, como es el caso del método basado en Unidades de Acción o AUs.
- Módulo de Aprendizaje, a partir del cual se ejecuta un algoritmo o método de aprendizaje automático a elegir, necesario para entrenar al sistema para su posterior reconocimiento de la emoción. Como por ejemplo Árbol de Decisión, Red Neuronal, SVM, etc.
- Clasificación de las emociones, es decir, el reconocimiento de la emoción a partir del método elegido en el módulo de aprendizaje.



Figura 1 Diagrama de bloques con las etapas para el reconocimiento de emociones

Para poder entender cada una de las etapas anteriores y conocer los diferentes métodos, algoritmos y técnicas que existen hoy en día se realizan estudios y análisis sobre el estado del arte en el ámbito del reconocimiento facial de emociones.

Inicialmente en los sistemas de reconocimiento cuando se captura la imagen se procede a detectar la cara y localizar cada una de las regiones que la componen, como son los ojos, las cejas y la boca. Uno de los algoritmos más utilizados para detectar la cara y sus regiones es el algoritmo de Viola & Jones [4], incluido en la *Computer Vision Toolbox* de Matlab, que utiliza

el algoritmo de *Adaboost* y la combinación de distintos clasificadores en cascada para descartar los objetos que no se consideran de interés.

El siguiente paso es la implementación del método de extracción de características con el cual se extraen las características o información relevante que posee el rostro para poder identificar y clasificar las distintas emociones. A continuación se destacan los métodos más relevantes encontrados en el ámbito de la **extracción de características**:

- Métodos de extracción de características de deformación: los cuales se basan en la extracción de información de deformación que se produce en la cara como por ejemplo la deformación geométrica o los cambios de textura del rostro. Algunos de estos principales métodos son los siguientes:
 - *AAM (Active Appearance Model)*: es un método creado por Cootes, Edwards y Taylor en 1998 [5], es un modelo estadístico de apariencia creado en una fase de entrenamiento donde se encuentran los parámetros del modelo que permiten reducir las diferencias entre una imagen y el modelo que se proyecta en la imagen. Este entrenamiento se realiza con imágenes que contienen puntos de referencia, en concreto 68 puntos, que se sitúan en posiciones representativas simbolizando la forma del rostro. A partir de estos puntos se crea una malla triangular en 2D, cuyas coordenadas se definen con la forma (1) donde n es el número de vértices.

$$s = [x_1 \ y_1 \ x_2 \ y_2 \ \dots \ x_n \ y_n] \quad (1)$$

Esta malla modelo se realiza haciendo la media de las mallas de cada una de las imágenes de entrenamiento, todas ellas normalizadas, y un subespacio vectorial que contiene las variaciones principales. El AAM se considera un modelo eficaz debido a que se aprovecha de las características de textura y forma que están presentes en los rostros.

- *ASM (Active Shape Model)*: este método consiste en un modelo estadístico que se utiliza para representar los objetos deformables, donde cada forma se representa por un conjunto de puntos característicos. Estos puntos primero se buscan de forma individual y posteriormente se aplica el Análisis de Componentes Principales (PCA) para que la forma del objeto solo se pueda deformar de manera específica. A su vez este modelo se compone de tres elementos: un modelo de contorno, un modelo de apariencia y un algoritmo de búsqueda para ajustar el modelo. El objetivo del modelo de contorno es asegurar que con la segmentación se producen contornos validos para la región facial. El modelo de apariencia se encarga de asegurar que la segmentación localiza el rostro en una posición donde la estructura de la imagen del contorno es similar a la de las imágenes de entrenamiento. Por último el algoritmo de búsqueda hace que se ajuste el contorno de la región facial mediante un proceso iterativo que comienza en el centro de la cara.

Estas características se extraen mediante el filtrado de la imagen o de una región concreta de interés de la misma. Para ello se pueden utilizar diferentes técnicas como las que se muestran a continuación:

- *Gabor Wavelets*: esta técnica utiliza los filtros Gabor, que permiten obtener información de frecuencia en una región de la imagen. Así se transforma la imagen de la cara en pequeñas ondas que facilitan el reconocimiento de la

característica deseada, como se puede ver en la Figura 2. Estos filtros son un banco de filtros cada uno de los cuales tiene distintas orientaciones y frecuencias.

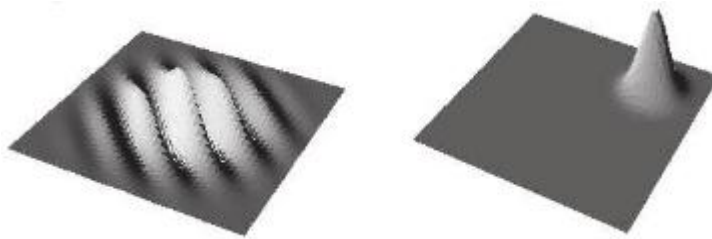


Figura 2 Filtro Gabor en el dominio del espacio (izquierda) y de la frecuencia (derecha) [6].

- DCT (*Discrete Cosine Transform*) es una técnica con la que se comprimen las imágenes mediante la eliminación de la información que no se va a tener en cuenta, para ello transfiere la imagen del dominio del tiempo al dominio de la frecuencia. Su principal finalidad es la de reducir el coste computacional que se produce con la extracción de las características. La DCT-2D se define como se muestra en (2) donde dada una imagen de entrada $N \times M$ en escala de grises, $I(i, j)$ es una imagen de N, M elementos y $X(i, j)$ es su transformada.

$$X(i, j) = \frac{2}{N} k_i k_j * \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} I(x, y) * \cos\left(\frac{\pi(2x+1)i}{2N}\right) * \cos\left(\frac{\pi(2y+1)j}{2M}\right) \quad (2)$$

$$k_i, k_j = \begin{cases} \frac{1}{\sqrt{2}}, & i, j = 0 \\ 1, & otherwise \end{cases}$$

- *Haar Wavelet Transforms*: es otro método de filtrado en frecuencia que se basa en la filtración de una imagen separando en dos grupos las bandas de frecuencia, altas y bajas. La *wavelet Haar* es discontinua y tiene la forma de una función escalón como se ve en la Figura 3.

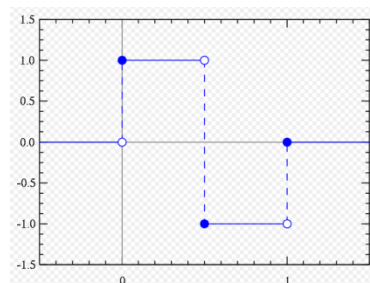


Figura 3 Wavelet de Haar [7].

- Métodos de extracción de características de estadística: este tipo de métodos describen las características que contienen las imágenes con el uso de la estadística, como pueden ser histogramas o momentos invariantes. Los métodos más utilizados son los siguientes:

- PCA (*Principal Component Analysis*): es una técnica muy popular que permite extraer vectores de características con menor dimensionalidad de las facciones más significativas de la cara, llamadas *eigenfaces* [8] o componentes principales, transformando la imagen a un subespacio. Se considera una de las técnicas que mayor rendimiento ofrece, sin una pérdida de información importante como se puede observar en la Figura 4.



Figura 4 Imágenes proyectadas en el espacio PCA para obtener eigenfaces

- ICA (*Independent Components Analysis*): esta técnica es una generalización de la técnica anterior PCA, cuya principal diferencia con esta es que no solo se basa en estadísticas entre píxeles sino que también es sensible a sus dependencias estadísticas. Su principal objetivo es descomponer la imagen de la cara en una combinación lineal de fuentes independientes.
 - LDA (*Linear Discriminant Analysis*): deriva del PCA y es una técnica de aprendizaje supervisado que se utiliza para clasificar los datos. El LDA tiene como misión conseguir una proyección de datos con una dimensión menor o igual que la de los datos que recibe como entrada. La principal diferencia con el método de PCA es que esta no intenta reducir el error que se comente en la representación de los datos, como si sucede en PCA.
- Método LBP (*Local Binary Patterns*): es uno de los métodos más usados en el ámbito de la extracción de características en el área del reconocimiento facial, con el fin de optimizar las etapas de identificación y clasificación de caras. Posee gran robustez frente a los cambios de iluminación. A partir de un pixel central se toman los píxeles vecinos y se binarizan en función de un valor umbral en escala de grises dependiendo si el valor del pixel vecino es mayor o menor que el del umbral, como se puede ver en la Figura 5.

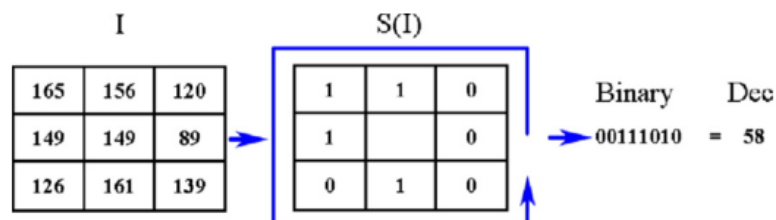


Figura 5 Ejemplo del método LBP [9]

- Método FACS (*Facial Action Coding System*): su objetivo consiste en identificar los músculos faciales que provocan cambios en la cara a la hora de expresar una emoción. A las acciones de estos músculos se las denominan AUs (Unidades de Acción). Es el método en el cual se basa este proyecto, por lo que se explica posteriormente con mayor detalle.

Después de obtener cada una de las características relevantes en el rostro a través de alguno de los métodos explicados anteriormente se procede a identificar la emoción presente en la cara por medio de un método de clasificación de la expresión. Pero para ello es necesario realizar un aprendizaje previo del sistema de reconocimiento facial. Para poder llevarlo a cabo se requiere el uso de un gran número de imágenes de rostros con las diferentes emociones que se desean identificar, lo que conlleva bases de datos grandes, así el sistema una vez que se aplican los métodos de extracción de características y clasificación de la emoción va aprendiendo las características típicas que se corresponden con cada emoción lo que le permite diferenciarlas y clasificarlas correctamente.

Respecto a los métodos de **clasificación de la expresión** se explican a continuación los más importantes:

- **Método SVM (*Support Vector Machine*):** es un método de clasificación de la emoción automático, creado en 1992, que no requiere supervisión y que se basa en el aprendizaje de máquinas (*Machine Learning*). Se considera un método útil en tareas de reconocimiento de patrones, incluyendo la cara, y de expresiones faciales. Permite determinar si un vector de entrada, como puede ser una Unidad de Acción, pertenece o no a un grupo (emoción) que se ha definido con anterioridad. Por medio de la construcción de un plano de N dimensiones que separa los datos en dos categorías diferentes. Por ejemplo la separación de los puntos de la cara que han sido detectados que pertenecen a una emoción en concreto frente a los puntos que no pertenecen a dicha emoción.
- **Métodos basados en el algoritmo *Adaboost*:** es un clasificador que entrena iterativamente una serie de clasificadores base para que cada clasificador nuevo ponga mayor atención en los datos que han sido clasificados como erróneos. El clasificador final se obtiene por tanto a partir de una ponderación de los clasificadores débiles. Se considera más rápido que SVM pero si el número de muestras es pequeño la clasificación que realiza no se considera óptima.
- **Métodos basados en Redes Bayesianas:** las redes bayesianas dado un conjunto de características presentes en los objetos permiten calcular la probabilidad de un conjunto de clases, en este caso emociones, para clasificar los objetos en la emoción más probable. Estas redes se componen de un grafo que posee una serie de nodos los cuales están asociados a una característica. Estos métodos tienen una buena precisión en la clasificación pero requieren gran cantidad de parámetros y muestras para que los resultados que se obtienen sean reales.
- **Método basado en Redes Neuronales (*Neural Networks*):** el cual ha evolucionado a ANN (*Artificial Neural Networks*), se inspira en la observación de los sistemas de aprendizaje que desarrollan las redes neuronales de los seres humanos, intentan imitar al cerebro humano en la tarea del reconocimiento. Las redes neuronales se componen de una serie de neuronas o nodos las cuales reciben una o más entradas y emite una o varias salidas, cuyo valor es 1 o 0 determinado mediante condiciones previas.

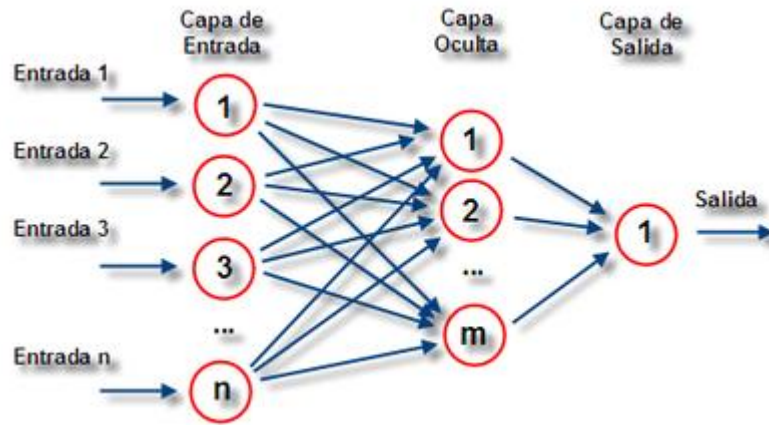


Figura 6 Red Neuronal [10]

- Métodos basados en Árboles de Decisión: se basan en una serie de atributos o características más relevantes de los objetos para llevar a cabo la clasificación. Se basa en una estructura tipo árbol donde el proceso comienza en el nodo superior o raíz del que parten las ramas principales que llevan a los nodos secundarios y así sucesivamente hasta llegar a las hojas que se corresponden con los posibles resultados como se puede ver en la Figura 7. Este es el método elegido en este proyecto para llevar a cabo la clasificación de la emoción y será explicado con más detenimiento en apartados posteriores.

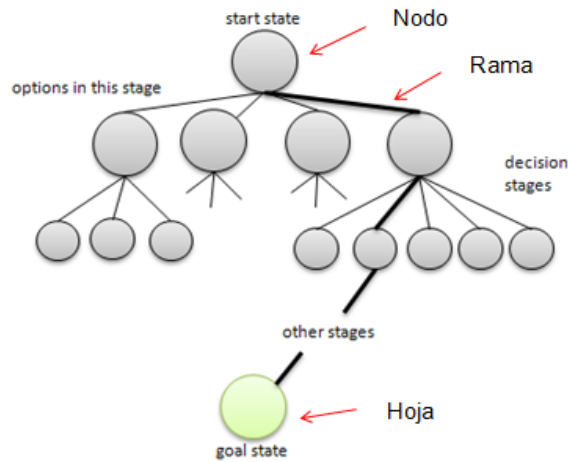


Figura 7 Árbol de decisión

De entre los métodos más relevantes comentados en este apartado el más utilizado en los últimos tiempos, en cuanto a la extracción de características se refiere, es el FACS (Sistema de Codificación de Acción Facial) y es por ello por lo que se decide utilizar en este proyecto.

2.2. Las Unidades de Acción como técnica elegida

El FACS fue creado por Paul Ekman y Wallace V. Friesen en 1978 [11], el cual se basa en la detección de cambios que se producen en la cara con las acciones de los músculos faciales. A estas acciones se las denominan AUs (Unidades de Acción) y hacen referencia a la acción de un único músculo o al conjunto de ellos.




















Ekman y Friesen diseñaron el FACS con el objetivo de caracterizar las expresiones faciales por medio de estas Unidades de Acción (AUs). Para desarrollar este sistema se utilizaron observaciones realizadas por expertos en el comportamiento humano en referencia a los cambios sutiles que se producen en la cara por la presencia de expresiones faciales. Por otro lado Ekman viajó durante cinco años alrededor del mundo con el propósito de comprobar científicamente si las expresiones eran distintas dependiendo de la cultura. Llegando a la conclusión de que las expresiones humanas son innatas y por tanto universales para todas las culturas, tal como sostenía Charles Darwin en 1872.

Según estos investigadores las emociones se clasifican en seis emociones básicas universales que son: *fear* (miedo), *anger* (ira), *happiness* (felicidad), *surprise* (sorpresa), *sadness* (tristeza) y *disgust* (aversión). Las cuales poseen una expresión facial estrechamente relacionada con cada una de ellas y estas a su vez se codifican por medio de un Sistema de Codificación de Acción Facial (FACS).

Del mismo modo Ekman y Friesen hacen referencia a las micro-expresiones descubiertas por *Haggard* y *Isaacs* en 1966 es decir, expresiones faciales involuntarias que se muestran en la cara de las personas de manera momentánea durante un tiempo inferior a un segundo, por lo que son muy difíciles de detectar. Siendo estas, según Ekman y Friesen, la clave para determinar si una emoción es real o fingida ya que al generar fugaces movimientos faciales se refleja la emoción original.

Con el FACS se definen 46 AUs, en la Tabla 1 se muestran las AUs que más se utilizan según este sistema para el reconocimiento facial de emociones. A su vez estas AUs se dividen en aquellas que se refieren a la parte superior del rostro y a las de la parte inferior del mismo. En cuanto a las que aluden a la parte superior se tienen: la AU 1, 2 y 4 para las cejas, la AU 5 y 7 para los párpados, la AU 6 para las mejillas y la AU 9 para la nariz. Para la parte inferior del rostro, como es la boca se tienen las AUs 10, 12, 15, 16, 17, 20, 22, 23, 24, 25, 26 y 27, tal como se muestra en la Tabla 1. Por otro lado el FACS se codifica según cinco niveles de intensidad, como se ve en la Tabla 2, para algunas AUs como por ejemplo las Unidades de Acción 25, 26 y 27. También es importante destacar que estos niveles no tienen intervalos iguales, es decir unos pueden cubrir una mayor cantidad de cambios en el rostro que otros como se representa en la Figura 8.

Tabla 1 Unidades de Acción (AUs) más importantes según el FACS

AU (FACS)	Imagen representativa	Definición	AU (FACS)	Imagen representativa	Definición
AU1		Interior de las cejas elevado	AU16		Labio inferior hacia abajo
AU2		Exterior de las cejas elevado	AU17		Barbilla elevada
AU4		Cejas bajadas	AU20		Labios estrechados y estirados en horizontal
AU5		Párpado superior elevado	AU22		Labios crateriformes
AU6		Mejillas elevadas	AU23		Labios tirantes, tensos
AU7		Párpados tensos	AU24		Labios presionados
AU9		Nariz arrugada	AU25		Labios separados
AU10		Labio superior elevado	AU26		Mandíbula caída (Boca entreabierta)
AU12		Comisuras de los labios elevadas	AU27		Boca abierta
AU15		Comisuras de los labios deprimidas (hacia abajo)			

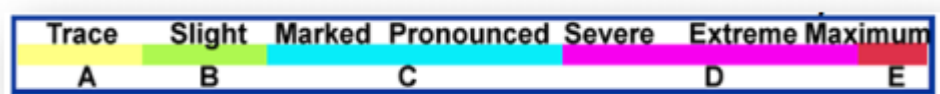


Figura 8 Escalas de los niveles de intensidad

Tabla 2 Niveles de intensidad según FACS

Niveles de intensidad	Denominación
<i>Trace(rastro ligero)</i>	A
<i>Slight(evidencia leve)</i>	B
<i>Marked/Pronounced (trazo marcado)</i>	C
<i>Severe/Extreme (nivel extremo, emoción intensa)</i>	D
<i>Maximum (máximo)</i>	E

El FACS permite relacionar estas Unidades de Acción con cada una de las expresiones que se pretenden detectar, es decir con las AUs que se encuentran presentes en cada una de las emociones básicas según este Sistema de Codificación de Acción Facial, como se muestra en la Tabla 3. Estas AUs se corresponden con todas aquellas que Ekman y Friesen observaron en los rostros tras su estudio, determinando que una expresión se puede componer de varias de ellas.

Tabla 3 Correspondencia de AUs para cada emoción según el FACS

Emoción	AUs (FACS)
Anger	4,5,7,10,17,22,23,24,25,26
Disgust	9,10,16,17,25,26
Fear	1,2,4,5,20,25,26,27
Happy	6,12
Sadness	1,4,6,11,15,17
Surprise	1,2,5,26,27

En cuanto a las ventajas y desventajas que presenta este sistema hay que mencionar que el Sistema de Codificación de Acción Facial posee una gran rapidez, siendo muy valioso a la hora de determinar información valiosa sobre las emociones así como su valencia (positiva o negativa) es decir, permite detectar si las reacciones emocionales son positivas o negativas. También se considera de fácil entendimiento y ejecución. En relación con las principales desventajas que presenta este sistema cabe destacar la dificultad para medir la intensidad emocional y la dificultad de identificar las AUs en ciertas situaciones como por ejemplo cuando la cara se encuentra inclinada o tapada.

Para el desarrollo de este proyecto se ha seleccionado el estudio del FACS con sus correspondientes AUs debido a que se considera uno de los métodos de extracción de características más sencillo, ya sea conceptualmente como de desarrollo. Por otro lado en el PFG de Sara González [3], en el que se inició el estudio de las AUs, no se llegó a encontrar las AUs o combinaciones de AUs adecuadas para el reconocimiento de cada emoción ni se

detectaron de manera correcta las AUs y los puntos característicos de la cara que son necesarios para determinar cada AU, como se explicará en apartados posteriores. En [3] se detectan 19 puntos de interés en la cara (Figura 9), un número de puntos que se considera adecuado para poder detectar las diferentes Unidades de Acción características de cada emoción. Sin embargo la localización de estos puntos en algunos casos no es lo suficientemente precisa y correcta como para poder identificar las AUs, debido a por ejemplo fallos en la localización de las regiones de la cara o en la segmentación, como se comentará en el apartado 3.1. Por estos motivos se decide realizar un estudio más amplio y detallado de estas Unidades de Acción, con el objetivo de establecer aquellas que puedan caracterizar las emociones, así como de la colocación de los puntos de la cara.



Figura 9 Detección de los 19 puntos característicos de la cara en [3]

2.3. Unidades de Acción

2.3.1. Estudio de las publicaciones

Tras un primer acercamiento al Sistema de Codificación de Acción Facial y a las Unidades de Acción (AUs) surge la necesidad de determinar cuáles de estas están presentes en cada emoción y cuál es la forma de combinarlas para poder detectar dicha emoción.

Para poder obtener una combinación adecuada de AUs para cada emoción, que permita reconocer las emociones básicas a detectar, se procede a realizar un estudio y análisis detallado de diferentes publicaciones existentes en este ámbito.

Las publicaciones que se eligen, de entre varias que han sido estudiadas, son aquellas que se consideran de mayor interés al proporcionar una mayor información en cuanto a la selección de AUs, a las combinaciones, a las bases de datos y métodos de clasificación utilizados. A continuación se muestra un resumen de las publicaciones en las que se utilizan las Unidades de Acción para detectar las emociones.

- [12]: Esta publicación se centra en la extensión de la base de datos Cohn-Kanade (CK) a Extended Cohn-Kanade (CK+), con un incremento del número de imágenes y sujetos. Esta base de datos se considera como una de las bases de datos más adecuadas para el uso de las AUs y la detección de la emoción. Así como una de las más utilizadas para el desarrollo y evaluación de algoritmos para el reconocimiento de expresiones faciales. Uno de los principales motivos por los que se lleva a cabo esta extensión de la CK, se debe a que las etiquetas de las emociones (*anger*, *contempt*, *disgust*, *fear*, *happy*, *sadness* y *surprise*) que se asocian a cada imagen de la base de datos no concuerdan con la emoción que realmente está presente en dicha imagen. Por lo tanto se añaden 107 imágenes y 28 sujetos nuevos, quedando como resultado un total de 593

imágenes con 123 sujetos de ambos sexos, con edades comprendidas entre los 18 y 50 años de edad y diferentes nacionales.

A partir de estas imágenes se procede a realizar el etiquetado de cada una de ellas codificándolas según el FACS [11]. Para ello se llevan a cabo tres pasos.

En primer lugar se comparan las imágenes con la Tabla 1 que muestra las Unidades de Acción según el FACS. Si una imagen satisface los criterios de AUs típicos de cada emoción de forma exacta, esta se codifica o etiqueta provisionalmente como perteneciente a esa emoción.

En segundo lugar se realiza una comparación menos rigurosa, es decir si una imagen contiene una AU que no pertenece a ninguna emoción se determina si es coherente o no con alguna de las emociones básicas que se pretenden detectar.

El último paso consiste en ver si la expresión o expresiones presentes en la imagen pertenecen definitivamente a esa emoción. Este paso se realiza debido a que en los códigos del FACS solo se describen las emociones en su pico de expresión y no se tienen en cuenta los cambios faciales que se producen hasta llegar a dicho pico.

Como resultado de este proceso se determinó que en 327 de las 593 imágenes se establecían los criterios de una de las emociones básicas.

Por otro lado en cuanto al método de extracción de características que utiliza es el AAM (*Active Appearance Model*) con el cual se realiza el seguimiento de la cara y se extraen las características faciales. Posteriormente se utiliza el método SVM (*Support Vector Machine*) para la clasificación de la emoción. Con estos métodos se pretende la detección de las Unidades de Acción y de la emoción, para ello se requiere diferentes imágenes para entrenamiento y prueba del sistema.

Para la detección de las emociones de *disgust*, *happiness* y *surprise* se usan las AUs que se encuentran en la malla AAM, ya que son emociones que causan una gran cantidad de deformación en la cara y afectan a zonas que son detectadas con facilidad por el sistema de reconocimiento. Sin embargo las emociones como *anger*, *sadness* y *fear* se reconocen por AUs difíciles de detectar y encontrar en la malla por lo que para su reconocimiento se utiliza información de textura.

Finalmente en la Figura 10 se puede ver el diagrama de bloques con el método de extracción de características y de clasificación, utilizados para el reconocimiento de la emoción.

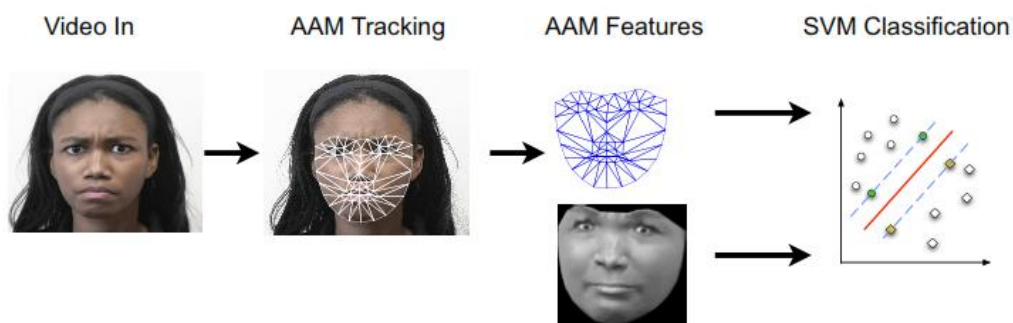


Figura 10 Diagrama de bloques usando los métodos AAM y SVM [12]

- En [13] se lleva a cabo un análisis comparativo de las diferentes técnicas de reconocimiento de las expresiones faciales de las publicaciones más destacadas de los últimos años.

Todo sistema de reconocimiento facial consta de cuatro etapas, como se ve en la Figura 11: detección de la cara en la imagen de entrada, pre-procesamiento,

extracción de características necesarias para describir una emoción y por último la clasificación de la emoción.

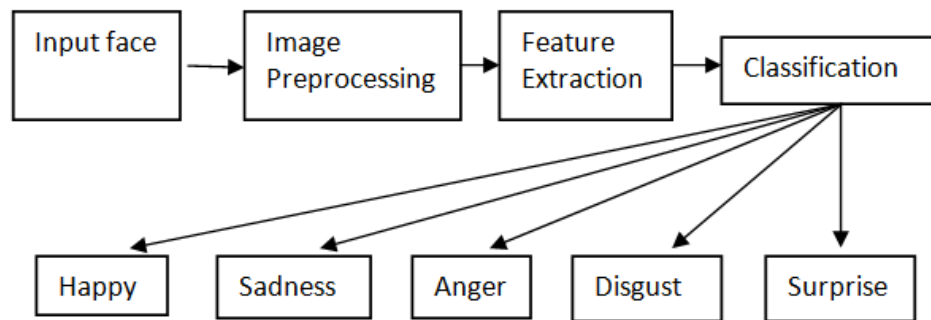


Figura 11 Etapas del reconocimiento facial [13]

Entre las distintas etapas cabe destacar el filtrado y la extracción de características, fundamental en estos sistemas de reconocimiento. Por este motivo esta publicación se centra en recabar toda la información referente a esta etapa, en cuanto a lo que a técnicas de extracción se refiere como por ejemplo: la técnica PCA (*Principal Component Analysis*), *Gabor Wavelets* [14], DCT (*Discrete Cosine Transform*), *Haar Wavelet Transforms*. Cada una de estas técnicas se explican en el apartado 2.1.

Por otro lado con el estudio de las diferentes publicaciones que se realiza en [13] se llega a la conclusión de que el sistema más completo para sintetizar la expresión facial es el desarrollado por Ekman y Friesen [15], basado como ya se ha comentado anteriormente en las Unidades de Acción o AUs. Hasta el momento solo los sistemas que se basan en marcadores son capaces de codificar de una manera fiable las AUs del FACS así como sus intensidades. De entre los muchos obstáculos y deficiencias que se han encontrado en el estudio realizado por [13] cabe destacar las deficiencias que se encuentran en las bases de datos, como por ejemplo la base de datos CMU [16] compuesta por 486 imágenes de 97 sujetos, con un número de imágenes diferente para cada emoción pudiendo provocar fallos en la clasificación de la emoción.

- En los sistemas de reconocimiento facial existen tres fases que caracterizan a las actividades faciales: el primer nivel consiste en inferir o localizar los puntos característicos faciales, el segundo nivel trata de localizar las Unidades de Acción facial (AUs) definidas en el FACS [11] a partir de los puntos anteriormente localizados y en el tercer y último nivel consiste en identificar las seis expresiones faciales básicas, con la ayuda de las AUs, que representan la contracción global de los músculos y los estados de emoción humana.

En [17], a diferencia de los estudios anteriores que solo se centran en una o dos de las fases anteriores por separado, se introduce una unificación probabilística con el uso de la Red Bayesiana Dinámica (*Dynamic Bayesian Network*, DBN) donde se representa la evolución facial en las diferentes fases llevando a cabo diferentes experimentos donde se ilustra la eficiencia y viabilidad del modelo propuesto en las tres fases anteriores. En el modelo que se propone en esta publicación, el cual se basa en DBN, el flujo de información es bidireccional es decir no solo el seguimiento de las características faciales puede ayudar al reconocimiento de la expresión o de la AU sino también este reconocimiento puede contribuir a mejorar el rendimiento del seguimiento de las características faciales.

Con los puntos característicos de la cara se codifica la información más importante sobre la forma y deformación de la misma. Las tecnologías de detección y seguimiento de los puntos se pueden clasificar en dos categorías: el modelo libre y el modelo basado en algoritmos de seguimiento. Un modelo es aquel que permite determinar un resultado a partir de unos datos de entrada. En el modelo libre se realiza el seguimiento sin el conocimiento a priori del objeto, con una búsqueda local e individual de la mejor posición coincidente con el mismo. En el modelo basado en algoritmos como por ejemplo ASM (*Active Shape Model*) [18], DAM (*Direct Appearance Model*) [19], AAM (*Active Appearance Model*) [5], se centran en el modelado de la forma de los objetos.

En cuanto al reconocimiento de las AUs y de las expresiones faciales se podrían clasificar los métodos actuales en dos categorías según [17]: métodos basados en imágenes y métodos basados en modelos. Los primeros se basan en el reconocimiento de las AUs por la observación de los cambios que se producen en la cara, clasificando las AUs que se encuentran en cada una de las imágenes que se analizan independientemente. A su vez este método basado en imágenes se lleva a cabo en dos etapas: en primer lugar se extraen diversas características faciales y en segundo lugar las expresiones y AUs se identifican a través de diferentes métodos de reconocimiento como son las Redes Neuronales [10], SVM [20], clasificadores AdaBoost [21]. La desventaja que presentan los métodos basados en imágenes es que se tienen que reconocer cada AU o combinaciones de AUs de manera individual y estáticamente, sin tener en cuenta las posibles relaciones semánticas entre AUs. En la segunda categoría correspondiente a los métodos basados en modelos se elimina este problema, con el uso de las relaciones entre Unidades de Acción y el reconocimiento de estas simultáneamente y no individualmente.

En [17] además de modelar las relaciones entre AUs, se utiliza la información de los puntos característicos, de las expresiones y el acoplamiento y las interacciones entre ellos para poder reconocer las emociones. Así pues en [17] se construye un modelo DBN (*Dynamic Bayesian Network*) (modelo gráfico que se dirige y compara con modelos dinámicos anteriores) para modelar las interacciones bidireccionales entre las diferentes fases de actividades faciales, reconociendo estas tres fases (puntos faciales, AUs y expresión) de manera simultánea. Para ello se detectan 26 puntos característicos y se reconocen 15 AUs más comúnmente usadas (AU 1, 2, 4, 5, 6, 7, 9, 12, 15, 17, 23, 24, 25, 26 y 27). La selección de las AUs adecuadas para el reconocimiento de cada emoción se basa principalmente en la frecuencia de aparición u ocurrencia de cada Unidad de Acción, y su importancia a la hora de caracterizar las 6 emociones básicas (*surprise, happiness, sadness, disgust, anger y fear*).

Las AUs describen el movimiento de las partes de la cara y por tanto el movimiento de los puntos característicos. A su vez la deformación de la cara refleja la acción de las AUs. Cuando se habla de las AUs y las combinaciones entre ellas hay que tener en cuenta que algunas de estas pueden ser no aditivas, es decir si se analiza la aparición de una AU en una combinación es diferente de si se analiza de forma independiente. Por ejemplo cuando se produce la AU1 por sí sola, indica tristeza y la AU5 indica sorpresa, sin embargo la combinación del AU1 y AU5 puede indicar miedo. También hay algunas Unidades de Acción que son mutuamente excluyentes ya que no se pueden dar ambas al mismo tiempo, como es el caso de las AU23 y AU25 al mismo tiempo. Estas relaciones e incertidumbres son representadas por el modelo propuesto en esta publicación.

En la Figura 12 se puede ver el diagrama de bloques del sistema de reconocimiento de acción facial con sus diferentes etapas y los métodos utilizados en cada una de ellas como ya se ha comentado.

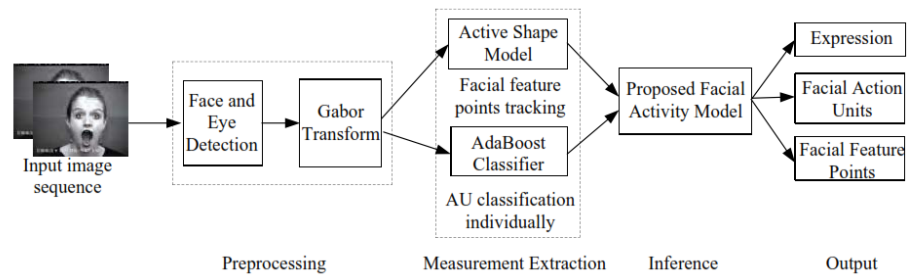


Figura 12 Diagrama de bloques del sistema de reconocimiento de acción facial [17]

El modelo propuesto se evalúa en dos bases de datos: la CK+ [12] y la MMI [22]. Una ventaja importante de la CK+ en comparación con la CK, como ya se ha dicho, es que las etiquetas de emoción en la CK+ se han revisado, asegurándose de que la expresión que se ve en cada imagen corresponde con su etiqueta, mientras que en la otra base de datos las etiquetas se corresponden con las que se les dijo a los actores de esta base de datos que expresaran sin una revisión posterior. Con el fin de evaluar la capacidad del modelo propuesto, se entrena el modelo con la base de datos CK+ y se prueba con MMI.

- El propósito de [23] consiste en investigar qué cambios faciales son los más frecuentes en las expresiones *happy*, *anger*, *sadness* y *fear* y cuáles de ellos son esenciales para un reconocimiento adecuado de cada una de las emociones. La emoción de *disgust* no se utiliza en este estudio porque se supone que no puede representar una emoción pura sino una combinación de otras emociones universales. Al igual que *surprise*, la cual se considera que su valencia depende de un evento desencadenante y por tanto puede ser cualquiera de las otras emociones en un inicio rápido.

Para ello se examinaron 128 imágenes con las expresiones faciales mencionadas anteriormente en 105 sujetos, usando un método de regresión logística. Los sujetos de las imágenes fueron reclutados de la Universidad de Drexel en Filadelfia. De los 105 participantes 63 eran hombres y 42 mujeres, 71 de raza caucásica, 9 afroamericanos, 4 hispanos y 21 de origen asiático. Para obtener las imágenes con las expresiones faciales los participantes tenían que pulsar un botón para indicar su cara neutra y otro cuando evocaban la emoción concreta, llegando a la conclusión de que cada emoción se puede identificar por un grupo de AUs características y distintos para cada emoción. Para determinar que AUs según el FACS estaban presentes en cada emoción, se calcula mediante un análisis visual el número de veces que aparece una AU para cada imagen agrupadas por emoción. Las Unidades de Acción (AUs) que se producen en al menos un 25% de las imágenes de una emoción se consideran características de la emoción es decir, importantes para reconocer dicha emoción.

En las pruebas que se llevan a cabo en esta publicación las AUs se definen como presentes o ausentes (1 o 0 respectivamente) para cada emoción. Las AUs características según el FACS que propone [23] para cada emoción se muestran en la Tabla 1.

Para el análisis de los datos se usan modelos de regresión logística, que se consideran los más populares para datos binarios (ausencia o presencia de una AU) correlacionados. La regresión logística es apta, con el uso de Ecuaciones de Estimación Generalizadas (GEE), para establecer la independencia o agrupación de Unidades de Acción pudiendo así evaluar si existe una relación entre el reconocimiento de una cierta emoción y la presencia de las AUs. Así pues con el uso de estos modelos y en

comparación con lo establecido anteriormente, se muestran en la Tabla 4 las siguientes relaciones entre emociones y AUs según [23]:

Tabla 4 AUs únicas y ausentes para cada emoción según [23]

	AUs únicas	AUs ausentes
Anger	9,16	1
Disgust	----	----
Fear	2	7,10
Happy	6,12	4,20
Sadness	17	26
Surprise	----	----

[23] establece también que las diferencias entre las AUs características establecidas en los diferentes estudios que se han realizado en este ámbito del reconocimiento facial de emociones, se pueden deber a las diferentes metodologías que se usan para obtener las expresiones faciales. Otro posible motivo se puede deber a la forma en la que se capturan las expresiones a la hora de generar las imágenes, ya que se puede dar el caso de que la imagen no se capture en el momento donde la expresión tiene su máxima intensidad.

Así como que las AUs tienen una sensibilidad limitada para detectar los pequeños cambios que se producen en la cara y las microexpresiones. Por lo tanto en [23] todas las emociones son de intensidad extrema para reducir la posible aparición de expresiones ambiguas.

- Basado en las investigaciones que desarrollan Paul Ekman [15] y Charles Darwin, en [24] se determinan los siguientes movimientos musculares o AUs asociados a cada una de las emociones siguientes:
 - *Anger*: donde se ven involucrados los músculos de las cejas como es el AU4 y los músculos de la boca ya sea abierta o cerrada e incluso enseñar los dientes. Por otro lado se dice que esta expresión se detecta más fácilmente en los rostros masculinos que los femeninos.
 - *Surprise*: según [15] esta emoción junto con *fear* son dos de las más difíciles de distinguir, afirmando que las culturas alfabetizadas son las únicas capaces de distinguirlas. Para que se puedan distinguir se señala que el movimiento fundamental es el de la boca, donde en *surprise* se mueve verticalmente mientras que en *fear* es horizontal. Las AUs implicadas para el reconocimiento de *surprise* son la AU 1, 2, 5 y 25.
 - *Fear*: en este caso se establece que las AUs más características son la AU 1, 2, 5, 20 y 25.
 - *Sadness*: según [24] es importante señalar que para el reconocimiento de esta emoción se tiene que prestar atención en el movimiento interior de las cejas hacia arriba (AU 1), pero no exterior, y el de los bordes de los labios hacia abajo (AU 15).
 - *Happy*: para esta emoción se considera fundamental el AU12 (comisuras de los labios elevadas). También se establece una posible diferenciación entre una sonrisa real y otra falsa. Esto se consigue con la contracción de los músculos que rodean la cuenca del ojo (movimiento ocasionado por la AU 6) es decir,

cuando una sonrisa es real las partes externas del ojo se aprietan hacia abajo y se generan arrugas o patas de gallo junto con la elevación de las comisuras de los labios (AU 12). Mientras que si es falsa solo tiene lugar la AU 12.

- *Disgust*: esta es una emoción no solo presente en seres humanos sino también en especies del reino animal, como por ejemplo perros y lobos cuando enseñan sus dientes cuando amenazan. Aquí tienen lugar la AU 9 (nariz arrugada) y AU 10 (labio superior elevado).

- En [25] se opta por analizar el reconocimiento de las emociones en la infancia. Con el estudio de este documento se pretende tener otra perspectiva y un mayor conocimiento en esta área no solo en adultos sino también en niños. Para ello se utiliza el método de análisis del FACS [11] y un modelo de aprendizaje visual.

El origen de las investigaciones sobre la expresión facial radica en Charles Darwin [26] que establece que las emociones son universales, resultando ser indispensables para la comunicación y la supervivencia de los individuos. Tras el estudio de numerosos investigadores en esta área se determina que las emociones están definidas por cambios faciales concretos, que permiten su identificación de manera precisa por todo tipo de individuos. Concretamente cabe destacar la labor realizada por los investigadores Ekman y Friesen [15] quienes elaboraron técnicas objetivas para la codificación de acciones faciales, FACS.

El principal objetivo de este estudio se centra en determinar si en los lactantes o niños se producen las emociones básicas e innatas. Para ello se realizan juicios subjetivos y códigos observacionales. Los juicios subjetivos se basan en solicitar a diversos observadores que indiquen que términos emocionales o características, sacados de una lista, son los que permiten identificar cada expresión. En cambio con los códigos o estudios observacionales, también llamados estudios de componentes expresivas, se busca la medición directa mediante la observación de los cambios faciales que determinan las emociones. La gran diferencia entre ambos es que los primeros se consideran menos precisos que los segundos, ya que en los primeros no se tienen en cuenta los elementos o criterios en los que se centra cada observador para determinar si se trata de una emoción u otra.

Para este estudio también se debe tener en cuenta los estímulos que se utilizan para generar las emociones en los lactantes o niños. Algunos de ellos son por ejemplo: la emoción de alegría se genera con la interacción con la madre y la realización de juegos, para la ira con la retirada de objetos de las manos de los niños, el desagrado con zumos ácidos y para el miedo con pinchazos, caídas y ruidos intensos.

El punto de partida de este análisis se apoya en los estudios realizados por Ekman y Friesen para las seis emociones básicas: *anger, happy, surprise, fear, sadness y disgust*. Cada una de las respuestas faciales se graba en video para su posterior análisis cuadro a cuadro, llegando a la conclusión que se muestra en la

Tabla 5 en cuanto a las Unidades de Acción según el FACS que se detectan tras el análisis visual de cada una de las imágenes.

Por último y como conclusión en [25] se establece que los niños son capaces de reproducir los mismos cambios faciales que los adultos.

- Con el propósito de encontrar formas de reconocimiento facial diferentes usando las AUs, se estudia la publicación [27]. Este se basa en el reconocimiento de emociones

faciales con el uso de Redes Neuronales (ref-redNeuronal) y Unidades de Acción, así como el análisis semántico de la voz con el uso de un robot humanoide.

La investigación se centra en la producción de un robot inteligente con la habilidad de inferir las emociones e interpretar contextos de diálogo. Con el fin de lograr este objetivo se usa el FACS que emplea AUs para representar las actividades musculares que describen las expresiones faciales. Así como Redes Neuronales (NN) que se utilizan para analizar respectivamente las Unidades de Acción de la parte superior e inferior del rostro y para la detección de la emoción.

El sistema de reconocimiento de emociones faciales que se propone incluye dos Redes Neuronales que se basan en el análisis de las características de la parte superior e inferior de la cara. Las AUs que se usan correspondientes a la parte superior del rostro son la AU 1, 2, 4, 5, 6 y 7 y para la parte inferior la AU 9,10, 12, 15, 16, 17, 23, 24, 25, 26 y 27, haciendo un total de 17 AUs utilizadas. Estas AUs se usan como entradas para la red neuronal de reconocimiento de las emociones faciales, en la Figura 13 se muestra el flujo de datos para el proceso de reconocimiento de emociones.

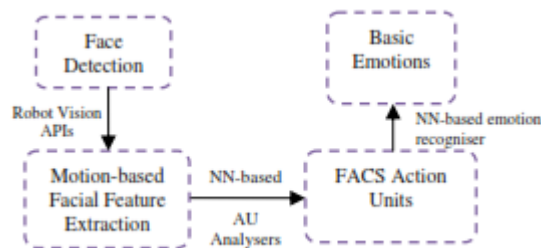


Figura 13 Flujo de datos para el proceso de reconocimiento usando en [27]

Para llevar a cabo el reconocimiento se les pide a sujetos que muestren en sus rostros la expresión facial neutra y la de las seis emociones que se pretende detectar durante 1 o 2 segundos, frente al robot. Con este se procesan en tiempo real los datos faciales y se obtiene la información de los puntos característicos de la cara. Después con las Redes Neuronales se llega a la detección de la emoción. A continuación el robot pide a los sujetos que indiquen con un 0 el rostro neutro que han mostrado y con un 1 la expresión de la emoción. Tras esto, el robot comunica al usuario si las emociones han sido o no reconocidas en base a la propia interacción con el mismo. Otro procedimiento se basa en mostrar al robot imágenes de la base de datos CK en una pantalla.

El robot es capaz de extraer las características faciales en tiempo real tanto de los sujetos como de las imágenes de la base de datos es decir, los puntos característicos de la cara. En concreto se extraen 31 puntos que proporcionan información de la boca, nariz, ojos y cejas. Para el entrenamiento y obtención se las AUs superiores del rostro se usan 20 puntos mientras que para las AUs inferiores se usan 11.

Los porcentajes de acierto para el analizador superior de AUs es del 71,3%, para el analizador inferior de Unidades de Acción es del 78,6% y para la red de reconocimiento se obtiene un 71,3% de precisión.

- Para poder determinar la producción de expresiones faciales en las personas en [28] se usan dos metodologías diferentes. La primera se basa en el uso de electrodos en el rostro con los que se realizan mediciones de los distintos movimientos faciales que se producen. La segunda se centra en el uso del FACS y sus Unidades de Acción muscular (AUs).

En este segundo método se ponen a prueba las expresiones prototípicas propuestas por Ekman y Friesen en situaciones reales. Llegando a la conclusión de que las AUs tienen una gran variabilidad para cada una de las emociones, ya que estas cumplen funciones adaptativas y no tienen porqué estar presentes en todas las situaciones.

Por este motivo en [28] se realizan estudios de diferentes sujetos como campeones olímpicos, fans del fútbol, toreros y niños en edad de vacunación entre otros. Para poder determinar así que Unidades de Acción se producen cuando estos sujetos muestran las emociones básicas, con el propósito de ver si se cumplen las configuraciones prototípicas de Ekman y Friesen.

Tras este estudio se observa que por ejemplo para la emoción de *fear* están presentes las AUs 4 y 25, en la mayoría de los casos como combinación de ambas 4+25. Para el caso de *happy* las AUs características son la 7, 12 y 26 y sus posibles combinaciones 12+26 o 7+12+26 y cuando se muestra la emoción de *sadness* se ven las AUs 1, 4, 15 y 17.

Las Unidades de Acción que se producen se evalúan por codificadoras expertas en el FACS. Estas se encargan de determinar de una manera visual, si las AUs están presentes o no en cada una de las emociones que los sujetos analizados muestran en sus rostros.

- También se procede a analizar el [29] donde se utilizan el FACS para el análisis de las expresiones faciales en sujetos con trastornos psiquiátricos. El sistema que se presenta es un sistema automatizado que se basa en el FACS, con el que se miden los movimientos musculares individuales o combinaciones a través de las AUs de forma automática sin la necesidad de la intervención de ningún operador.

A pesar de las ventajas que ofrece el uso del FACS para el análisis de las expresiones faciales, se requiere una amplia formación y un gran consumo de tiempo.

Este método se aplica en videos donde se proporciona la información detallada de los movimientos de los músculos para cada sujeto y con ellos se computa la frecuencia de aparición de las AUs, ya sea de forma individual o en combinación de varias.

Con [29] se establece una nueva línea de investigación que se centra en la clasificación automatizada de AUs, consistente en múltiples etapas como son la detección del rostro, el seguimiento y extracción de características y la clasificación de la emoción.

El sistema de análisis automatizado se inicia con la detección de la cara y sus regiones, usando el algoritmo de *Viola & Jones* [4], ya sea para imágenes o video. Después se localizan los componentes faciales y se alinea la cara. En este caso los puntos de referencia o puntos característicos son seguidos y localizados con el modelo ASM (*Active Shape Model*) [18] siendo este un modelo de caras deformables. Para entrenar al modelo ASM se recogen 159 puntos de referencia de un conjunto de imágenes fijas que también se utilizan para entrenar al clasificador de AUs. El número típico de puntos característicos que se suele usar oscila entre 30 y 80 puntos en cambio en [29] se usan 159 para que se detecte con una mayor precisión el movimiento de cada uno de los componentes faciales.

Por otro lado se usan los puntos anteriormente detectados para extraer las AUs características, tanto geométricas como de textura. Por ejemplo las AUs 1 y 2 pueden ser detectadas de los cambios geométricos que a su vez tienen asociados cambios de textura como arrugas en la frente. Para alinear los puntos de referencia y sacar las características geométricas se crea una malla de 158 vértices y 436 bordes, con la deformación de esta malla por compresión o expansión de los bordes se refleja la contracción o relajación de los músculos faciales. En el caso de las características de textura se usan los filtros Gabor [14].

Posteriormente un algoritmo clasificador toma los puntos de como entrada y produce como salida una decisión binaria, donde se indica si una AU está presente o no. El clasificador que se utiliza es AdaBoost [21] ya que se adapta tanto a las características geométricas como a las de textura.

En cuanto a la adquisición de los videos con las emociones se recurre a cuatro pacientes sanos y cuatro con esquizofrenia, dos hombres y dos mujeres. Se detectan cinco de las emociones básicas: *happy, sadness, fear, anger* y *disgust*. Previo a la grabación se les pide a los sujetos que describan situaciones emocionales, con estas se realiza un resumen en viñetas para así poderse enseñar de manera narrativa.

Por último dos evaluadores se encargan de determinar la presencia o ausencia de las AUs en cada una de las imágenes sacadas de los videos para poder comparar los resultados con los del algoritmo clasificador para cada emoción.

- Por último se utilizan las AUs y combinaciones de AUs que se proponen en [30], en el cual se basa el trabajo que se ha realizado durante las prácticas en el centro de investigación CITSEM [1] junto con mi compañero Diego Zapatero.

2.3.2. Análisis de las publicaciones para la extracción de reglas comunes

Una vez que se ha resumido y explicado cada una de las publicaciones que se consideran interesantes y que se basan en el reconocimiento de emociones mediante Unidades de Acción, se decide comparar lo que se dice en cada una de ellas para intentar obtener unas conclusiones comunes para todas las publicaciones o expertos en la medida de lo posible.

Para ello en la

Tabla 5 se pueden ver las AUs que se consideran importantes para reconocer cada emoción en las diferentes publicaciones y en la Tabla 6 las combinaciones de AUs que se han publicado como reglas. El motivo de diferenciar estas dos tablas se debe a que en algunas de las publicaciones no se especifica claramente la forma en la que combinan las AUs es decir, simplemente citan aquellas que consideran importantes o necesarias para detectar una emoción sin llegar a establecer con cuales de ellas se pueden combinar porque no con todos los métodos es posible extraer estas reglas. A partir de toda la información que proporcionan estas tablas se irán realizando distintos estudios para poder llegar a un análisis final acerca del uso de las Unidades de Acción.

A la izquierda de cada una de las tablas (

Tabla 5 y Tabla 6) se muestra la referencia de cada una de las publicaciones en las que aparecen las AUs que consideran más importantes o sus combinaciones.

Tabla 5 AUs presentes en cada emoción, establecidas por las publicaciones estudiadas, según el FACS

		Anger	Disgust	Fear	Happy	Sadness	Surprise
[12]	AUs	23,24	9,10	1,2,4	12	1,4,6,11,15	1,2,5
[13]		4,5,22,24	10	1,2,5	12	1,7	1,2,5,26
[17]		4,5,7,23,24	9,17	1,4,5,7	6,12, 25	1,4,15,17	1,2,5,26, 27
[23]		4,5,7,9,10, 16,20, 25, 26	-----	1,2,4,5,20, 26	1,6,7,10,12,2 5,26	1,4,7,10,15, 17,20,25	-----
[24]		4,23,26	9,10	1,2,5,20,25	6,12	1,15	1,2,5,25
[25]		4,6,20,25, 26,27	9,15,16, 17,25,26, 27	1,2,4,5, 20,25,26, 27	6,12	1,4,15,17,25, 26,27	1,2,26,27
[27]		4,5,7,10,17, 23, 25	9,10,15,16,17, 25	1,2,4,5,20,25	6,12	1,4,6,15	1,2,5,26, 27
[28]		4,5,7,10,17, 22,23,24,25, 26	9,10,15,16,17, 25,26	1,2,4,5,20,25, 26,27	6,12	1,4,6,11,15	1,2,5,26, 27
[29]		9,16	-----	2	6,12	17	-----
[30]		4,5,7,10,17, 22,23,24,25, 26	9,10,16,17,25, 26	1,2,4,5,20,25, 26,27	6,12	1,4,6,11,15, 17	1,2,5,26, 27

Para poder interpretar la Tabla 6 se explica el significado de los símbolos utilizados: el “+” indica un AND es decir una suma de AUs y “|” indica un OR es decir el uso de una AU u otra. Con el uso de estos símbolos se muestran las posibles combinaciones con distintas Unidades de Acción para cada una de las seis emociones básicas que se quieren reconocer.

Tabla 6 Combinaciones de AUs para cada emoción, establecidas por las publicaciones estudiadas, según el FACS

		Anger	Disgust	Fear	Happy	Sadness	Surprise
[12]	Combinaciones de AUs	23 24	9 10	1+2+4	12	1+4+ (11 15) 6+15	1+(2 5)
[13]		4+5+(22 24)	10	1+2+5	12	1+7	1+2+5+ 26
[27]		4+5+7+17+23 4+5+7+10+23+25	9 9+15+16 9+17 10+16+25	1+2+4+5+20+ 25 1+2+4+5+25	12 6+12	1+4+15 6+15	1+2+5+ (26 27)
[28]		4+5+7+10+22+23+25+ 26 4+5+7+10+23+25+26 4+5+7+23+25+26 4+5+7+17+(23 24) 4+5+7+(23 24)	9 10 (9 10)+17 9+16+15+26 10+16+ (25 26)	1+2+4+5+20+ (25 26 27) 1+2+4+5+ (25 26 27)	6+12	1+4+11+ 15 1+4+15 6+15	1+2+5 +(26 27)
[30]		4+5+7+10+22+23+ (25 26) 4+5+7+10+23+(25 26) 4+5+7+17+(23 24) 4+5+7+ (23 24) 4+(5 7) 17+24	9 10 (9 10)+17 (9 10)+16+ (25 26)	1+2+4 1+2+4+5+20+ (25 26 27) 1+2+4+5+ (25 26 27) 1+2+4+5 1+2+5+ (25 26 27) 5+20+ (25 26 27) 5+20 20	12 6+12	1+4 1+4+ (11 15) 1+4+15+ 17 6+15 11+17 1	1+2+5+ (26 27) 1+2+5 1+2+ (26 27) 5+(26 27)

Para que el contenido que se muestra en la

Tabla 5, acerca de las AUs que se pueden utilizar para cada emoción, quede más claro y para reducir la cantidad de información que se obtiene de todo este análisis se realiza la Tabla 7 donde se cuenta el número de veces que se menciona cada AU en las publicaciones para cada emoción. Por ejemplo para el caso de la emoción de *anger* la AU4 es mencionada en 8 de las 10 publicaciones estudiadas. Con este recuento se consigue ver la frecuencia de aparición de las mismas, siendo también un procedimiento realizado en algunas de las publicaciones estudiadas como por ejemplo en [17] y [23].

Esta información se estima de interés ya que a partir de ella se quiere averiguar cuáles son las AUs que se consideran más importantes para detectar cada emoción, eligiendo aquellas que se usan en un mayor número de publicaciones es decir las que más se detectan.

Tabla 7 Estadística de recomendaciones de las AUs a detectar por emoción sobre 10 publicaciones

AU	Anger	Disgust	Fear	Happy	Sadness	Surprise
1			9	1	9	8
2			9			8
4	8		7		7	
5	6		8			7
6	1			8	4	
7	5		1	1	1	
9	2	7				
10	4	6		1	1	
11					3	
12				10		
15		3			8	
16	2	4				
17	3	5			5	
20	2		6		1	
22	3					
23	5					
24	5					
25	5	4	5	2	2	1
26	5	3	4	1	1	6
27	1	1	3		1	5

Para poder determinar las AUs más características de cada emoción se debe establecer cuál es el límite de apariciones a partir del cual se puede determinar que una AU es importante para llevar a cabo la detección de la emoción. Ya que una AU que solo se menciona en una única publicación puede ser menos importante que aquellas que se mencionan en un mayor número de publicaciones.

Por tanto para establecer un criterio que indique cual debe ser el número mínimo de veces que una AU debe ser utilizada o mencionada en las diferentes publicaciones estudiadas se decide realizar una media aritmética como se muestra en la Tabla 8, también llamada promedio o media que se obtiene con la forma (3).

$$\bar{x} = \frac{\sum x_n}{N} \quad (3)$$

Donde x_n es el número de veces que se menciona una AU en las publicaciones estudiadas para cada emoción y N es el número total de AUs que se utilizan en cada emoción, como se puede ver en la Tabla 7 a partir de la cual se sacan los datos para realizar cada una de las medias

aritméticas. Para precisar más que AUs son las características para cada emoción y para conocer en que emociones destaca más una AU, se realiza también la media aritmética en horizontal, donde en este caso la N hace referencia a las emociones en las que se usa una AU.

Tabla 8 Medias aritméticas (en vertical) de la frecuencia de aparición de AUs para cada emoción

Emoción	N (vertical)	Media aritmética vertical (\bar{x})
Anger	15 AUs	4
Disgust	8 AUs	4
Fear	9 AUs	6
Happy	7 AUs	3
Sadness	12 AUs	4
Surprise	6 AUs	6

Una vez que se realizan estas medias se propone que el número de veces que se debe mencionar una AU como importante en la Tabla 7 es de 6 o más para *fear* y *surprise*, 4 o más veces para el caso de *anger*, *disgust* y *sadness*, y 3 o más para *happy*. Para mostrarlo más claramente en la Tabla 7 se indican en rojo las que superan las medias verticales y sombreadas en naranja las que superan las medias horizontales. Así pues con todo lo comentado el *set* o conjunto de AUs más importantes que quedaría como resultado, tras la aplicación de las medias aritméticas, y que permite llevar a cabo el reconocimiento de emociones se muestra en la Tabla 9. Las casillas sombreadas y marcadas con una X se corresponden con las AUs que pertenecen a cada emoción.

Tabla 9 Conjunto o set de AUs características para el reconocimiento de las emociones según las publicaciones estudiadas

Descripción de las AUs	AUs (FACS)	Anger	Disgust	Fear	Happy	Sadness	Surprise
Interior de las cejas elevado	1			X		X	X
Exterior de las cejas elevado	2			X			X
Cejas bajadas	4	X		X		X	
Parpado superior elevado	5	X		X			X
Mejillas elevadas	6				X	X	
Párpados tensos	7	X					
Nariz arrugada	9		X				
Labio superior elevado	10	X	X				
Comisuras de los labios elevados	12				X		
Comisuras de los labios hacia abajo	15					X	
Labio inferior hacia abajo	16		X				
Barbilla elevada	17		X			X	
Labios estrechados y estirados en horizontal	20			X			
Labios crateriformes	22	X					
Labios tirantes, tensos	23	X					
Labios presionados	24	X					
Labios separados	25	X	X	X			
Mandíbula caída (Boca entreabierta)	26	X	X	X			X
Boca abierta	27			X			X

Para poder llevar a cabo el reconocimiento de las emociones, se realiza una clasificación de AUs atendiendo a distintos criterios, tras haber establecido aquellas AUs que más se detectan en las publicaciones. Estos criterios servirán de ayuda a la hora de realizar el análisis final (apartado 2.3.4) con todas las conclusiones del estudio de las AUs que se han ido obteniendo en cada uno de los apartados.

El primer criterio que se aplica se basa en aquellas AUs que funcionan “exclusivas” es decir, las AUs que determinan una emoción de forma clara. Pudiendo reconocer la emoción con la sola presencia de la misma, independientemente de las demás ya que solo aparecen en una de las emociones que se quiere reconocer. Teniendo en cuenta la información de [29] y [23] se establece que las AUs exclusivas son las que aparecen únicamente en una de las emociones, como se muestran en la Tabla 10.

Tabla 10 AUs exclusivas

	AUs exclusivas
Anger	7,22,23,24
Disgust	9,16
Fear	20
Happy	12
Sadness	15
Surprise	----

El segundo, teniendo en cuenta las AUs que no deben estar presentes en una emoción para que pueda ser detectada, detección por exclusión presente en la Tabla 11. Para ello se utiliza la información presente en las publicaciones de [29] y [23] en cuanto a lo que a las AUs que no deben estar presentes se refiere.

Tabla 11 AUs de exclusión o no presentes

	AUs de exclusión o no presentes
Anger	1
Disgust	----
Fear	7,10
Happy	4,20
Sadness	25,26
Surprise	----

El tercero se basa en las AUs que se pueden detectar con los 19 puntos que se localizan en la cara, descartando aquellas que se refieren a puntos que no se detectan en este proyecto de reconocimiento facial de emociones. Por tanto el conjunto de AUs de la Tabla 9, tras aplicar este criterio, quedaría tal como se ve en la Tabla 12.

Tabla 12 Set o conjunto de AUs que se pueden detectar con los 19 puntos de la cara

	AUs para los 19 puntos de la cara
Anger	4,5,7,10,23,24,26
Disgust	9,10,17,26
Fear	1,2,4,5,26,27
Happy	6,12
Sadness	1,4,6,15,17
Surprise	1,2,5,26,27

Como se puede ver se han eliminado la AU 16 (labio inferior hacia abajo), AU 20 (labios estrechados y estirados en horizontal), AU 22 (labios crateriformes) y la AU 25 (labios separados) ya que son AUs que entra en conflicto a la hora de poder distinguirlas con el resto de AUs que se refieren a la boca y que aportan mayor información, como por ejemplo el AU 23, 24, 25, 26 o 27. Por lo que se seleccionan aquellas que se van a poder implementar y distinguir teniendo en cuenta los puntos faciales de los que se disponen. Todos estos aspectos se han teniendo en cuenta debido a las limitaciones que existen en la herramienta desarrollada en [3] en cuanto al número de puntos característicos de la cara de los que se dispone.

Para poder interpretar las AUs características establecidas en el conjunto de la Tabla 12 de una manera más visual e intuitiva se realiza un análisis mediante *clusters*, como se muestra en la Figura 14. El análisis *cluster* es una técnica estadística que se utiliza para clasificar un conjunto de elementos en grupos o *clusters*. Con esta técnica se puede ver la dependencia o independencia de cada una de las AUs que se utilizan para llevar a cabo el reconocimiento. Así se puede ver que AUs se corresponden con cada emoción, que AUs son las exclusivas para cada emoción o las que no deben estar presentes, tal y como se ha indicado en los criterios anteriores a tener en cuenta.

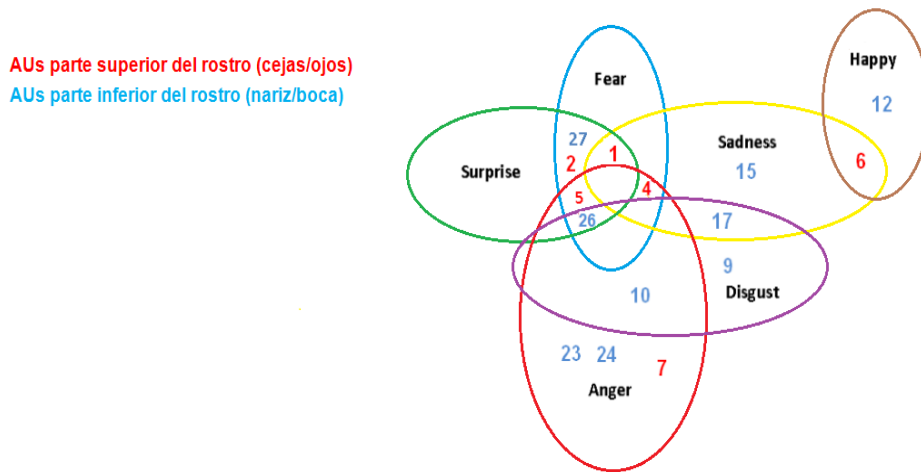


Figura 14 Cluster para el set o conjunto de AUs de la Tabla 12

En el *cluster* de la Figura 14 cabe destacar que cada emoción se ha dividido en grupos con sus correspondientes AUs características. Los AUs que corresponden a la parte superior del rostro, como son las cejas y los ojos, se han sombreado en color rojo y los correspondientes a la parte de la nariz y boca en azul, para poder distinguirlos de manera visual.

El cuarto y último criterio o aspecto que se tiene que tener en cuenta en este estudio de las AUs son las combinaciones de AUs más frecuentes una vez que se seleccionan las AUs características de cada emoción. Para ello se parte de la Tabla 6 donde se muestran las combinaciones propuestas por las publicaciones y se adaptan dichas combinaciones según el *set* o conjunto de AUs característico de cada emoción de la Tabla 9, quedando como resultado la Tabla 13 con todas las posibles combinaciones de AUs. Donde como ya se dijo anteriormente, el + indica la suma de AUs y el || la posibilidad de elegir una u otra AU.

Tabla 13 Combinaciones de AUs para el set o conjunto de AUs de la Tabla 9

	Anger	Disgust	Fear	Happy	Sadness	Surprise
Combinaciones de AUs			20			
			5+20			
	23 24	9 10	1+2+(4 5)			
	4+5+(22 24)	9+17	1+2+4+5		1+4+15	1+2+5+(26 27)
	4+5+7+10+23+(25 26)	10+16+25	1+2+4+5+20+		6+15	1+(2 5)
	4+5+7+10+22+23+25+26	10+16+	(25 26 27)	12	1+4	1+2+5
	4+5+7+10+22+23+(25 26)	(25 26)		6+12	1+7	
		(9 10)+17	1+2+4+5+(25 26 27)		1+4+15+1	1+2+(26 27)
	4+5+7+23+25+26	(9 10)+16+(25 26)			7	5+(26 27)
	4+5+7+(23 24)	(9 10)+26	1+2+5+(25 26 27)		1	
4+(5 7)		5+20+(25 26 27)				

Por otro lado se elabora una tabla (Tabla 14) donde se muestran las combinaciones de AUs no solo adaptadas para las AUs características (Tabla 9, que contiene el conjunto de AUs características para cada emoción) sino también para aquellas que se pueden detectar con los 19 puntos de la cara que se localizan (Tabla 12, que contiene el conjunto de AUs que se detectan con los 19 puntos de los que se dispone).

Tabla 14 Combinaciones de AUs para el set o conjunto de AUs de la Tabla 9 teniendo en cuenta las AUs que se pueden detectar con los 19 puntos de la cara

	Anger	Disgust	Fear	Happy	Sadness	Surprise
Combinaciones de AUs para los 19 puntos de la cara	23 24	9 10	1+2+(4 5)	12	1+4+15	1+2+5+(26 27)
	4+5+24	9+17	1+2+4+5	6+12	6+15	1+(2 5)
	4+5+7+10+23+26	(9 10)+17	1+2+4+5+(26 27)		1+4	1+2+5
	4+5+7+10+23	(9 10)+26	1+2+5+(26 27)		1+7	1+2+(26 27)
	4+5+7+23+26				1+4+15+17	5+(26 27)
	4+5+7+(23 24)				1	27)
	4+(5 7)					

2.3.3. Estudio subjetivo de las AUs

Por otro lado para tener una mayor conciencia de las AUs que pueden aparecer en cada una de las emociones y contrastar el análisis realizado en el apartado anterior, basado en distintas publicaciones estudiadas, se decide crear un segundo set o conjunto de posibles AUs y combinaciones de AUs características para cada emoción. Al hacer este estudio subjetivo se consigue ver si la selección de AUs de la Tabla 9 puede ser correcta o si todavía quedan AUs que se pueden descartar o añadir.

Para establecer esta segunda combinación de AUs se realiza un estudio detallado de múltiples bases de datos de imágenes que contienen las seis emociones básicas que se pretenden detectar. Las imágenes que se analizan son de las bases de datos CK+[12], MMI [22], JAFFE [31] y de [32] imágenes encontradas en Internet. Estas imágenes son de hombres, mujeres y niños de diferentes nacionalidades y edades. Por ejemplo en la Figura 15 se pueden observar algunas de las imágenes observadas para el reconocimiento de la emoción de *happy*.



Figura 15 Imágenes con la emoción de *happy* de [12], [22], [31] y [32]

Este análisis se lleva a cabo por comparación visual entre la imagen neutral y la imagen con la emoción, anotando en cada caso las AUs del FACS que se reconocen. El resultado del análisis se puede ver en la Figura 15.

Tabla 15 Set o conjunto subjetivo con todas las AUs encontradas en el estudio de diferentes imágenes

Descripción de las AUs	AUs (FACS)	Anger	Disgust	Fear	Happy	Sadness	Surprise
Interior de las cejas elevado	1			X		X	X
Exterior de las cejas elevado	2			X			X
Cejas bajadas	4	X	X	X		X	
Parpado superior elevado	5	X		X			X
Mejillas elevadas	6				X		
Párpados tensos	7						
Nariz arrugada	9	X	X				
Labio superior elevado	10		X				
Comisuras de los labios elevados	12				X		
Comisuras de los labios hacia abajo	15					X	
Labio inferior hacia abajo	16		X				
Barbilla elevada	17		X			X	
Labios estrechados y estirados en horizontal	20			X			
Labios crateriformes	22	X					
Labios tirantes, tensos	23	X	X			X	
Labios presionados	24	X					
Labios separados	25			X	X		
Mandíbula caída (Boca entreabierta)	26				X		X
Boca abierta	27	X				X	X

A continuación para determinar las AUs que se estiman características de las emociones es decir aquellas con las que se puede reconocer una emoción de una manera más clara y precisa, al igual que se hacía en el apartado anterior (Tabla 7), se tiene en cuenta la frecuencia de aparición u ocurrencia de cada una. Dando como resultado la Tabla 16, en la que se muestra el número de veces que se ha detectado una AU en las imágenes analizadas para cada emoción.

Cabe destacar que este recuento puede variar según las imágenes analizadas o el criterio subjetivo del evaluador, siendo aconsejable analizar las imágenes por múltiples evaluadores para tener una mayor cantidad de opiniones. En este caso el número de imágenes analizadas ha sido de 20 imágenes para cada emoción básica.

Tabla 16 Estadística de la presencia de AUs tras un estudio visual subjetivo

AU	Anger	Disgust	Fear	Happy	Sadness	Surprise
1			10		9	20
2			10			20
4	18	16	10		11	
5	2		10			20
6				20		
7						
9	5	17				
10		11				
11						
12				20		
15					17	
16		3				
17		12			16	
20			3			
22	3					
23	6	6			1	
24	9					
25				3		
26			17	17		2
27	2				2	18

Para seguir los mismos pasos realizados que con el análisis de AUs del apartado anterior, para determinar las AUs más características de cada emoción en este estudio subjetivo se debe establecer cuál es el mínimo número de apariciones que debe tener una AU para considerarla como importante. Por lo que con el uso de la ecuación (2) se obtienen las medias aritméticas tanto en vertical para cada emoción como en horizontal para cada AU, quedando como resultado la Tabla 17.

Tabla 17 Medias aritméticas de la frecuencia de aparición de AUs para cada emoción para el estudio subjetivo

Emoción	N (vertical)	Media aritmética vertical (\bar{x})
Anger	7 AUs	6
Disgust	6 AUs	11
Fear	6 AUs	10
Happy	4 AUs	15
Sadness	6 AUs	9
Surprise	5 AUs	16

Tras obtener la media vertical adecuada para cada emoción se estima que el número de veces que como mínimo se debe visualizar una AU en las imágenes analizadas para cada expresión es: para *anger* de 6 veces o más, para *disgust* 11 veces o más, *fear* mayor o igual a 10 veces, *happy* 15 o más, *sadness* 9 o más y finalmente para *surprise* de 16 veces o más. Los resultados de la media horizontal se muestran en la Tabla 16 de manera más visual con sombreado en naranja y en rojo las medias verticales, al igual que en apartado anterior. Estas medias cambian en función del número de imágenes analizadas y el número de AUs detectadas.

Finalmente en la Tabla 18 se puede ver el set o conjunto de AUs más importantes o características para el reconocimiento de cada emoción tras el estudio visual de imágenes y la aplicación de las medias aritméticas.

Tabla 18 Set o conjunto subjetivo de AUs características mayormente utilizadas

	AUs (FACS)
Anger	4,22,23,24
Disgust	4,9,10,16,17,23
Fear	1,2,4,5,20,26
Happy	6,12,25,26
Sadness	1,4,15,17
Surprise	1,2,5,27

Como en el apartado anterior del estudio de las publicaciones, se lleva a cabo un análisis del *set* o conjunto subjetivo de AUs atendiendo a diferentes criterios que permiten distintas formas de clasificación de las AUs, y que pueden servir de ayuda para una mejor detección de la emoción y para la realización del análisis final de las AUs (apartado 2.3.4).

Al igual que en el estudio de las publicaciones el primer criterio que se tiene en cuenta son las AUs que funcionan “exclusivas” es decir, que aparecen única y exclusivamente en una emoción como se muestra en la Tabla 19.

Tabla 19 AUs exclusivas, análisis subjetivo

	AUs exclusivas
Anger	22,24
Disgust	9,10,16
Fear	20
Happy	6,12,25
Sadness	15
Surprise	27

El segundo aspecto o criterio a tener en cuenta son las AUs que no deben estar presentes según el análisis visual de las imágenes realizado. Detección por exclusión como se puede ver en la Tabla 20.

Tabla 20 AUs de exclusión o no presentes del análisis subjetivo

	AUs de exclusión o no presentes
Anger	1,5,12
Disgust	12,27
Fear	12
Happy	4,15
Sadness	12
Surprise	4,20,23,24

Partiendo de las AUs de la Tabla 18, se obtienen las AUs que se pueden detectar con los 19 puntos característicos de la cara dando como resultado la Tabla 21 como tercer criterio a tener en cuenta para poder sacar futuras conclusiones. Al igual que en el apartado anterior se elimina el AU 16 (labio inferior hacia abajo), AU 20 (labios estrechados y estirados en horizontal), AU 22 (labios crateriformes) y la AU 25 (labios separados), ya que se consideran difíciles de distinguir por su gran parecido con otras AUs de la boca que aportan una mayor información para la detección.

Tabla 21 Set o conjunto de AUs del análisis subjetivo que se pueden detectar con los 19 puntos de la cara

	AUs para los 19 puntos de la cara
Anger	4,23,24
Disgust	4,9,10,17,23
Fear	1,2,4,5,26
Happy	6,12,26
Sadness	1,4,15,17
Surprise	1,2,5,27

Al igual que en el caso anterior se procede a realizar el análisis por *clusters* para poder apreciar las relaciones entre AUs y emociones, como se observa en la Figura 16. Por ejemplo en este caso se puede ver que las exclusivas para happy son la AU 6 y 12, para surprise la AU 27, en anger la AU 24, disgust la AU 9 y 10, para sadness la AU 15 y para el caso de fear no se tendría ninguna AU única para esta emoción.

AUs parte superior del rostro (cejas/ojos)
 AUs parte inferior del rostro (nariz/boca)

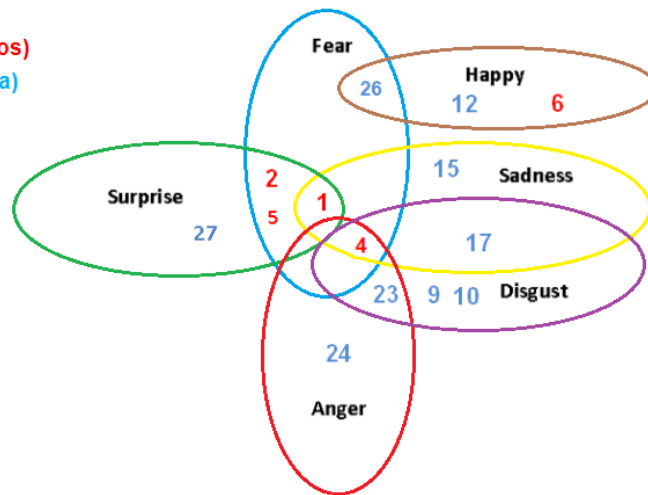


Figura 16 Cluster para el Set o conjunto subjetivo de AUs de la Tabla 21

Como cuarto y último criterio a tener en cuenta, las combinaciones de AUs más frecuentes una vez seleccionadas las AUs más importantes (Tabla 18). Estas combinaciones se obtienen tras un análisis visual de las AUs encontradas en cada imagen ya sea de manera individual o en combinación con otras como se ve en la Tabla 22.

Tabla 22 Posibles combinaciones de AUs para el set o conjunto de AUs subjetivo de la Tabla 18

	Anger	Disgust	Fear	Happy	Sadness	Surprise
Combinaciones de AUs	4+(22 23 24)	9 10 4+9+10+(16 17) 4+9+23 4+10 (9 10)+17	1+2+5+(20 26) (1 4)+(20 26)	12 6+12 6+12+(25 26)	(1 4)+15+17 (1 4)+15	1+2+5+27

Adaptado para las AUs que se pueden detectar con los 19 puntos de la cara, como se hace en la Tabla 14 pero en este caso para el análisis subjetivo, se obtiene la Tabla 23 donde se muestran las posibles combinaciones de AUs para estos 19 puntos disponibles.

Tabla 23 Combinaciones de AUs para el set o conjunto de AUs subjetivo de la Tabla 18 teniendo en cuenta las AUs que se pueden detectar con los 19 puntos de la cara

	Anger	Disgust	Fear	Happy	Sadness	Surprise
Combinaciones de AUs para los 19 puntos de la cara	4+(23 24)	9 10 4+9+10+17 4+9+23 4+10 (9 10)+17	1+2+5+26 (1 4)+26	12 6+12 6+12+26	(1 4)+15	1+2+5+27

2.3.4. Análisis final

Tras llevar a cabo por un lado un estudio de todas las publicaciones mencionadas anteriormente y por otro un estudio personal y subjetivo, en el ámbito de reconocimiento facial a través de Unidades de Acción o AUs, se procede a llegar a una conclusión en cuanto a las AUs características y combinaciones para cada emoción. Determinando si las AUs y combinaciones elegidas son razonables, y si pueden llevar a determinar resultados como se explicará en el apartado 5 de resultados y pruebas.

En primer lugar se procede a realizar una comparativa los dos *sets* o conjuntos de AUs que se han establecido previamente para aquellas AUs posibles de localizar con los 19 puntos de la cara (Tabla 24). Para ello se llama “conjunto 1” al *set* basado en el estudio de publicaciones (Tabla 12) y “conjunto 2” al *set* subjetivo (Tabla 21). Se decide elegir los *sets* o conjuntos que hacen referencia a las AUs que se pueden detectar con los 19 puntos que se localizan actualmente en la cara ya que son los que se podrán implementar posteriormente.

Tabla 24 Comparación de AUs "conjunto 1" de las publicaciones y "conjunto 2" subjetivo, para los 19 puntos

	AUs “conjunto 1” (publicaciones)	AUs “conjunto 2” (subjetivo)
Anger	4,5,7,10,23,24,26	4,23,24
Disgust	9,10,17,26	4,9,10,17,23
Fear	1,2,4,5,26,27	1,2,4,5,26
Happy	6,12	6,12,26
Sadness	1,4,6,15,17	1,4,15,17
Surprise	1,2,5,26,27	1,2,5,27

En segundo lugar, tras el estudio de los análisis realizados (Tabla 24) y viendo la dificultad de distinguir AUs muy parecidas entre sí como por ejemplo la AU 23 (labios tirantes, tensos) y AU 24 (labios presionados) o AU 26 (boca entreabierta) y AU 27 (boca abierta), se decide simplificar la lista de AUs aún más. Para ello se propone agrupar 11 de las AUs seleccionadas formando cinco CAUs (Unidades de Acción Combinadas), tal y como se puede ver en la Tabla 25.

Tabla 25 Selección de AUs y definición de CAUs

	AU	CAU	Descripción de la AU
Parte posterior del rostro	1	1	Interior de las cejas elevado
	2		Exterior de las cejas elevado
	5		Párpado superior elevado
	4		Cejas bajadas
	6		Mejillas elevadas
	7		Párpados tensos
Parte inferior del rostro	9	2	Nariz arrugada
	10		Labio superior elevado
	12		Comisuras de los labios elevados
	15	3	Comisuras de los labios hacia abajo
	17		Barbilla elevada
	23	4	Labios tirantes, tensos
	24		Labios presionados
	26	5	Mandíbula caída (Boca entreabierta)
	27		Boca abierta
H		Boca sonriente abierta	

Estas CAUs se han establecido según las siguientes observaciones:

- La AU 1, AU 2 y AU 5 se observa que generalmente aparecen juntas y suelen estar presentes en las emociones de *fear* y *surprise*, por lo que se decide unir las formando la **CAU 1**.
- Aunque la descripción de la AU 9 se refiere a la nariz y la AU 10 al labio superior elevado se observa que son AUs muy parecidas. Por lo que si se produce los movimientos de los músculos de una AU se producen los de la otra, por lo que se asocian como **CAU 2**.
- La **CAU 3** se forma con la AU 15 y AU 17, presente en la emoción de *sadness*.
- Por último como ya se han comentado los problemas para distinguir ciertas AUs de la boca se decide unir la AU 23 y AU24 en la **CAU4** y la AU 26 y AU 27 en la **CAU5**. Esto se debe a que visualmente son AUs difíciles de distinguir, por lo que más difícil será distinguirlas con el uso de un sistema de reconocimiento.

Por otro lado se crear una nueva AU que se considera importante para poder detectar la emoción de *happy*, ya que no se incluye en el FACS (Tabla 1). Esta AU se nombra como AU-H y representa una boca abierta sonriente, un ejemplo de la misma se puede ver en la cuarta imagen de la derecha de la Figura 15. Es importante no confundir la AU-H con la AU 27 ya que esta se trata de una abertura de la boca sin una sonrisa en ella y generalmente se usa para representar la emoción de *surprise*.

Finalmente con las AUs y CAUs que se van a utilizar e implementar ya definidas, se llega a un conjunto decisivo que es necesario probar para comprobar si con él se puede llevar a cabo el reconocimiento de cada una de las emociones, tal como se muestra en la Tabla 26. Para llegar a este *set* o conjunto decisivo se analiza de una en una todas las AUs y CAUs, teniendo en cuenta los conjuntos de la Tabla 24, y se intenta que se detecten las emociones de forma independiente es decir, que haya el menor número posible de Unidades de Acción que valgan para identificar a varias emociones.

Tabla 26 Set o Conjunto de AUs decisivo

	Set o conjunto de AUs decisivo
Anger	CAU 2, CAU 4, CAU 5, AU 4, AU 7
Disgust	CAU 2, CAU 3, AU 4
Fear	CAU1, CAU 5, AU 4
Happy	AU 6, AU 12, AU-H
Sadness	CAU 3, AU4, AU 6
Surprise	CAU 1, CAU 5

Para poder tener una mejor interpretación y visión de las AUs características de este conjunto decisivo se realiza de nuevo un análisis mediante *clusters*, como se muestra en la Figura 17.

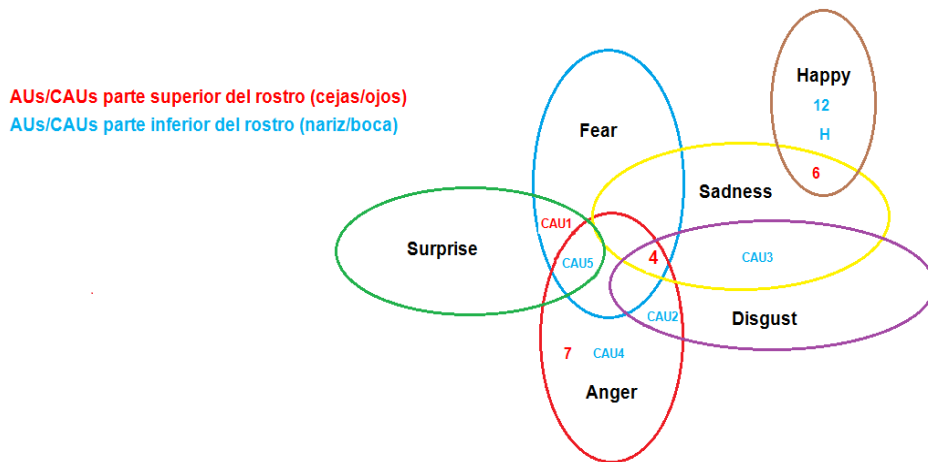


Figura 17 Cluster para el set o Conjunto de AUs decisivo de la Tabla 26

Finalmente una vez que se tiene el conjunto decisivo de la Tabla 26 y el *cluster* de la Figura 17, se establecen las combinaciones de AUs y CAUs para cada emoción (Tabla 27) teniendo en cuenta la Tabla 14 del estudio de las publicaciones y Tabla 23 del estudio subjetivo. Estas tablas no están adaptadas a las nuevas AUs y CAUs establecidas en la Tabla 25, por lo que es necesario adaptar las combinaciones que se proponen tanto en el estudio de las publicaciones como en el estudio subjetivo. Para poder llegar a las combinaciones de este set o conjunto decisivo se van analizando las que intervienen en cada emoción y se van adaptando según las nuevas AUs y CAUs, teniendo en cuenta que por ejemplo en una misma combinación hay AUs o CAUs que no es lógico que puedan darse a la vez. Un ejemplo de ello sería por ejemplo la CAU 5 con la CAU 4 o la CAU1 con la AU4, las cuales no sería lógico que estuvieran en la misma combinación.

Una vez hecho esto, para determinar el conjunto decisivo se juntan las posibles combinaciones que surgen del análisis de las publicaciones y del análisis subjetivo, siendo en muchas ocasiones iguales para ambos sets.

Tabla 27 Combinaciones de AUs para el set o conjunto decisivo

	Anger	Disgust	Fear	Happy	Sadness	Surprise
Combinaciones de AUs para set o conjunto decisivo	AU4+CAU4	AU4+CAU2	AU4+CAU5	AU6+AU12	AU4+CAU3	CAU1+CAU5
	AU4+AU7+CAU2	CAU2	CAU1+CAU5	AU6+AUH	AU6+CAU3	CAU1
	AU4+AU7+CAU4	AU4+CAU2+CAU3	CAU 1	AU12		CAU5
	AU4+AU7+CAU5			AUH		
	AU4+AU7 CAU4					

3. Reconocimiento de las AUs

3.1. Colocación de puntos

En todo sistema de reconocimiento facial de emociones el primer paso a realizar consiste en la detección de la cara y sus diferentes regiones, para posteriormente llegar a la localización de los puntos característicos de cada una de ellas. Es una de las etapas más importantes, ya que si la detección de las regiones y la localización de los puntos falla se pueden ocasionar fallos en el reconocimiento de las emociones. Cuanto más preciso sea el detector de puntos faciales mejor será la calidad de estos sistemas.

Por este motivo se dispone a realizar un análisis comparativo de la colocación de puntos que se lleva a cabo en los métodos propuestos en el PFG de S. González [3] con lo realizado en el método que propone [33] (al cual nos referiremos como método comparativo *Landmarks*), siendo este último un código libre disponible en la comunidad de MathWorks [34] para la detección de puntos de referencia o *landmarks*. En [3] se desarrolla un método de extracción de características basado en regiones fijas y detección de bordes de *Canny* y se inicia el desarrollo de otro método basado en regiones adaptativas (*Thres*), para obtener 19 puntos de interés en la imagen de una cara. A continuación se procede a explicar de forma resumida en qué consisten cada uno de estos métodos, los cuales se quieren comparar con [33], y los pasos que se llevan a cabo hasta que se obtienen cada uno de los puntos característicos.

Cuando se tiene una imagen de entrada en un sistema de reconocimiento el primer paso que se realiza es la localización de la cara o caras presentes en dicha imagen. En el **primer método** utilizado, para encontrar la cara en la imagen se usa el algoritmo de *Viola & Jones*, dicho algoritmo se encuentra disponible en la *Computer Vision Toolbox* de MATLAB. Una vez que se ha hecho esto se divide la cara en seis regiones de una manera fija como se propone en [35] realizando algunos ajustes, es decir cada una de las regiones de cara (ceja izquierda, ceja derecha, boca, nariz, ojo izquierdo y ojo derecho) tienen un tamaño fijo independientemente de la imagen de entrada a analizar.

Previo a la localización de los puntos (12 en la versión inicial de la herramienta y ampliado a 19 puntos en la versión final de la herramienta para poder realizar la normalización y poder detectar correctamente las AUs) se desarrolla un método de segmentación consistente en tres

fases. Inicialmente la imagen se transforma a escala de grises y se aumenta el contraste de la imagen, para que después se detecten los bordes de cada uno de los componentes faciales dentro de las regiones de interés con el uso del algoritmo de detección de bordes *Canny* y en la tercera y última fase se utilizan distintas operaciones morfológicas como pueden ser la dilatación o erosión de la imagen. Estas operaciones son diferentes para cada una de las regiones de la cara o con parámetros diferentes. Sirven para eliminar el ruido presente en las imágenes permitiendo que se mejore la delimitación de las regiones.

Posteriormente a la segmentación de la imagen (separar los objetos de interés de una imagen del resto) se localizan los puntos característicos calculando los extremos de las regiones detectadas. Los puntos que se detectan son: tres para cada una de las cejas, cuatro para la boca, uno para la nariz y cuatro para cada ojo. Para detectar los puntos superiores e inferiores de la boca y de los ojos y el superior de las cejas se busca el punto medio del eje x situado entre los extremos (puntos exteriores de las cejas, ojos y boca) y después se recorre el eje y, hasta encontrar una variación entre dos píxeles consecutivos. Para el caso de la nariz, se hallan los extremos inferiores de la región nariz y se calcula el punto medio. En la Figura 18 se puede ver el proceso seguido en este primer método de extracción de características.

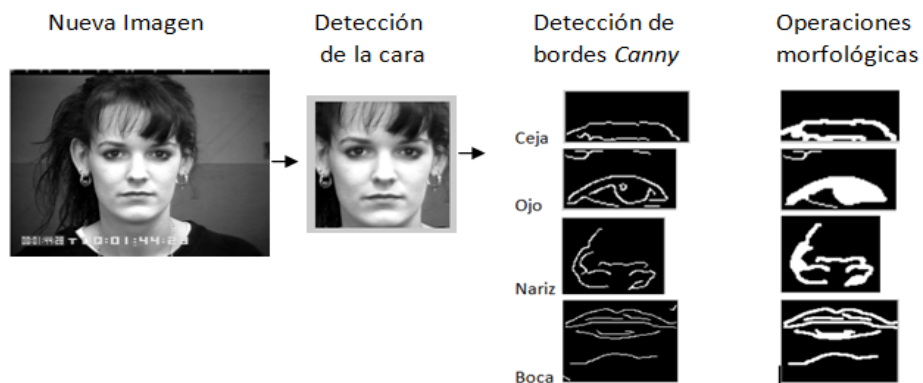


Figura 18 Proceso de extracción de características para el primer método [3]

El **segundo método** que se propone analizar y mejorar (ya que en [3] se realiza una versión inicial a falta de mejoras) para su posterior comparación en la detección de puntos, se lleva a cabo debido a que con el método anterior se observa que la detección de puntos no es correcta. Esto sucede puesto que al tener regiones fijas hay algunas regiones en las que se corta la zona de interés si el movimiento de los ojos, cejas, nariz o boca sobrepasa los límites establecidos para las mismas. Por tanto la detección de bordes *Canny* y la posterior segmentación no se realiza correctamente, llevando en algunos casos a que los puntos coincidan con los bordes de las regiones que se detectan. Por este motivo en lugar de dividir la cara en regiones fijas se establecen nuevas regiones que se adaptan a cada componente de la cara (ojos, cejas, nariz y boca) evitando que puedan ser cortados al delimitar la región de una manera fija. En las mejoras introducidas en este segundo método con respecto a su versión inicial en [3] se precisa y mejora la forma en la que se detectan las regiones a partir de los puntos de los ojos y se mejora la segmentación de algunas de ellas.

La primera región que se localiza es la de los ojos que se obtiene con el algoritmo de *Viola & Jones* implementado en la *Computer Vision Toolbox* de Matlab, al igual que el de localización de la cara. Como este algoritmo comprende la región de ambos ojos, para poder localizar las regiones de cada ojo por separado se divide esta región por la mitad. Después se detectan los puntos de cada ojo y se calculan las regiones de las cejas, boca y nariz en función de la posición de estos puntos. En concreto para la región de las cejas se determina una altura igual que la de la región de los ojos y en el punto superior del ojo se posiciona el borde inferior de la región de las cejas, también se desplaza en horizontal 1/6 hacia la izquierda o derecha según la ceja que se analice con respecto al punto externo del ojo. En el caso de la boca se establecen los

extremos de la región coincidiendo con los puntos extremos de los ojos y el borde inferior con el inferior de la región de la cara y el borde superior con el punto de la nariz. La última región corresponde a la de nariz cuyos extremos coinciden con los puntos medios de los ojos y la altura es el 50% o 70% de la región de la cara. También puede ocurrir que el algoritmo no llegue a detectar la región de los ojos, por lo que si ocurre esto se extrae dicha región de manera fija como se hace en el primer método anteriormente explicado y el resto de zonas se procesan de la misma manera a partir de los puntos de los ojos.

En cuanto a la segmentación que se realiza en cada una de las regiones para el caso de los ojos, la nariz y la boca se mejora el contraste usando ecualización del histograma, es decir se extienden los valores de intensidad, y se convierte la imagen a una imagen binaria. Después se invierten los unos en ceros y viceversa, se dilatan los bordes para extraer el esqueleto, se elimina el ruido y se vuelven a dilatar los bordes de nuevo. Para las cejas se ajusta el contraste es decir, los niveles de intensidad de la imagen, se detectan los bordes usando el detector de bordes de *Canny*, se elimina el ruido, se dilatan los bordes para extraer el esqueleto, se unen los bordes que están ligeramente separados y por último se rellena el objeto. Esta segmentación es la misma que la que se realiza en el primer método. Cuando se tiene todo esto se procede a la detección de los puntos de la misma manera que en el primer método. En la Figura 19 se muestra el proceso seguido en este segundo método de extracción de características.

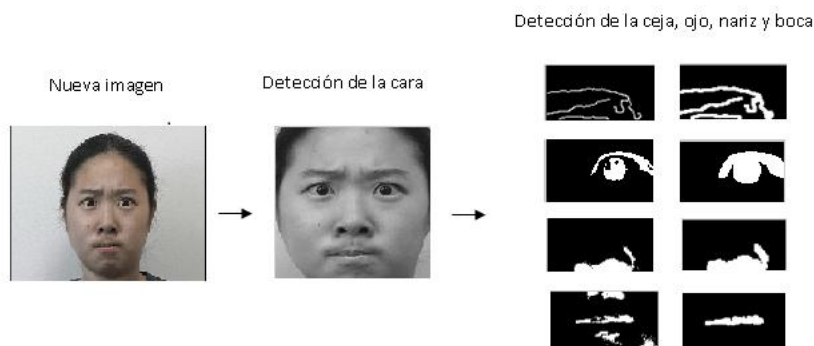


Figura 19 Proceso de extracción de características para el segundo método

La diferencia fundamental a parte de las regiones adaptativas en el segundo método, es la segmentación para la separación de las cejas, ojos, nariz y boca del resto por medio de un umbral de valores de grises en lugar del uso de la detección de bordes *Canny*. Al introducir las regiones adaptativas se evita que algunas de las regiones puedan ser cortadas y que para la posterior localización de los puntos se obtenga una mayor precisión, influyendo a su vez en la detección de las Unidades de Acción. Por tanto a partir de ahora se denominará al primer método como *Canny* y al segundo como *Thres*.

El **tercer y último método** (el método comparativo *Landmarks*), que se pretende analizar para poder compararlo con los dos anteriores, tiene como objetivo la detección de 19 puntos faciales de interés en la cara y para ello se compone de diferentes etapas de procesamiento para los ojos, cejas y boca.

El primer paso se centra en la localización de la imagen de la cara, para ello se usa un detector basado en las características *Haar-like features* (características invariantes a la iluminación y robustas frente al ruido de la imagen) como método de extracción de características y *Adaboost* como clasificador. Este es el algoritmo de *Viola & Jones*, que se usa además para detectar las regiones de los ojos y de la boca. Para el caso de las cejas se estiman a partir de la ubicación de la región de los ojos.

Con la detección de las regiones de interés se procede a realizar la localización de los puntos de referencia o *landmarks*, pero previamente es necesario realizar técnicas de procesamiento de imágenes para mejorar y extraer información relevante en la imagen. Las técnicas usadas en el caso de los ojos son las siguientes: inicialmente se calculan los límites del histograma y a partir de este se realiza un ajuste del contraste, a continuación se convierte a escala de grises y se aplican umbrales concretos para después buscar los píxeles conectados en la imagen para tratar el objeto de mayor área y por último se aplican operaciones morfológicas como la dilatación y el relleno de los espacios vacíos o huecos. Para la región de las cejas (calculada a partir de la del ojo) la imagen se convierte a escala de grises, se ecualiza, se selecciona el mayor área y se dilata y rellenan los huecos. Para finalizar la región de la boca, zona bastante conflictiva por las regiones oscuras o brillantes que pueden aparecer al sonreír por ejemplo, se aplica un filtro *Gaussiano* para suavizar la imagen con el fin de eliminar el ruido. Inicialmente se convierte a escala de grises pero no se obtienen buenos resultados para esta región, por lo que se decide convertir al espacio de color HSV (*Hue, Saturation, Value*) en la que el color se divide en tres componentes. A continuación se aplican las operaciones morfológicas de apertura, erosión y dilatación, se umbraliza, se invierte la imagen umbral y se vuelven aplicar operaciones morfológicas.

Para encontrar los 19 puntos de referencia en las regiones de los ojos, cejas y boca se aplica el detector de bordes *Canny* en ellas y se dividen en partes iguales. Los ojos y la boca se dividen en cuatro partes y las cejas se dividen en dos. Inicialmente se buscan los 10 puntos de referencia que se encuentran situados en los extremos derecho e izquierdo de cada región, los 9 puntos restantes se corresponden con los puntos superiores, inferiores y centrales de la boca y ojos. Los superiores e inferiores se obtienen calculando la Distancia Euclidiana entre los puntos de los extremos derecho e izquierdo para cada una de estas regiones, y los centrales se encuentran calculando en este caso la Distancia Euclidiana de los puntos superiores e inferiores. En la Figura 20 se muestra el proceso seguido por este último método analizado para la localización de los puntos de referencia.

En [33] se utiliza la base de datos MUG [36] para evaluar los resultados. Esta se compone de 86 imágenes en color con rostros en posición frontal con diferentes expresiones faciales. Para su elaboración participaron 35 mujeres y 51 hombres todos de raza blanca y en sus rostros no hay oclusiones.

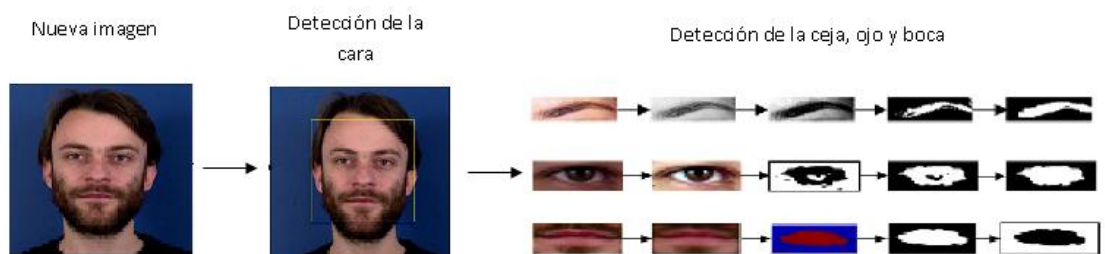


Figura 20 Proceso de extracción de características para el tercer método [33]

Para poder comparar de forma rápida las diferentes formas con las cuales se detectan las regiones y los puntos de referencia para cada uno de los dos métodos desarrollados *Canny* y *Thres*, y el método comparativo *Landmarks* se muestran a continuación la Figura 21 y Figura 22 respectivamente.

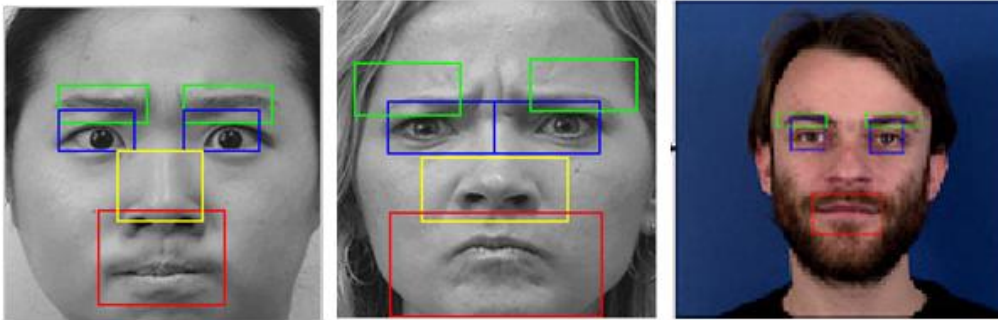


Figura 21 Detección de las regiones para los métodos de Canny, Thres y Landmarks

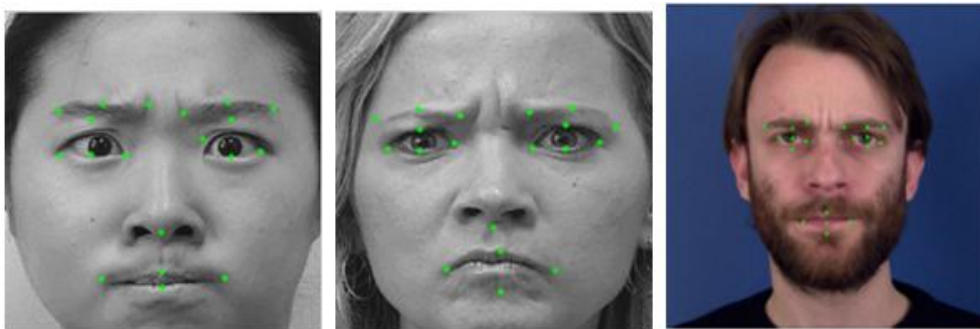


Figura 22 Colocación de puntos para los métodos de Canny, Thres y Landmarks

Una vez que se han explicado los dos métodos anteriores (*Canny*, *Thres*) y el método comparativo *Landmarks* se pretende evaluar el porcentaje de acierto en la colocación de los puntos de referencia para cada uno de los métodos. Con esta evaluación se quiere comprobar cuál de los tres métodos es el mejor en la detección de los puntos, ya que como ya se ha dicho la detección es una etapa muy importante en los sistemas de reconocimiento de emociones.

Para ello se eligen imágenes en color de mujeres y de hombres que se obtienen de las bases de datos *CohnKanade+*, *MUG*, *MMI* e imágenes de internet y se posicionan en ellas los puntos de manera correcta y precisa manualmente. A continuación se compara la posición de los puntos de cada uno de los métodos anteriores con esta localización para evaluar el porcentaje de acierto como se muestra en la Tabla 28.

Tabla 28 Porcentajes de acierto en la detección de puntos para imágenes en color

Detección de Puntos	Canny (%)	Thres (%)	Landmarks (%)	Detección de Puntos	Canny (%)	Thres (%)	Landmarks (%)
Esquina interior de la ceja izquierda	60	90	70	Esquina exterior del ojo izquierdo	80	85	65
Esquina interior de la ceja	75	80	55	Esquina exterior del ojo derecho	85	90	60
Centro de la ceja izquierda	85	95	---	Esquina interior del ojo	60	50	70
Centro de la ceja derecha	95	100	---	Esquina interior del ojo derecho	75	45	60
Esquina exterior de la ceja izquierda	80	90	60	Parte inferior del ojo izquierdo	70	100	70
Esquina exterior de la ceja	85	90	35	Parte inferior del ojo derecho	90	100	65
Esquina izquierda de la	65	70	30	Parte superior del ojo	50	80	65
Esquina derecha de la boca	55	55	40	Parte superior del ojo derecho	65	70	55
Parte inferior de la boca	55	35	75	Punto de la nariz	75	80	---
Parte superior de la boca	70	45	70	Centro de la boca	---	---	70
Centro del ojo derecho	---	---	60	Centro del ojo izquierdo	---	---	75

Las imágenes que se han analizado son en color ya que el tercer método (método comparativo *Landmarks*) sólo procesa imágenes en color por lo que las imágenes en grises de la base de datos *CohnKanade+* no son válidas. El porcentaje de imágenes que se elige para el análisis es el mismo independientemente de cuál sea su procedencia, la base de datos *CohnKanade+*, *MUG* etc. Esto se determina así para que ninguno de los tres métodos evaluados se vea beneficiado por el uso de una base de datos u otra, ya que por ejemplo el método de *Landmarks* está ideado para la base de datos *MUG* y un mayor uso de imágenes de esta base de datos podría hacer que sus porcentajes mejorasen frente al resto.

En la Tabla 29 también se puede ver como se detectan los puntos únicamente con imágenes en grises de la base de datos *CohnKanade+* para el caso del primer y segundo método. El número de imágenes elegido es el mismo que en el caso anterior de la Tabla 28.

Tabla 29 Porcentajes de acierto en la detección de puntos para imágenes en grises [3]

Detección de Puntos	Canny (%)	Thres (%)	Detección de Puntos	Canny (%)	Thres (%)
Esquina interior de la ceja izquierda	25	90	Esquina exterior del ojo izquierdo	50	75
Esquina interior de la ceja derecha	55	95	Esquina exterior del ojo derecho	40	75
Centro de la ceja izquierda	75	80	Esquina interior del ojo izquierdo	55	55
Centro de la ceja derecha	85	95	Esquina interior del ojo derecho	40	60
Esquina exterior de la ceja izquierda	70	85	Parte inferior del ojo izquierdo	50	90
Esquina exterior de la ceja derecha	60	70	Parte inferior del ojo derecho	50	95
Esquina izquierda de la boca	55	80	Parte superior del ojo izquierdo	35	85
Esquina derecha de la boca	55	55	Parte superior del ojo derecho	40	95
Parte inferior de la boca	40	30	Punto de la nariz	85	85
Parte superior de la boca	60	60			

Como se puede observar en la Figura 22 los tres métodos detectan 19 puntos de interés en la cara pero el método comparativo *Landmarks* no detecta el punto de la nariz a diferencia de los otros dos métodos. De la misma manera en los dos primeros métodos (*Canny* y *Thres*) no se detecta el punto central de la boca y de los ojos. Para el caso de *Canny* y *Thres* el punto de la nariz es necesario e indispensable para realizar la normalización de las imágenes y los puntos centrales de los ojos y de la boca se considera que no dan información relevante para la detección de las AUs y posteriormente de las emociones.

Tras este análisis se puede determinar y concluir cual de los tres métodos se considera más adecuado en la colocación de puntos. En el caso del método de *Thres* se ha observado tanto para los resultados obtenidos en la Tabla 28 como para los de la Tabla 29 que los puntos de la región de la boca no se detectan correctamente, esto se debe a que la boca no se segmenta correctamente debido a la segmentación mediante valores de grises que se emplea. En esta zona de la boca se suelen producir confusiones debido a las arrugas y sombras que aparecen por ejemplo en caras sonrientes con la emoción *happy*, y esto crea problemas a la hora de llevar a cabo la segmentación. En cambio para el resto de regiones detectadas (nariz, cejas y ojos) los porcentajes de aciertos son mucho mejores que en los otros dos métodos tanto para el análisis de la Tabla 28 como para el de la Tabla 29, rondando entre el 70 - 100 % de precisión.

En cuanto a la detección usando el método de *Canny* a diferencia del método de *Thres* y el comparativo *Landmarks* la localización de los puntos característicos de la boca es mejor, teniendo unos porcentajes de acierto entre el 55 – 70 % para el análisis de la Tabla 28 con imágenes en color y entre el 40 – 60 % para las imágenes en grises de la Tabla 29. Esto lleva a pensar que la detección de la región de la boca usando un tamaño fijo es mejor que la utilizada en *Thres* a partir de regiones adaptativas según la situación de los extremos de los ojos o que

la usada en *Landmarks* con el uso del algoritmo de *Viola & Jones*, así como la segmentación que se lleva a cabo.

Finalmente con el método comparativo *Landmarks* se obtienen unos porcentajes de acierto que rondan entre el 30 y 70 %, siendo estos resultados peores con respecto a cualquiera de los otros dos métodos que son analizados y usados en la herramienta desarrollada en [3]. Por otro lado se ha observado que el método comparativo *Landmarks* requiere un mayor tiempo a la hora de encontrar los puntos que los otros dos métodos analizados. El uso de un correcto número de puntos y de la elección de los mismos es fundamental para que el tiempo de localización de los mismos se reduzca. También hay que destacar como ya se ha comentado anteriormente que este método desarrollado por [33] está pensado para el uso de la base de datos *MUG*, con lo cual a la hora de realizar el análisis de la detección de puntos se observa que la localización de los mismos no es adecuada para cualquiera de las imágenes que no sean de la base de datos *MUG*.

Como conclusión y partiendo de la base de que ningún método es totalmente perfecto y preciso, se puede determinar tras lo mencionado anteriormente que el método más adecuado para la localización de los puntos de interés en el rostro es el método *Thres*. Tanto para el análisis que se realiza con imágenes en color en la Tabla 28 como para imágenes en grises en la Tabla 29, como se puede ver de una manera más visual en la Figura 23 y Figura 24 respectivamente. En ambas se pueden observar las mejoras que se obtienen con respecto a los otros métodos, salvo en el caso de la boca como ya se ha comentado. Por otro lado se tiene que tener en cuenta que estos resultados pueden variar según el número y las imágenes que se elijan para realizar el análisis.

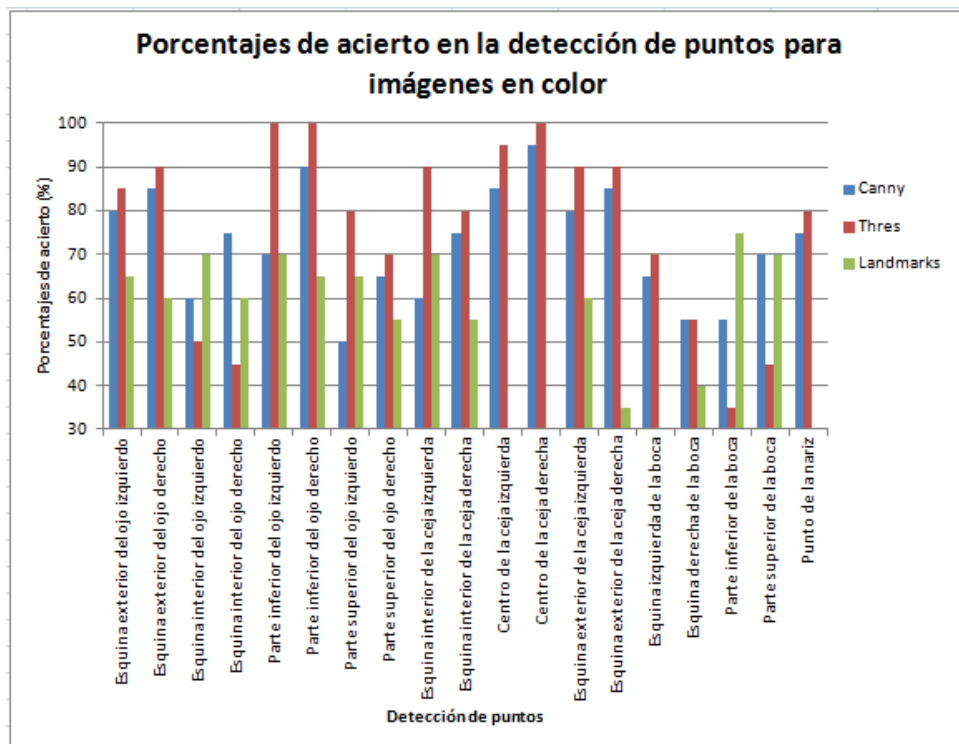


Figura 23 Gráfica comparativa de la detección de puntos imágenes en color

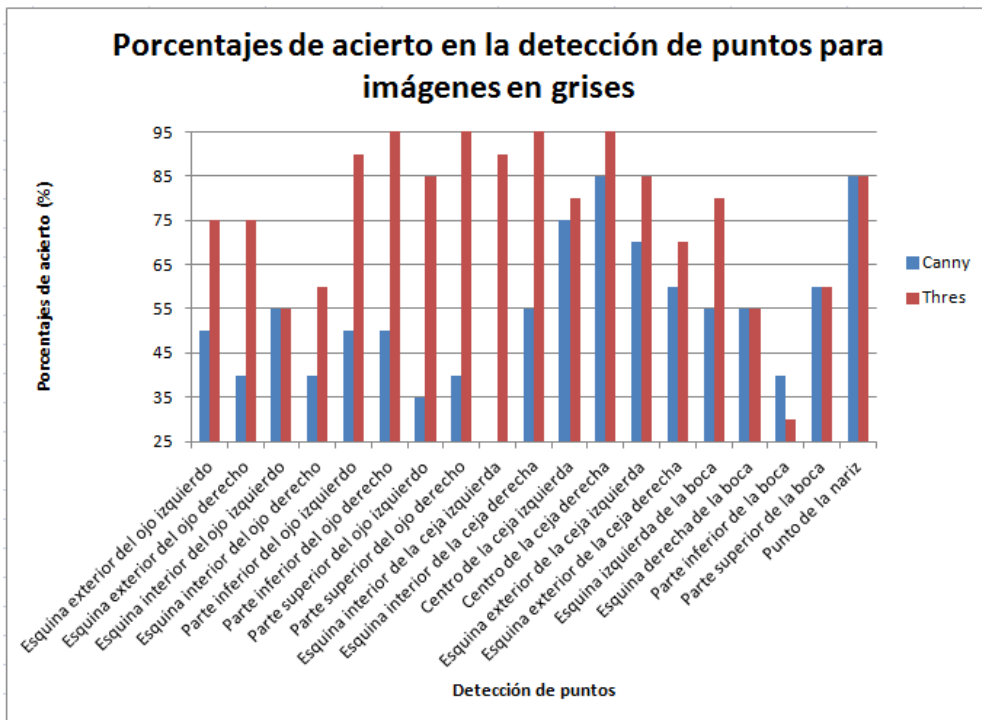


Figura 24 Gráfica comparativa de la detección de puntos imágenes en grises

3.2. Detección de AUs y CAUs con 19 puntos

Una vez que se han identificado los puntos característicos del rostro para cada una de las regiones detectadas, se determinan los grupos de músculos que se consideran importantes para poder obtener información de la emoción que se pretende detectar ya que en ellos es donde se encuentran las AUs.

Una AU aislada se puede considerar como un mero movimiento de un músculo pero una combinación de AUs puede llevar al reconocimiento de una emoción en el rostro. En primer lugar se detectan las AUs que están presentes en la cara y que se han seleccionado como características de cada emoción (Tabla 26) y después con el uso de un método de clasificación se procesan y se clasifican dentro de una emoción.

En este punto es donde se encuentran los problemas, ya que los clasificadores de emoción deben determinar a partir de qué momento se considera que una AU está presente o no. Por ejemplo ¿cuál es el grado de curvatura que tienen que tener los labios para determinar si el rostro tiene o no una sonrisa?, o ¿cuánto tienen que descender las cejas para determinar que están bajadas? Por este motivo, y al no poderlo establecer de una manera precisa, se recurre a realizar planteamientos para establecer los criterios con los cuales se pueda determinar con una mayor certeza cuando se detecta o está activa una AU.

La detección de las AUs o CAUs se lleva a cabo con la comparación entre la imagen neutral y la imagen que posee la emoción que se pretende detectar. Una AU o CAU se construye a través de un grupo de puntos que se pueden caracterizar por medio de cuatro posibles descriptores como son: áreas cerradas, ángulos, distancias entre puntos y vectores de movimiento.

La detección de las AUs y CAUs se podría llevar a cabo con el uso de algunos de estos descriptores. Por ejemplo con el uso de las **áreas cerradas** se podría calcular el área que encierra un conjunto de puntos, pudiendo formar distintas figuras geométricas (rombos, triángulos etc.), que se compara con el área de la imagen neutral para esos mismos puntos. Para el caso de los **ángulos** consiste en calcular el ángulo que forma la unión de tres puntos,

estableciendo si este es mayor o menor que el de la neutral. Por último para el caso de las **distancias entre puntos** y los **vectores de movimiento**, consiste en determinar la variación de la distancia entre dos puntos o el movimiento de uno de ellos con respecto de la imagen neutral.

Para este proyecto se decide implementar la detección de las AUs o CAUs en las imágenes con la emoción por medio de las áreas cerradas que crean los grupos de puntos y con las distancias y el movimiento de dichos puntos, con respecto a la imagen neutral. Se eligen estos descriptores ya que se cree que son los más adecuados para los 19 puntos faciales de los que se dispone y dan información suficiente para llevar a cabo el reconocimiento de la emoción, sin necesidad de crear un sistema de detección largo y complejo que ralentice la herramienta.

Las áreas cerradas que se detectan son triángulos sobre la parte de la cara que está en acción (ojo, ceja o boca). Estos triángulos se pueden ver en la Figura 25 y se nombran de la siguiente manera:

- **LU** (*left-up*), **RU** (*right-up*): triángulo superior del ojo izquierdo y derecho, formados por los puntos (7, 8, 9) y (11, 12, 13).
- **LD** (*left-down*), **RD** (*right-down*): triángulo inferior del ojo izquierdo y derecho, formados por los puntos (7, 9, 10) y (11, 13, 14).
- **LI** (*left-inner*), **RI** (*right-inner*): triángulo interior del ojo izquierdo y derecho, puntos (3, 8, 9) y (4, 11, 12) respectivamente.
- **UL** (*up-left*), **UR** (*up-right*): triángulo superior de la boca para el lado izquierdo y derecho de la misma, formado por los puntos (15, 16, 19) y (16, 17, 19).

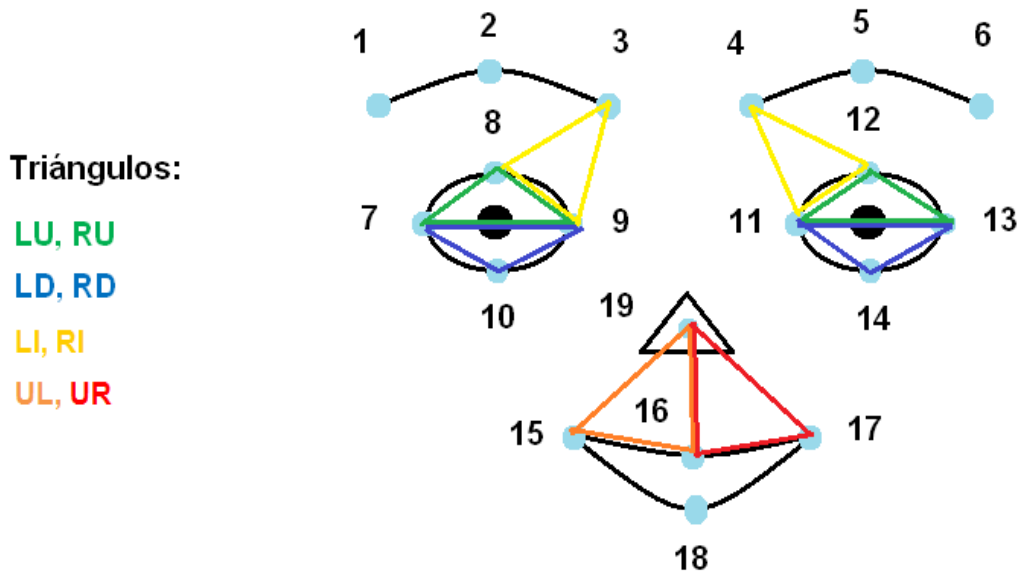


Figura 25 Triángulos utilizados para detección de las AUs y CAUs.

En la Figura 25 también se puede ver cuáles son los puntos característicos que se localizan, así como la numeración que tienen para poder determinar las distancias entre los mismos y sus movimientos. A continuación en la Tabla 30 se muestra cuales son los triángulos y los puntos de la Figura 25 que intervienen en cada AU y CAU para su detección.

Tabla 30 Triángulos y puntos implicados en la detección de cada AU y CAU

AUs y CAUs	Triángulos implicados	Puntos implicados
CAU 1	----	2, 5, 10, 14
CAU 2	UL, UR	15, 17, 19
CAU 3	UL, UR	15, 16, 17, 18, 19
CAU 4	----	15, 16, 17, 18, 19
CAU 5	UL, UR	16, 17, 18, 19
AU 4	LI, RI	3, 4, 9, 11
AU 6	LU, RU, LD, RD	8, 10, 12, 14
AU 7	LU, RU, LD, RD	8, 10, 12, 14
AU 12	----	15, 16, 17, 18, 19
AU-H	----	15, 16, 17, 18, 19

En la Tabla 30 se puede observar que en muchos casos hay varias AUs y CAUs que comparten los mismos triángulos y puntos para su detección, pero las reglas que se establecen para cada una son diferentes como se explicará posteriormente.

Para que se pueda entender mejor, en la Tabla 31 se muestran los dibujos correspondientes a cada AU y CAU con los triángulos y puntos implicados tanto para la imagen neutral como para la imagen con la emoción. Se ha tratado de simular los movimientos de los músculos que están asociados con sus respectivas AUs y CAUs.

Tabla 31 Dibujos para cada AU y CAU

	Imagen neutral	Imagen con emoción
CAU 1		
CAU 2		
CAU 3		
CAU 4		

	Imagen neutral	Imagen con emoción
CAU 5		
AU 4		
AU 6		
AU 7		
AU 12		

	Imagen neutral	Imagen con emoción
AU-H		

Tras establecer cómo se lleva a cabo la detección de las AUs y CAUs y que parámetros son los que se utilizan para ello, se explica cómo se realiza la implementación de esta detección en la herramienta de detección de emociones. En [3] la detección de las Unidades de Acción se realiza por medio de matrices de diferencias entre los 19 puntos extraídos en la imagen neutral y en la imagen analizada, para determinar si una AU está activa o no. De esta forma se evalúan las variaciones de los puntos implicados en cada AU en dichas matrices es decir, el movimiento de los puntos implicados. Como ya se ha comentado los clasificadores tienen que determinar cuándo se puede considerar que una AU esté activa o no, antes de poder realizar el reconocimiento de la emoción.

Es por ello por lo que se procede a explicar cuáles son las reglas de detección que se han establecido tras el estudio realizado para que el clasificador, en este caso el Árbol de Decisión, determine la emoción que está presente en la imagen analizada. Estas reglas hacen referencia tanto a las áreas de los triángulos establecidos como a las distancias entre puntos y movimiento de los mismos como se puede ver en la Tabla 31, siempre con referencia a la imagen neutral.

- **CAU 1:** esta CAU conlleva que el interior y exterior de las cejas y el parpado superior se eleven con respecto de la imagen neutral. Para ello se establece que para que esta CAU se active la distancia entre los puntos 2 y 10 y 5 y 14 debe ser mayor que en la imagen neutral.
- **CAU 2:** donde el área de los triángulos UL y UR es menor que en la imagen neutral y el labio superior se eleva es decir, la diferencia entre los puntos 15 y 19 y 17 y 19 disminuye.
- **CAU 3:** en esta CAU el área de los triángulos UL y UR aumenta y las comisuras de los labios (puntos 15 y 17) descienden.
- **CAU 4:** en ella los labios se aprietan y tensan, por lo que los puntos 16 y 18 se acercan y las comisuras (puntos 15 y 17) se estiran hacia los lados.
- **CAU 5:** las áreas de los triángulos UL y UR aumentan, así como la distancia entre los puntos 16 y 18 es dos veces mayor que en la neutral.
- **AU 4:** con esta AU las cejas se bajan y por tanto los triángulos LI y RI disminuyen, o los puntos 3 y 4 se acercan o la diferencia entre los puntos 3 y 9 y 4 y 11 es más pequeña que en la neutral.

- AU 6: para que esta AU se active es necesario que al menos tres de los triángulos LU, RU, LD, y RD disminuyan, que la distancia entre los puntos 8 y 10 y 12 y 14 sea menor que en la neutral.
- AU 7: para la AU 7, se cumplen los mismos criterios que la AU 6 además de que los puntos 10 y 14 suban con respecto a la imagen neutral.
- AU 12: conlleva que las comisuras (puntos 15 y 17) se eleven y se desplacen en hacia los lados, expresando una sonrisa con la boca cerrada.
- AU-H: al igual que la AU 12 las comisuras se elevan y se desplazan en hacia los lados, además de que la diferencia entre los puntos 16 y 18 sea mayor que en la neutral expresando una sonrisa con la boca abierta.

Cabe destacar que todo el proceso de detección e implementación de AUs y CAUs se ha realizado en cooperación con el compañero Diego Zapatero Olmedillo, cuyo PFG [2] se basa en el reconocimiento facial de emociones para Android.

3.3. Árbol de Decisión

Una vez que se han establecido las reglas de detección es el Árbol de Decisión el que se encarga de reconocer las emociones, previamente entrenado. Un Árbol de Decisión es un modelo de aprendizaje con el que se construye un diagrama de decisiones secuenciales, para ello es necesario una serie de entradas o atributos a partir de las cuales y junto con un algoritmo de decisión devuelve una salida.

De igual modo que un árbol, el Árbol de Decisión está formado por una raíz, ramas y hojas. En la Figura 26 se puede observar cómo se estructura.

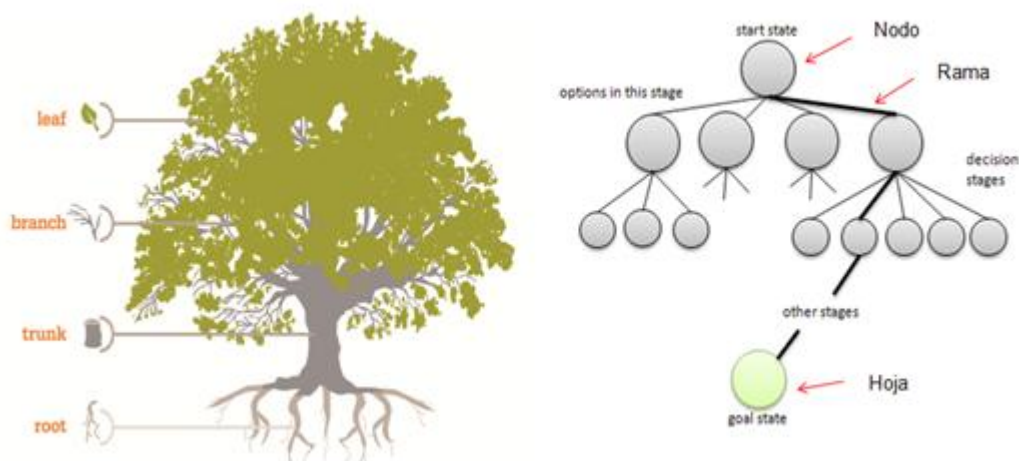


Figura 26 Estructura de un Árbol de Decisión

Para poder explicar la estructura se parte de la raíz del árbol, esta constituye el nodo principal y más importante, para llegar a las ramas principales que conducen a los nodos secundarios y así sucesivamente para concluir en las hojas que se corresponden con los posibles resultados o salidas. Esta estructura se crea con el uso del algoritmo ID3 [37], utilizado en el ámbito de la inteligencia artificial por su sencillez frente a otros.

El algoritmo ID3 permite crear un Árbol de Decisión con una estructura descendente, con él se identifica cuál es el atributo o entrada de mayor importancia e influencia que se situará más próximo a la raíz del árbol. Este atributo es aquel que tiene el mayor poder discriminatorio

para poder así ir creando subconjuntos a partir del mismo. Para determinar la influencia de cada atributo se lleva a cabo un análisis estadístico como es el cálculo de la Entropía (E) o de la Ganancia de información, siendo este último parámetro más preciso que el anterior para identificar cual es el mejor atributo.

En resumen los pasos que sigue el algoritmo para crear el Árbol de decisión son los siguientes: en primer lugar teniendo en cuenta el conjunto de ejemplos o características se selecciona el mejor atributo (nodo principal), en segundo este atributo divide el conjunto de ejemplos en subconjuntos de ejemplos, en tercer lugar se vuelve a repetir los pasos anteriores a partir de cada subconjunto pero con un atributo menos (el seleccionado en el primer paso) y con menos ejemplos, este proceso se repite sucesivamente hasta que en cuarto y último lugar todos los ejemplos de un subconjunto tengan el mismo valor (la salida de la rama u hoja tomaría ese valor) y se hayan analizado todos los atributos (en este caso la hoja se determina en función del mayor número de valores iguales dentro del subconjunto). Cabe destacar que los atributos que se seleccionan como mejores atributos de cada subconjunto se descartan para las subdivisiones inferiores.

Tras la implementación de este algoritmo con la ayuda de la matriz y por tanto la creación del Árbol de Decisión, se va realizando el entrenamiento previo a la detección de la emoción por parte de este clasificador. Para realizar este entrenamiento se requiere una gran cantidad de imágenes de las que se conoce la emoción presente en ellas y así el árbol pueda ir aprendiendo las características que pertenecen a cada una de las seis emociones básicas que se pretenden detectar. Estas características son las AUs y CAUs que se han establecido en la Tabla 25. El entrenamiento se lleva a cabo con el uso de una matriz de entrada que contienen las posibles combinaciones de AUs y CAUs para cada emoción (Tabla 27), y también es necesario un vector que asocie la emoción real con cada una de las combinaciones de la matriz. El ejemplo de la matriz de entrenamiento se puede ver en la Tabla 32, en ella se muestran algunas de las posibles combinaciones de AUs y CAUs de la Tabla 27 para cada emoción. Este entrenamiento corresponde al modulo de aprendizaje de la Figura 1.

Tabla 32 Ejemplo sencillo de matriz de combinaciones

Combinaciones	CAUs/AUs									Emoción real
	CAU 1	CAU2	CAU 3	CAU 4	CAU 5	AU 4	AU 6	AU 7	
	0	0	0	1	0	1	0	1		Anger
	0	1	0	0	0	1	0	0		Disgust
	0	0	0	0	1	1	0	0		Fear
	0	0	0	0	0	0	1	0		Happy
	0	0	1	0	0	1	0	0		Sadness
	1	0	0	0	1	0	0	0		Surprise

Por tanto cuando una nueva muestra o imagen se tenga que analizar el Árbol de Decisión analizará sus entradas o atributos (en este caso las AUs o CAUs). Según el orden en el que se haya creado dicho árbol en la fase de entrenamiento, y en función de los valores de entrada de la imagen que se quiere analizar (las AUs o CAUs que se activan en dicha imagen) seguirá un camino u otro por el árbol hasta llegar a la salida u hoja del árbol concreta. La detección de la emoción por parte del clasificador se corresponderá con el camino que contenga más AUs o CAUs activos. En la Figura 27 se puede ver un ejemplo de cómo quedaría la estructura de un

ejemplo de Árbol de Decisión, los números entre paréntesis indican si la AU o CAU está activa o no.

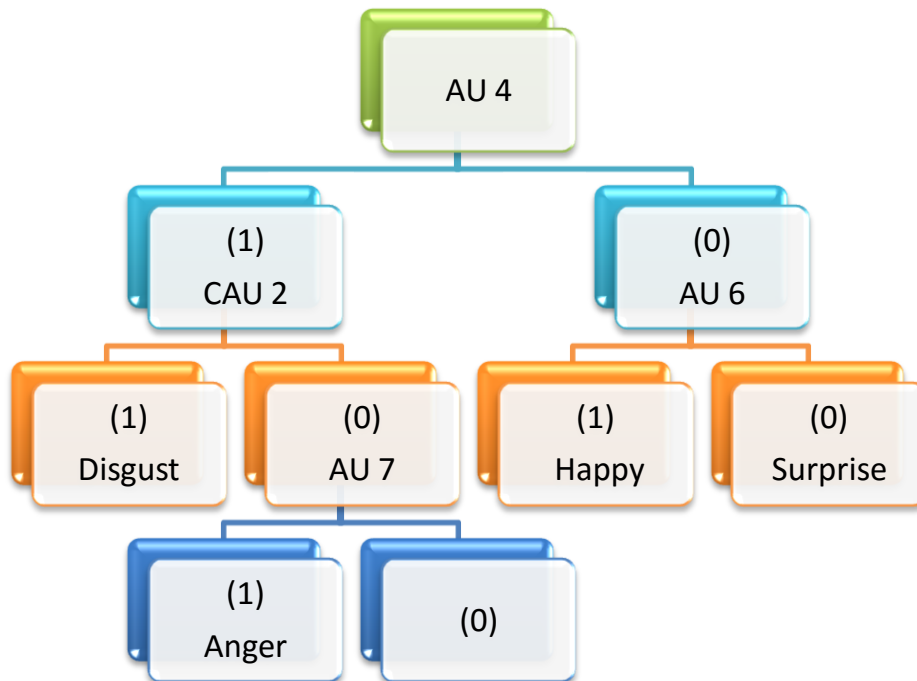


Figura 27 Ejemplo de estructura simple de Árbol de Decisión

En el PFG de Diego Zapatero [2] se puede encontrar más información acerca del Árbol de Decisión, explicando en mayor detalle cómo se crea y cómo es su funcionamiento interno, ya que este proyecto no se centra en los métodos de clasificación de la emoción sino en la parte de extracción de características.

4. Herramienta de pruebas

4.1. Antecedentes

Para el desarrollo de este proyecto se describe y amplía un herramienta que permite la implementación de diferentes métodos de extracción de características y de clasificación, usados para el reconocimiento facial de emociones. Esta herramienta se maneja a través de una interfaz gráfica en Matlab, disponible con la función GUIDE de Matlab que permite crear una ventana en la cual se pueden seleccionar los diferentes métodos y controles desarrollados y visualizar los resultados obtenidos.

La versión inicial de la herramienta se creó durante el curso escolar 2013/2014 en el centro CITSEM [38] por alumnos en prácticas y posteriormente se amplió en el PFG de Sara González Martín [3] durante el 2014/2015. En esta última versión, de la cual se parte para el desarrollo de este proyecto, se realizan distintos cambios que se comentan en el siguiente apartado a partir de la Figura 28 donde se puede ver la interfaz gráfica de la herramienta.

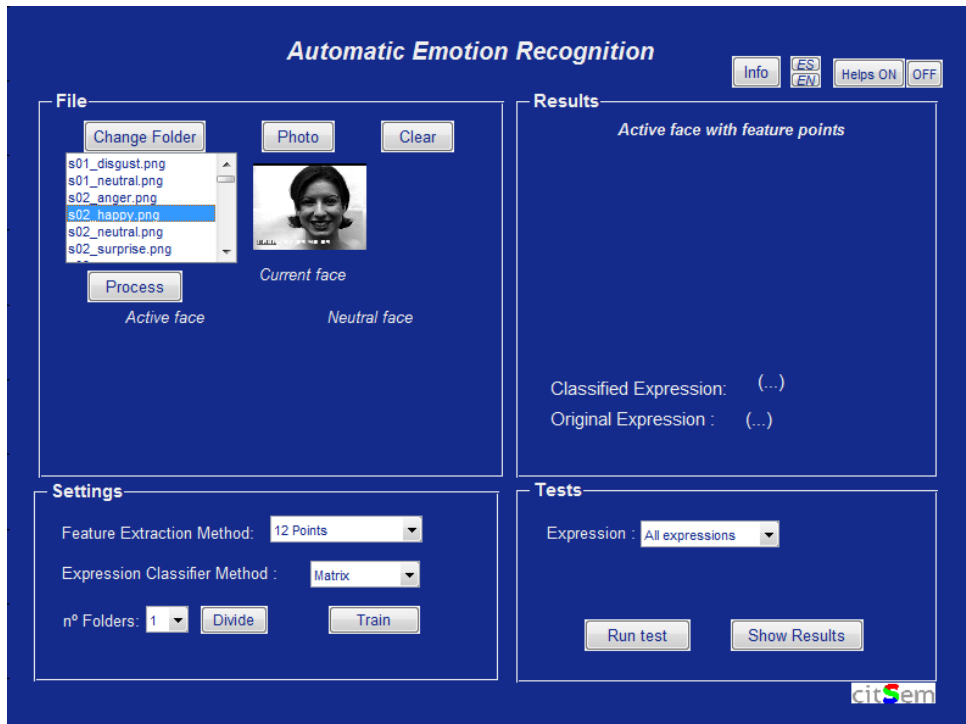


Figura 28 Interfaz gráfica de la herramienta del PFG Sara González

La interfaz se encuentra dividida en cuatro ventanas para representar las distintas etapas de un sistema de reconocimiento facial de emociones, permitiendo un manejo rápido e intuitivo. Se procede a explicar cada una de las ventanas:

- **FILE** (archivo): ventana donde se obtiene la imagen que se desea analizar, ya sea una base de datos o una fotografía por medio de la webcam del ordenador (botón *Photo*). Si se desea se puede cambiar la base de datos que se quiere analizar con el botón *Change folder* que permite cambiar el directorio en el que se encuentra dicha base de datos. En *Current face* se previsualiza la imagen de la base de datos que se ha seleccionado en la lista de la izquierda de la misma. Para ejecutar la herramienta se pulsa el botón *Process* (procesar) de la interfaz o el botón *Enter* del teclado y posteriormente se visualiza la imagen neutral (*Neutral face*) y la imagen con la emoción (*Active face*). Además se pueden eliminar todos los datos e imágenes que se han seleccionado para llevar a cabo el reconocimiento de una emoción presente en una imagen por medio del botón *Clear*, y así la interfaz vuelve a su estado original (Figura 28).
- **SETTINGS** (parámetros): aquí se seleccionan los métodos de extracción de características y de clasificación de la expresión, para poder determinar la emoción que el sujeto expresa. Los métodos de extracción de características de los que se dispone en esta versión de la herramienta son los siguientes: *12 Points*, *19 Points Canny*, *19 Points Thres*, *12 Points+normalize*, *19 Points+normalize_Canny* y *19 Points+normalize_Thres*. Estos métodos hacen referencia a los puntos característicos de la cara que se extraen (12 o 19 puntos), a la segmentación que se utiliza para la detección de los puntos (*Canny* o *Thres*) comentada en el apartado anterior y a si las imágenes analizadas se han normalizado o no. En cuanto a los métodos de clasificación se tienen *Matrix* y *AUs*. Tras obtener los puntos de interés es necesario crear un algoritmo que permita clasificar las emociones a partir de la información que se obtiene de ellos. Por lo que se desarrolla un método llamado *Matrix* (solo disponible para 12 puntos de la cara) que consiste en la comparación de matrices con las posiciones de los puntos característicos entre la imagen neutral y la que contiene

la expresión analizada, y otro método llamado *AUs* (disponible para 19 puntos) que consiste en reconocer una serie de Unidades de Acción e identificarlas para clasificar la emoción. Por otro lado la herramienta permite realizar un entrenamiento previo que sirve de aprendizaje al sistema de reconocimiento, cuando un método es seleccionado aparece el botón *Train* que permite realizar esta acción. También se permite dividir la base de datos que se ha seleccionado en la ventana FILE en diversos grupos o *folders* para que unos sirvan de entrenamiento al sistema (90 % de las imágenes) y otros de prueba (10 % restante), pulsando el botón *Divide*.

- *RESULTS* (resultados): ventana con la cual se ven los resultados de la extracción de características con la representación de los puntos faciales en la imagen que posee la emoción y los resultados de la clasificación indicando si se ha reconocido correctamente la emoción. En *Original Expression* se indica la emoción real que contiene la imagen y en *Classified Expression* la emoción que detecta el sistema, de esta manera se puede comprobar si se ha detectado la emoción adecuada.
- *TEST* (pruebas): donde se obtienen los porcentajes de acierto y fallo de todas las imágenes de la base de datos para cada una de las emociones básicas, pulsando el botón *Run Test*. Posteriormente se almacenan en un documento Excel, el cual se puede ver pulsando el botón *Show Results*. Se incluye también la opción de evaluar si se desea cinco de las seis emociones básicas en el menú desplegable de *Expression* en lugar de todas las emociones y en el desplegable de *Excluded expression* se tendrá que indicar la emoción de la cual no se quieren hacer las pruebas.

En la parte superior derecha de la interfaz se pueden observar distintos botones que facilitan el manejo de la interfaz a cualquier usuario. En primer lugar se encuentra el botón de *Info* que proporciona información de la herramienta de reconocimiento facial, como por ejemplo los autores y fechas de actualización. En segundo lugar los botones de *ES* y *EN* que permiten elegir el idioma de la interfaz, ya sea en Español o en Inglés y en tercer y último lugar los botones de *Helps ON* y *OFF* que posibilitan la opción de visualizar o no la ayuda que se proporciona al lado de cada botón en las diferentes ventanas (?). En la Figura 29 se puede ver cómo queda la interfaz tras ser ejecutada.

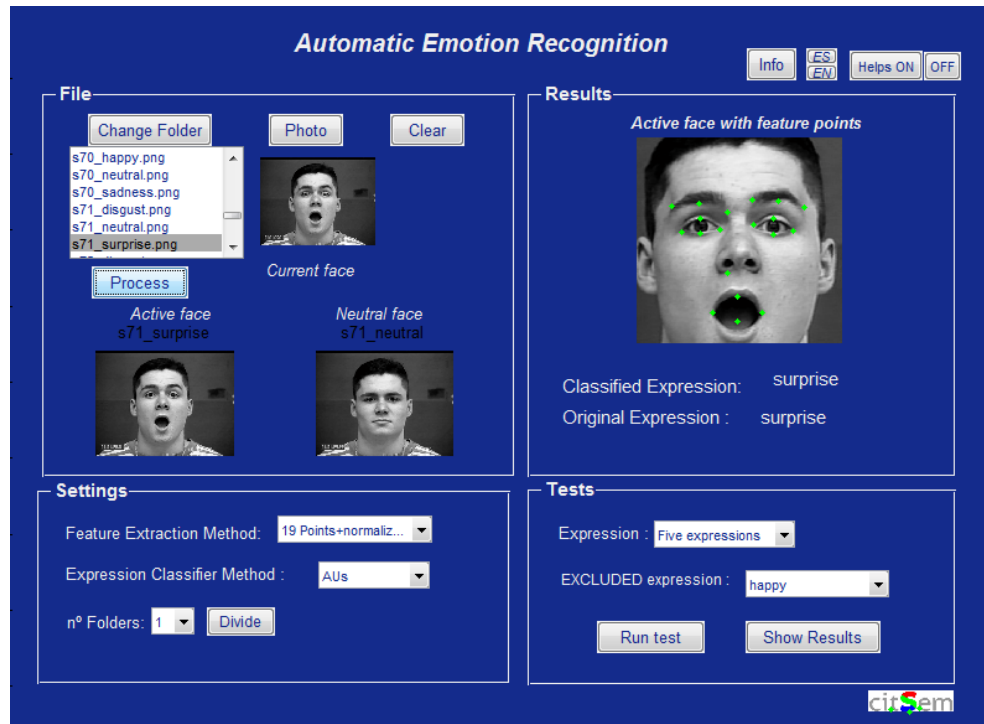


Figura 29 Interfaz gráfica de la herramienta del PFG Sara González, tras su ejecución

En el proyecto de Sara González [3] del cual se parte para el presente proyecto, se propone como trabajo futuro implementar nuevos métodos de extracción de características y de clasificación de la emoción como puede ser el uso de un Árbol de Decisión basado en las Unidades de Acción. Por otro lado se plantea la posibilidad de la mejora de la precisión en la localización de los puntos, ya que como ya se ha comentado en apartado anteriores es una etapa fundamental de todo sistema de reconocimiento facial.

Con motivo de poder llevar a cabo algunas de las propuestas mencionadas en [3] se decide intentar mejorar la localización de los puntos al introducir algunas mejoras y cambios en el método de extracción de características *Thres* (como se explica en el apartado 3.1). En cuanto al método de clasificación el estudio se centra en la implementación de las AUs y el Árbol de Decisión para poder clasificar las emociones en base a las mismas.

4.2. Ampliación de la herramienta

Para poder realizar los avances que se han comentado se necesita adaptar y ampliar la herramienta de la que se dispone, los objetivos de mejora que se llevan a cabo en este proyecto son los siguientes:

- Cambio del aspecto de la interfaz: se ha modificado tanto el color como la distribución de algunos de los controles.
- Se renombra la ventana de *SETTINGS* a *SETTINGS & TRAINING* (parámetros y entrenamiento) y en ella se realizan los siguientes cambios:

- Se eliminan las opciones de *12 Points*, *19 Points Canny* y *19 Points Thres* del desplegable del método de extracción de características y se han renombrado las opciones de *19 Points+normalize_Canny* y *19 Points+normalize_Thres* por *19 Points_Canny* y *19 Pointst_Thres* respectivamente. Ya que para el reconocimiento de las AUs los 12 puntos no son válidos y se requiere a su vez la normalización de las imágenes.
- Se suprime el método de clasificación *Matrix* puesto que el objetivo de este proyecto se centra en las Unidades de Acción. Se añade el método de *DecisionTree* que por defecto utiliza las AUs.
- Se incluyen en el código las funciones relativas al método de *DecisionTree* (Árbol de Decisión) y se lleva a cabo la implementación de cómo detectar las diferentes Unidades de Acción en las imágenes, siguiendo las reglas establecidas en los apartados anteriores.
- Se añade el control de *Training Data* (datos de entrenamiento) con su correspondiente desplegable para poder seleccionar la matriz de AUs con la que se quiere entrenar al Árbol de Decisión es decir, al método de clasificación de la emoción.
- Se renombra el botón *Train* a *Train System* y se adapta el código de la función *Train* para que en este caso se entrene el *DecisionTree* (Árbol de Decisión), ya que en la versión anterior de la herramienta este botón servía para crear las matrices representativas de cada una de las emociones. Así para clasificar la emoción de una imagen de entrada en la herramienta se calculaba su matriz de puntos y se comparaba con las matrices representativas (también de puntos) de las seis emociones. Ahora al pulsar este botón el árbol se entrena con la ayuda de la matriz que se ha seleccionado en el control de *Training Data*.

➤ En *TEST* (pruebas):

- Se elimina el desplegable de *Expression* y de *Excluded expression*, estos controles se utilizaban principalmente para el método de clasificación *Matrix* y en el presente proyecto se requiere siempre la obtención de las pruebas para todas las emociones.
- Se mueve el *nº Folders* y el botón de *Divide* a esta ventana, como ya se ha comentado estos controles sirven para poder hacer pruebas con distintos grupos de imágenes de la base de datos. Para el caso del Árbol de Decisión la división de la base de datos en diferentes grupos de imágenes, unos para entrenamiento y otros para pruebas, no es necesaria ya que el árbol utiliza una la matriz para entrenar al árbol. Aun así se decide mantener en la herramienta estos controles para futuras implementaciones de nuevos métodos que puedan necesitarlo. El botón *Run Test* y *Show Results* se mantienen tras una previa adaptación del código de sus funciones. Estos botones inicialmente no se muestran visibles hasta que no se ha entrenado por lo menos una vez al sistema, así posteriormente se realizarán las pruebas con las imágenes de la base de datos seleccionada y las exportará a un Excel con los resultados de las mismas.

- En *RESULTS* (resultados) se añaden:
 - Los botones *ON-Points* y *OFF-Points* que permiten activar o desactivar la visualización de los puntos característicos de la cara, por defecto se muestran desactivados.
 - Los botones *ON-Regions* y *OFF-Regions* con los que se pueden ver las regiones (cajas) que se detectan en la cara y que contienen a las cejas, ojos, nariz y boca.
- Actualización de todos los botones de ayuda, información e idiomas (*Info*, *Helps ON*, *OFF*, *ES*, *EN*) añadiendo y quitando la información que no se considera de interés para este proyecto.
- Se realiza una limpieza del código para que sea más eficiente y rápido y se han eliminado los elementos y funciones innecesarios. También se ha comentado todo el código que se ha añadido para que sea más fácil la futura ampliación de la herramienta.

Para que los cambios que se han realizado queden lo más claros posibles se procede a explicar paso por paso como se realiza el manejo de la herramienta de reconocimiento facial de emociones.

1. Para poder ejecutar la herramienta correctamente es necesario tener instalado el programa Matlab versión 2012b junto con la *Computer Vision Toolbox* de dicho programa. La función principal con la que se puede comenzar a usar la herramienta es la denominada *Interface_Faces.m*, en ella se debe ejecutar el botón *Run* de Matlab. A su vez esta función tiene asociada su correspondiente imagen *Interface_Faces.fig* que permite el acceso a la interfaz gráfica.
2. Una vez que se ejecuta la herramienta el aspecto inicial de la misma se muestra en la Figura 30 y es entonces donde se elige la base de datos *Change Folder* que contiene las imágenes que se quieren analizar o en su defecto se puede realizar el análisis de una imagen tomada con la webcam botón *Photo*.

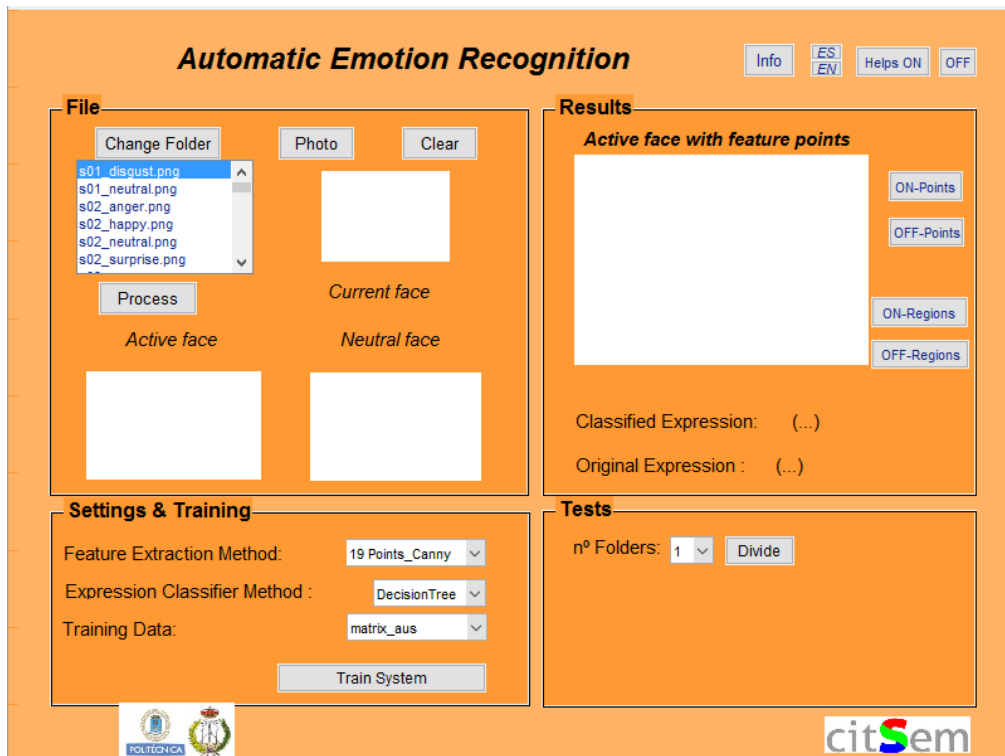


Figura 30 Interfaz gráfica de la herramienta ampliada tras ser ejecutada

3. En la ventana de *SETTINGS* a *SETTINGS & TRAINING* se elige: el método de extracción de características que se desee (*19 Points_Canny* o *19 Pointst_Thres*), el método de clasificación en este caso *DecisionTree* y la matriz con la que se quiere realizar el entrenamiento.
4. Una vez hecho esto se pulsa el botón *Train System* para poder entrenar al árbol. Como ya se ha dicho es necesario entrenar al Árbol de Decisión al menos una vez para que pueda realizar posteriormente el reconocimiento de la emoción. A medida que se va realizando el entrenamiento se va creando el árbol con el que posteriormente se reconocerá la emoción.
5. Con el árbol entrenado ya se pueden hacer las pruebas que se estimen oportunas. Los botones *Run Test* y *Show Results* ya se ven en la interfaz y por tanto se pueden realizar pruebas de la base de datos completa pulsando en *Run Test* y ver sus resultados pulsando *Show Results*, o bien se pueden seleccionar una a una las imágenes de la lista situada debajo del botón *Change Folder*. Si se selecciona una imagen de esta lista se debe pulsar en el botón *Process* o bien en el *Enter* del teclado, así se verá en la parte de *RESULTS* la expresión que se ha detectado junto con la original que posee la imagen. En la *Figura 31* se puede ver como quedaría la interfaz tras el reconocimiento de una emoción.

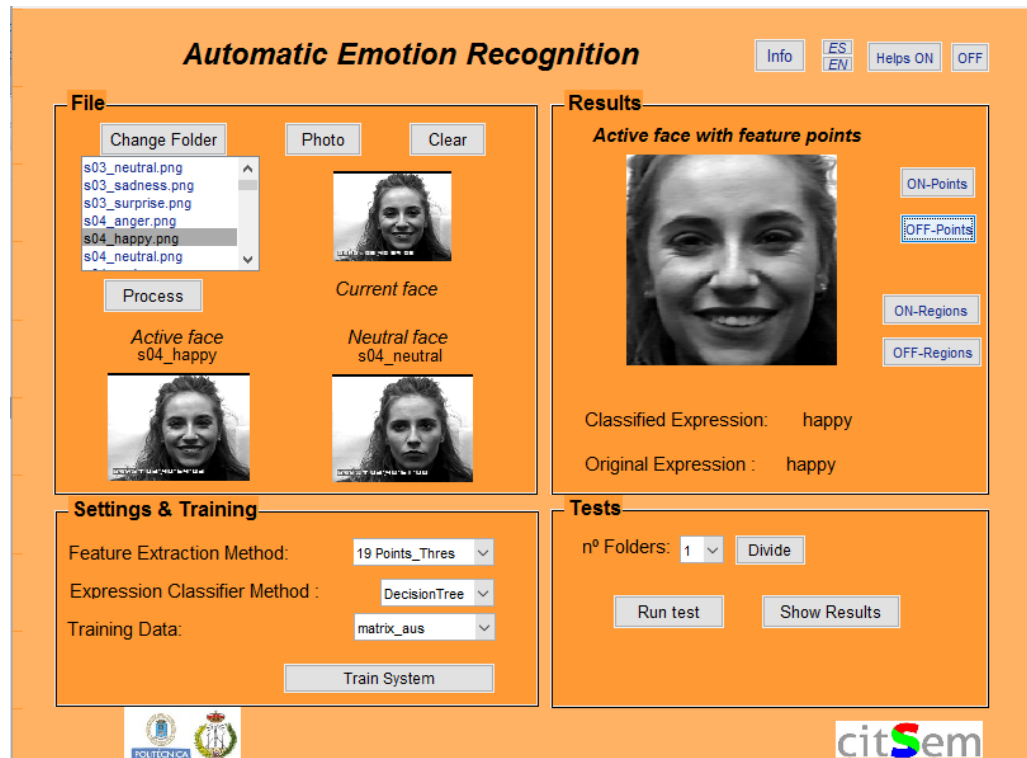


Figura 31 Interfaz gráfica de la herramienta ampliada tras el reconocimiento de la emoción

6. Para terminar en la Figura 32 se pueden ver distintas imágenes del aspecto de la interfaz tras activar algunos controles que posee la herramienta.

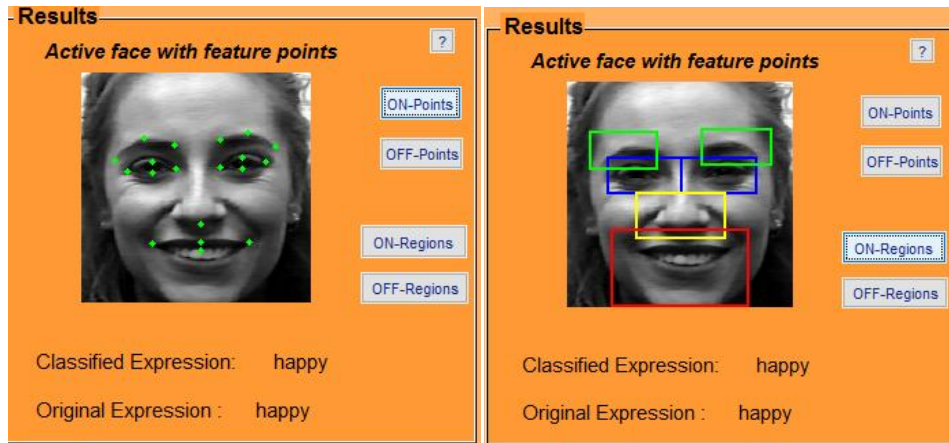
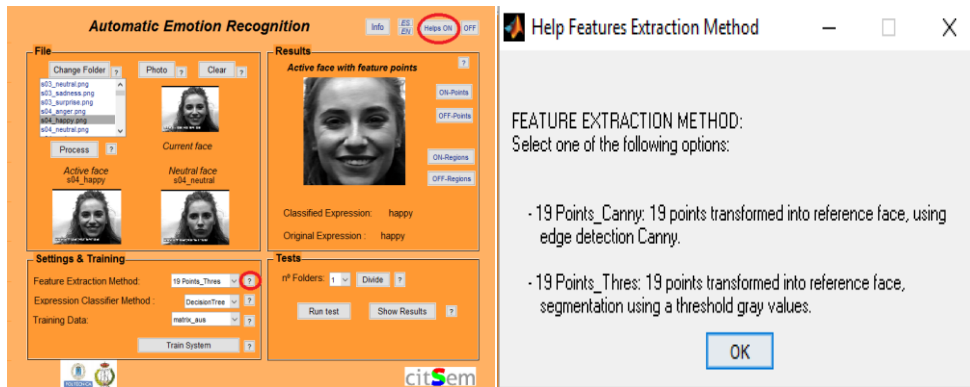


Figura 32 Interfaz gráfica de la herramienta ampliada tras pulsar distintos botones, de arriba abajo de izquierda a derecha: tras pulsar el botón Helps On, al pulsar el botón (?) con su correspondiente mensaje de ayuda, al pulsar ON-Points, al seleccionar ON-Regions y por último con la interfaz en español tras pulsar el botón ES.

5. Pruebas y resultados

Para poder probar el funcionamiento de la ampliación de la herramienta que se ha llevado a cabo y el método implementado, se realizan pruebas con las bases de datos *CohnKanade+* [12] y *JAFFE* [31]. Estas pruebas se realizan de forma automática con el uso de los botones *Run Test* y *Show Results* de la interfaz gráfica, que permiten seleccionar las imágenes de la base de datos con la expresión que se quiere analizar y hacer los cálculos oportunos para guardarlos posteriormente en una tabla Excel con los resultados de acierto y fallo. También se puede ver como se ha clasificado la emoción presente en cada imagen eligiéndolas de una en una y pulsando en *Process* o *Enter*.

Las pruebas que se van a realizar son para las seis emociones básicas (*anger, disgust, fear, happiness, surprise* y *sadness*), ya que son las emociones más analizadas por la mayor parte de los trabajos de reconocimiento facial existentes hasta el momento y para las que se ha creado el método del Árbol de Decisión.

Los resultados se evalúan para las dos formas de segmentación explicadas en el apartado 3.1 y que se pueden seleccionar en el desplegable del Método de extracción de características (*Feature Extraction Method*) de la herramienta, estos métodos son el *19 Points_Canny* y *19 Points_Thres*. La forma en la que se presentan los resultados que se van a obtener consiste en mostrar tablas de confusión, donde se puede ver el porcentaje de acierto y fallo para cada una de las emociones analizadas. Para el caso de fallo se puede conocer con qué emoción o emociones se confunde.

5.1. Base de datos CohnKanade+

La *CohnKanade+* ha sido la base de datos más usada por muchos investigadores en el área del reconocimiento automático de emociones, contiene 123 sujetos entre hombres y mujeres de diferentes nacionalidades. Para cada uno de ellos se tiene la imagen con expresión y la neutral, las expresiones son las seis básicas más otra con la expresión denominada *contempt* (desprecio)

Es importante comentar que la base de datos *CohnKanade+* (*CK*) no posee el mismo número de imágenes para cada emoción como se puede observar en la Tabla 33 y esto influye de manera considerable en los resultados que se puedan obtener.

Tabla 33 Número de imágenes de la *CohnKanade+*

Emoción	Nº de imágenes en la CK+
Anger	45
Disgust	58
Fear	24
Happy	68
Sadness	27
Surprise	81
Contempt	6
Neutral	105
Total de imágenes	414

Antes de probar el Árbol de Decisión se realizan una serie de pruebas con las AUs y CAUs características de cada emoción (Tabla 26) para comprobar que las reglas de detección que se han establecido son correctas, y permiten detectar si una AU o CAU está activa o no. Así se pueden realizar los ajustes oportunos en el algoritmo de detección implementado, para concretar lo máximo posible las reglas de detección e intentar evitar la confusión entre las diferentes AUs y CAUs. Estas pruebas se realizan con la base de datos *CohnKanade+* y con el método de *19 Points_Thres*, ya que este método proporciona una mejor localización de los puntos como se vio en el apartado 3.1. En la Tabla 34 se muestra el porcentaje de AUs y CAUs detectadas en 24 imágenes elegidas aleatoriamente para cada emoción.

Tabla 34 Porcentajes de AUs y CAUs detectadas en cada emoción

CAUs/AUs	Anger	Disgust	Fear	Happy	Sadness	Surprise
CAU 1	4%	4%	30%	29%	37%	92%
CAU 2	46%	34%	12%	17%	29%	8%
CAU 3	12%	9%	8%	0%	29%	0%
CAU 4	71%	58%	8%	20%	12%	0%
CAU 5	17%	0%	59%	4%	29%	88%
AU 4	55%	67%	34%	20%	29%	8%
AU 6	40%	33%	37%	63%	33%	29%
AU 7	38%	50%	0%	8%	17%	0%
AU 12	8%	0%	4%	59%	4%	0%
AU-H	0%	0%	8%	13%	0%	0%

Como se puede observar existe confusión entre algunas de las AUs y CAUs detectadas, provocando unos resultados bajos en la detección de las AUs o CAUs que son características para cada emoción (sombreadas en naranja). Hay que tener en cuenta que en muchas de las 24 imágenes analizadas se observan AUs o CAUs que no son frecuentes en la emoción que se analiza, pero que visualizando la imagen se puede ver que esa AU o CAU puede estar activa y por tanto ser detectada por el algoritmo implementado para la detección de las mismas. También hay que destacar que la detección de los puntos de la boca no es correcta en varias de las imágenes analizadas, por lo que las AUs y CAUs de esta región se detectan erróneamente al no estar bien situados los puntos. Así pues para comprobar que el algoritmo de decisión implementado es correcto, a pesar de los resultados de la Tabla 34, se tiene que analizar cada una de las imágenes de manera individual fijándose en si las AUs o CAUs que se detectan en cada una de ellas se corresponden con los movimientos de los músculos que se ven en la imagen. Tras este análisis se realizan algunos ajustes del algoritmo de detección, y se pone a prueba el Árbol de Decisión para comprobar si es capaz de clasificar las emociones.

5.1.1. Método Thres

En primer lugar se realizan pruebas con el método de *19 Points_Thres*, haciendo uso de la matriz de entrenamiento que contiene las combinaciones que se establecen en la Tabla 27, en concreto la matriz posee 21 filas de combinaciones. Con esta matriz se entrena el Árbol de Decisión, previamente al reconocimiento de la emoción de las imágenes de esta base de datos. En la Figura 33 se puede ver el Árbol de Decisión que se obtiene tras el entrenamiento y en la Tabla 35 los resultados de la matriz de confusión una vez se detecta la emoción en todas las imágenes de la CK+.

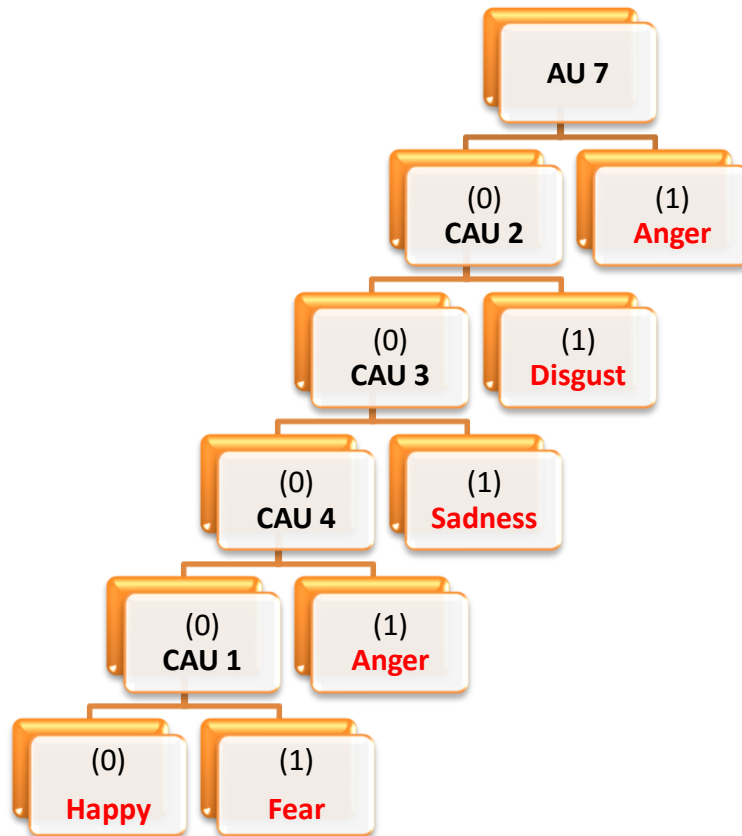


Figura 33 Árbol de Decisión con las combinaciones de la Tabla 27 para la CK+

Tabla 35 Matriz de confusión para las combinaciones de Tabla 27

Tested expression-->	Anger	Disgust	Fear	Happines	Sadness	Surprise
Detected expression						
Anger	67	68	8	24	26	6
Disgust	20	23	13	26	22	7
Fear	0	0	17	11	11	75
Happines	4	7	54	39	19	7
Sadness	9	2	8	0	22	4
Surprise	0	0	0	0	0	0

Como se puede observar los resultados de aciertos que se obtienen (marcados en negrita) son muy bajos excepto en *anger* que se consigue un porcentaje de acierto del 67%. Pero peor aún es el caso de *surprise* donde no se detecta esta emoción en ninguna de las imágenes analizadas. Esto se debe principalmente a dos motivos: el primero es que en el Árbol de

Decisión que se genera tras el entrenamiento (Figura 33) esta emoción no aparece en ninguna de las hojas del árbol es decir, como una posible solución por lo que nunca va a poder llegar a reconocer esta emoción en las imágenes que analiza. El otro motivo es que *fear* y *surprise* tienen algunas combinaciones en común (Tabla 27) y este es un hecho que no puede suceder porque puede llevar a que ambas emociones se confundan y no se reconozcan, como es el caso de *surprise* que se confunde con *fear* en un 75% de las veces. El que se haya llegado a las mismas combinaciones para ambas emociones se debe a que ya inicialmente algunas de las AUs que se detectan en estas emociones (Tabla 24) son las mismas es decir, los movimientos de los músculos que se producen en la cara con estas emociones son en muchas ocasiones los mismos. Partiendo de esta base y teniendo en cuenta que se ha reducido el número de AUs, formando CAUs, lleva a que *fear* y *surprise* tengan combinaciones en común con las AUs y CAUs que se han establecido inicialmente para la detección de las emociones.

Al ver estos resultados de acierto tan bajos se decide realizar una serie de cambios en cuanto a las AUs y CAUs que se detectan para llevar a cabo el reconocimiento y a las posibles combinaciones presentes en la matriz de entrenamiento.

Un factor importante a tener en cuenta para el reconocimiento de las emociones por parte del árbol es el entrenamiento del mismo con una matriz adecuada, es decir una matriz que aporte la información suficiente para que la estructura del árbol que se cree permita un buen reconocimiento.

Como ya se dijo en el apartado 3.3 el árbol utiliza un algoritmo para poder crear la estructura del árbol y una serie de fórmulas de probabilidad para determinar que atributos son los más importantes, todo ello con la matriz que se le proporciona. Así pues el tamaño de la matriz influirá a la hora de crear el árbol y de determinar las AUs o CAUs (atributos) más importantes en cada rama.

Dicho esto en primer lugar se decide eliminar la CAU5, que inicialmente se estableció como una combinación de las AUs 26 y 27 debido a que son AUs difíciles de distinguir, para implementar por separado dichas AUs. Con esta nueva implementación se pretende poder diferenciar *fear* y *surprise*, ya que se observa que en la mayoría de las imágenes la AU que predomina para *surprise* en cuanto a la región de la boca es la 27 y para *fear* la AU26.

En segundo lugar se decide modificar la matriz de combinaciones introduciendo las nuevas AUs implementadas así como nuevas posibles combinaciones para *disgust* y para *happy* tal como se muestra en la Tabla 36. Estas nuevas combinaciones se han introducido porque para el caso de *disgust* se había olvidado y porque para el caso de *happy* al analizar de nuevo las AUs y CAUs de las que se dispone, se determina que la AU 4 conlleva la bajada de las cejas y el acercamiento de los puntos interiores de las cejas, siendo este un movimiento que puede ocurrir al mostrar una cara *happy* en un rostro. También se decide ampliar la matriz de combinaciones aumentando el número de filas para la emoción de *fear* y de *surprise* es decir, duplicando las filas de las combinaciones de estas emociones para que el árbol pueda darles más importancia y pueda reconocerlas sin necesidad de confundirlas. Todo esto hará que el árbol que se obtenga sea mejor y que tenga como posible solución (hoja) la emoción de *surprise*, evitando que se produzcan resultados de acierto del 0% en alguna de las emociones.

Tabla 36 Combinaciones de AUs para el set o conjunto decisivo, tras realizar ajustes

	Anger	Disgust	Fear	Happy	Sadness	Surprise
Combinaciones de AUs para set o conjunto decisivo	AU4+CAU4	AU4+CAU2	AU4+AU27	AU6+AU12	AU4+CAU3	CAU1+AU27
	AU4+AU7+CAU2	CAU2	CAU1+AU26	AU6+AUH	AU6+CAU3	CAU1
	AU4+AU7+CAU4	AU4+CAU2+CAU3	AU26	AU4+AU6+AU12		AU27
	AU4+AU7+AU27	CAU2+CAU3		AU4+AU6+AUH		
	AU4+AU7			AU12		
	CAU4			AUH		

Por último tras aplicar estos cambios se obtiene la nueva estructura del árbol (Figura 34) y la matriz de confusión (Tabla 37) con los porcentajes de acierto y fallo para cada una de las emociones de las que se dispone.

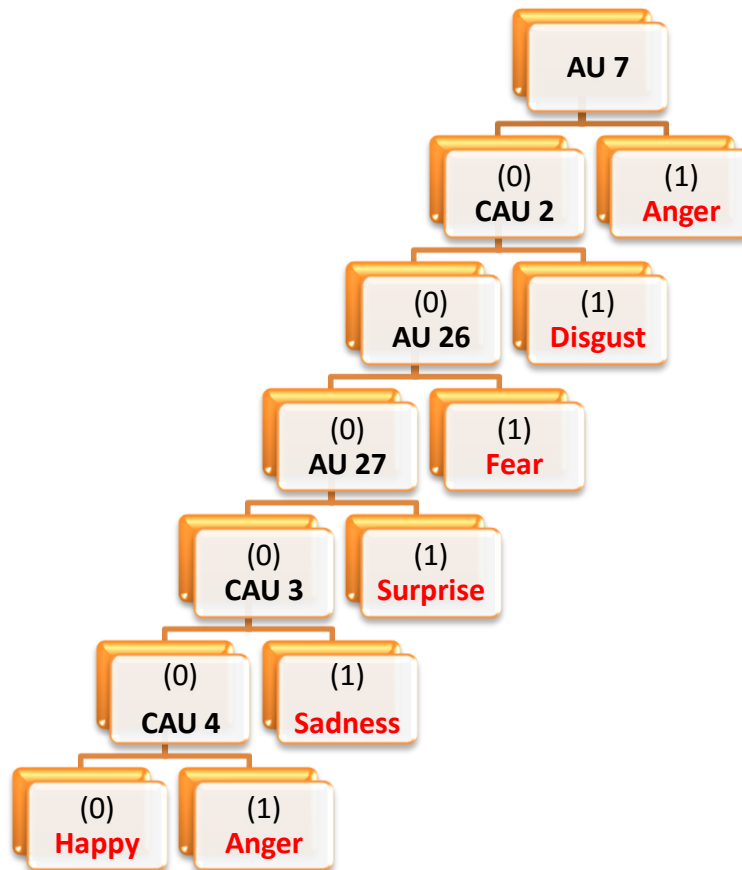


Figura 34 Árbol de Decisión con las combinaciones de la Tabla 36 para la CK+

Tabla 37 Matriz de confusión CK+ para las combinaciones de Tabla 36, Thres

Tested expression-->	Anger	Disgust	Fear	Happiness	Sadness	Surprise
Detected expression						
Anger	67	66	8	23	26	6
Disgust	20	23	13	26	22	7
Fear	2	5	42	3	7	19
Happiness	2	0	13	43	4	0
Sadness	9	2	8	0	22	4
Surprise	0	4	17	6	19	64

Como se puede observar al introducir estos cambios la estructura del árbol que se obtiene tiene un mayor número de atributos (AUS/CAUs) y por tanto ramas, lo que hace que ya aparezca la emoción de *surprise* como una posible solución u hoja del árbol. Por otro lado cabe destacar que en estos árboles no aparecen todos los atributos (las 11 AUs/CAUs de las que se dispone), ya que el árbol considera que no son importantes con motivo de los cálculos probabilísticos que realiza.

En cuanto a los resultados de acierto que se obtienen con la matriz de confusión (Tabla 37) se ha conseguido que el porcentaje de acierto para *anger* y *surprise* supere el 60% y que para *fear* y *happiness* supere el 40%, manteniéndose igual para *disgust* y *sadness*. También hay que comentar que el grado de confusión de *surprise* con *fear* ha pasado de un 75% en la Tabla 35 a un 19 % en la Tabla 37. Pero la confusión de *fear* con *surprise* se ha aumentado ya que antes no se tenía la emoción de *surprise* como posible solución del árbol (Figura 33) y por tanto no se podía producir ninguna confusión. Para el caso de *disgust* solo se ha conseguido mejorar su confusión con *anger* en un 2% (de un 68% a un 66%) provocando que los resultados de acierto para *disgust* sean muy bajos. Para *sadness* se ha reducido el porcentaje de confusión con *fear* (de 11% a un 7% de confusión) y *happiness* (de un 19% a un 4%) pero se sigue confundiendo en la misma medida con *anger* y *disgust*.

5.1.2. Método Canny

Del mismo modo se quiere comprobar los resultados que se obtienen con el uso del método de *Canny* (19 *Points_Canny*) el cual ofrece otra forma de segmentación y por tanto de precisión en la localización de puntos de la cara.

Para ello al igual que en el apartado anterior se utiliza la matriz de combinaciones que se crea a partir de las combinaciones de la Tabla 36, y con ella se genera la matriz de confusión de la Tabla 38 . En este caso no se muestra el Árbol de Decisión ya que es el mismo que el de la Figura 34 porque la matriz de entrenamiento sigue siendo la misma, lo único que cambia es la precisión en la localización de los puntos y por tanto esto influirá a la hora de detectar las AUs o CAUs que están presentes en cada imagen.

Tabla 38 Matriz de confusión CK+ para las combinaciones de Tabla 36, Canny

Tested expression-->	Anger	Disgust	Fear	Happiness	Sadness	Surprise
Detected expression						
Anger	60	52	33	36	37	12
Disgust	4	14	17	17	11	9
Fear	11	18	25	24	15	21
Happiness	11	2	8	16	0	0
Sadness	7	5	0	0	30	15
Surprise	7	9	17	7	7	43

En este caso los resultados que se muestran en la Tabla 38 con el uso del método *Canny* indican unos porcentajes de acierto en el reconocimiento menores que los que se obtienen con el método *Thres* (Tabla 37). La única emoción en la que se consigue mejorar el resultado de acierto es en *sadness* donde se obtiene un porcentaje del 30% con respecto al 22% para el método *Thres*, siendo un porcentaje poco significativo.

Por tanto como ya se dijo en el apartado 3.1, donde se analizaban los diferentes métodos de segmentación, la colocación de puntos que se obtiene con *Canny* es peor que la que se obtiene con *Thres* y esto influye de manera muy considerable a la hora de detectar las AUs y CAUs que están presentes en cada una de las imágenes. Una mala colocación de puntos puede hacer que se detecten AUs que no están presentes en la imagen, como se puede ver por ejemplo en la Figura 35.



Figura 35 Imágenes de la CK+ para el método de *Canny*, neutral (izquierda) y emoción (derecha)

En este caso los puntos que se localizan en la imagen neutral para la región de la boca no están bien situados por las sombras que se producen debajo del labio inferior, y en la imagen con la emoción de *surprise* el punto inferior del labio se sitúa en el labio superior ya que la región que se ha delimitado corta la boca. Así pues cuando se analiza esta imagen para detectar que AUs o CAUs están presentes el algoritmo de detección determina que la CAU que está presente en la imagen con la emoción de la Figura 35 es la CAU3 en lugar de la AU 27 que debería detectar, al tener las comisuras de los labios hacia abajo.

5.2. Base de datos JAFFE

Otra de las bases de datos más utilizadas por los investigadores es la *JAFFE*, la cual consta de 10 sujetos femeninos de etnia japonesa que representan las seis emociones básicas en cuatro grados diferentes de intensidad.

Para esta base de datos se realizan las mismas pruebas que en el apartado anterior para poder realizar una comparación de las mismas y ver con cuál de las dos bases de datos se obtienen mejores resultados. En este caso las pruebas que se llevan a cabo son ya con los nuevos cambios introducidos es decir, la implementación de la AU 26 y 27 y la nueva matriz de combinaciones (Tabla 36).

Una vez realizadas las pruebas con esta base de datos se procederá a analizar de manera conjunta todos los resultados obtenidos, tanto para esta base de datos como para la CK+ y para los dos métodos de segmentación utilizados.

Hay que destacar que el Árbol de Decisión sigue siendo el mismo para ambos métodos, ya que la matriz de entrenamiento es la misma que en el apartado anterior.

5.2.1. Método Thres

A continuación en la se obtiene la matriz de confusión con la base de datos *JAFFE* para el método indicado en este apartado. Para ello con el uso de la herramienta de la que se dispone se selecciona la base de datos *JAFFE* con el botón de *Change Folder*, se selecciona el método de extracción de características *Canny* junto con el *Árbol de Decisión* como clasificador y tras entrenar al mismo se pulsa el botón *Run test* que permite obtener dicha matriz de confusión.

Tabla 39 Matriz de confusión *JAFFE* para las combinaciones de Tabla 36, *Thres*

Tested expression-->	Anger	Disgust	Fear	Happiness	Sadness	Surprise
Detected expression						
Anger	42	48	39	10	59	33
Disgust	31	24	43	38	33	19
Fear	4	24	14	0	4	30
Happiness	8	0	4	41	0	4
Sadness	0	0	0	0	0	0
Surprise	15	4	0	10	4	15

Si comparamos esta matriz de confusión con la obtenida para la base de datos *CK+* (Tabla 37) se puede ver que los resultados de acierto para el método de *Thres* son peores con la base de datos *JAFFE* que con la *CK+* a nivel general. Para el caso de *sadness* se producen grandes confusiones con *anger* y *disgust* provocando que no detecte esta emoción en ninguna de las imágenes analizadas. *Fear* y *surprise* también se ven afectadas llegando a reducir su porcentaje de acierto del 42% para *fear* al 14% y del 64% al 15% para *surprise*. Esto sucede así ya que en la base de datos *JAFFE* existen grandes diferencias a la hora de expresar la emoción de *fear* en cada uno de los sujetos, incluso llegando a ser difícil identificar esta emoción de manera visual como se muestra en la Figura 36 . Lo mismo ocurre con *surprise*, ya que algunos de los sujetos asiáticos no abren la boca a la hora de expresar sorpresa, de ahí a que el sistema realice mal la clasificación y se produzca tanta confusión.



Figura 36 Imágenes de la base de datos *JAFFE* con la emoción de *fear*

5.2.2. Método Canny

Para concluir la parte de obtención de resultados se saca la matriz de confusión con el uso del método de *Canny*, tal como se muestra en la Tabla 40.

Tabla 40 Matriz de confusión JAFFE para las combinaciones de Tabla 36, Canny

Tested expression-->	Anger	Disgust	Fear	Happines	Sadness	Surprise
Detected expression						
Anger	54	56	54	38	67	33
Disgust	15	4	18	28	19	22
Fear	12	16	21	7	4	22
Happines	4	0	0	28	4	0
Sadness	12	20	7	0	4	19
Surprise	4	4	0	0	4	4

Al igual que sucedía con el método *Thres*, los porcentajes de acierto descenden con respecto al uso de la base de datos *CK+* para el mismo método. Llegando a no reconocer prácticamente ni *disgust*, ni *sadness*, ni *surprise* con un porcentaje de acierto del 4% y un porcentaje de confusión con *anger* entre el 67% y el 33%. En la se pueden observar imágenes de la base de datos *JAFFE* con la emoción de *anger* y *sadness* (emociones muy confundidas entre sí) en las que se puede apreciar que los sujetos asiáticos no mueven las comisuras de los labios hacia abajo (CAU 3), sino que encogen los labios como sucede para la emoción de *anger*.



Figura 37 Imágenes de la base de datos JAFFE con la emoción de *anger* (izquierda) y *sadness* (derecha)

5.3. Discusión de resultados

Con relación a las tablas mostradas en los apartados anteriores para diferentes bases de datos y para los dos métodos de extracción de características que se tienen, se lleva a cabo un análisis y discusión de los resultados.

Antes de nada hay que destacar que los porcentajes de acierto no son los que se esperaba encontrar al inicio de esta investigación, y es por ello por lo que se decide realizar un análisis en profundidad de cuáles pueden ser los motivos que han llevado a estos resultados tan bajos.

En primer lugar la precisión en la localización de los puntos es un factor que se tiene que tener muy en cuenta, ya que como ya se ha comentado una mala localización lleva a una detección de AUS y CAUs incorrecta y por tanto un reconocimiento de la emoción erróneo. Esto se ve claramente al comparar los resultados que se obtienen con el uso del método *Thres* (Tabla 37 y Tabla 39) y con el método *Canny* (Tabla 38 y Tabla 40), donde los porcentajes de acierto son generalmente mejores para el primer método que para el segundo, siendo el método *Thres* mejor a la hora de localizar los puntos característicos de la cara.

En segundo lugar se determina que el algoritmo de detección de AUS y CAUs no detecta de manera correcta las AUS o CAUs que se observan en algunas imágenes, lo que puede ocasionar equivocaciones a la hora de clasificar la emoción por parte del Árbol de Decisión.

En tercer lugar la matriz de combinaciones con la que se entrena el árbol, a partir de las combinaciones de la Tabla 36 que se han establecido, se considera pequeña para crear un árbol con información suficiente para un correcto reconocimiento de emociones. Como se puede ver en Figura 34 el árbol que se crea con esta matriz no contiene todas las AUs o CAUs y las AUs o CAUs que se muestran en los nodos principales (las más importantes) son las que aparecen en emociones que pueden ser más difíciles de reconocer y cuyas AUs son menos características. Por ejemplo en el Árbol de Decisión (Figura 34) no aparece la AU12 o AU-H para detectar *happy* dos Unidades de Acción muy características de esta emoción. Observando las tablas se puede ver que los porcentajes de acierto que se obtienen son aproximadamente mayores para las emociones de *anger*, *fear* y *surprise*, coincidiendo con las emociones que tienen un mayor número de filas de combinaciones en la matriz para su reconocimiento. Esto no sucede con la base de datos *JAFFE* por los problemas comentados a la hora de expresar estas emociones por parte de los sujetos asiáticos.

En cuarto lugar también se tienen que tener en cuenta las bases de datos que se utilizan para llevar a cabo el reconocimiento. Comparando las matrices de confusión que se muestran en la Tabla 37 y la Tabla 38 para la *CK+* con las de la Tabla 39 y la Tabla 40 para la *JAFFE*, se puede determinar que los resultados son mejores con el uso de la base de datos *CK+*. Esto se debe a que los sujetos asiáticos son menos expresivos que los sujetos occidentales (*CK+*) y a que el número de imágenes que posee la *JAFFE* es menor que el de la *CK+* (240 *JAFFE* y 414 *CK+*), a pesar de que la *CK+* tiene diferente número de imágenes para cada emoción (Tabla 33) hecho determinante en el reconocimiento.

En quinto y último lugar no hay que olvidar que de las seis emociones que se pretenden detectar existen algunas de ellas que tienen características comunes o poco diferenciadas, como puede ser *anger*, *fear* y *sadness* y esto puede hacer que se confundan entre ellas. Todo lo contrario ocurre con las emociones de *happy*, *surprise* y *disgust* que tienen características diferentes entre sí y no se deben confundir con el resto. Si se observan los resultados que se obtienen en las tablas de los apartados anteriores se puede ver que esto si se cumple, es decir *surprise* no se suele confundir ni con *happy* ni con *disgust* (porcentajes de confusión del 0% - 4% para *happy* y del 7% al 22% para *disgust*) y *disgust* tampoco se suele confundir ni con *happy* ni con *surprise* (porcentajes de confusión del 0% al 2% para *happy* y del 4% al 9% para *surprise*). Sin embargo sí se produce confusión de *happy* con *disgust* (del 17% al 38%), pero no de *happy* con *surprise* (del 0% al 10% de confusión).

5.3.1. Pruebas de diferenciación de emociones efusivas y no efusivas

Tras la discusión de los resultados y tras haber determinado que la matriz de entrenamiento tiene un número de combinaciones muy bajo (30 filas) para las seis emociones que se quieren clasificar, se puede estimar que el árbol está sub-entrenado.

Así pues con el objetivo de solventar este problema se propone una posible solución. Se decide reducir el número de emociones a detectar a dos (emociones efusivas y no efusivas) para que el árbol disponga de información suficiente para la clasificación. De las seis emociones que se tienen las efusivas serían *happy* y *surprise* y las no efusivas las otras cuatro: *anger*, *fear*, *sadness* y *disgust*.

Una vez que se ha reducido el número de emociones a detectar y aplicando la misma matriz de entrenamiento (matriz con las combinaciones de la Tabla 36), se obtiene el Árbol de Decisión que se muestra en la Figura 38 .

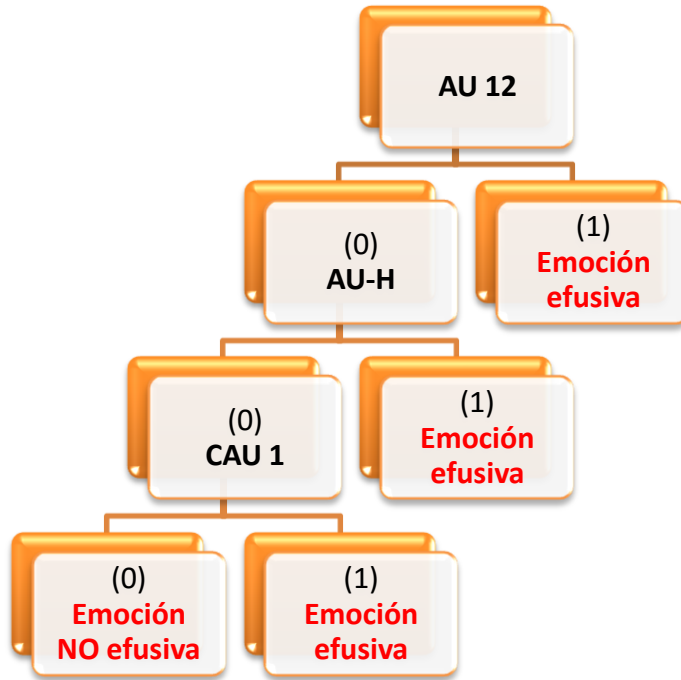


Figura 38 Árbol de Decisión para emociones efusivas y no efusivas

A diferencia de los árboles obtenidos en las pruebas anteriores, en este árbol se puede diferenciar de una manera clara las dos emociones. Los resultados de la matriz de confusión con estas dos emociones para imágenes de la base de datos CK+ y con el uso del método Thres, se pueden ver en la

Tabla 41 Matriz de confusión CK+ para emociones efusivas y no efusivas, Thres

Tested expression-->	Anger	Disgust	Fear	Happiness	Sadness	Surprise
Detected expression						
Emoción efusiva	13	5	38	61	41	90
Emoción no efusiva	87	95	63	39	59	10

Los resultados que se obtienen de la solución propuesta se han presentado en un documento oficial [39] en el cuál se propone un sistema que permita reconocer estas emociones en los sujetos y en función de la emoción detectada (efusiva o no efusiva) el sistema pueda realizar modificaciones. De tal manera que este sistema se ejecute en un segundo plano para poder realizar modificaciones sobre el sistema que está ejecutándose en primer plano en tiempo real.

6. Conclusiones

Este proyecto se ha centrado en el estudio de las Unidades de Acción (AUs) como método para el reconocimiento de emociones. Para ello se parte del trabajo realizado en el PFG se Sara González [3] y de las investigaciones realizadas durante el periodo de prácticas en el CITSEM [1].

Las emociones que se pretenden detectar son las seis emociones básicas, comúnmente utilizadas en la mayoría de los sistemas de reconocimiento. Inicialmente se ha tratado de analizar y resumir los métodos y técnicas que existen hoy en día en el reconocimiento automático de emociones, para centrarse en el análisis de diferentes publicaciones que usan las Unidades de Acción.

Con este estudio se han determinado cuales son las AUs que se consideran más importantes en cada una de las emociones, así como sus combinaciones posibles. Paralelamente se ha realizado un análisis subjetivo que permite contrastar los análisis que se obtienen con el estudio de las publicaciones y de igual manera se han obtenido tanto las AUs características como las combinaciones para cada emoción. Para finalmente llevar a cabo un análisis final en el que se han determinado de manera definitiva las AUs características y las combinaciones, introduciendo el concepto de CAUs (Unidades de Acción Combinadas) para reducir la lista de AUs y poder simplificar el proceso.

Por otro lado se ha hecho un análisis de la precisión en la colocación de los puntos característicos del rostro en los métodos de extracción de características (*Canny* y *Thres*) que se proporcionan en [3], así como la introducción de ciertas mejoras en la segmentación en el método *Thres*. A su vez se ha desarrollado e implementado un algoritmo que permite la detección de las AUs y CAUs presentes en los rostros que se quieren analizar, y se han indicado las reglas que se utilizan para ello.

Como método de clasificación de la emoción se propone el uso del Árbol de Decisión, el cual utiliza las AUs y CAUs que se han establecido para clasificar la emoción. Así pues se lleva a cabo la implementación del Árbol de Decisión a partir del trabajo realizado en el PFG de Diego Zapatero [2]. Una vez que se ha implementado el árbol es necesario realizar el entrenamiento del mismo a partir de una matriz que contiene las combinaciones de AUs y CAUs para cada emoción, para posteriormente realizar la clasificación de la emoción.

Para poder probar todas estas implementaciones se utiliza la herramienta del [3] a partir de la cual se introducen ciertas mejoras y avances, y con ella se evalúan los distintos métodos de extracción de características y el de clasificación, con diferentes combinaciones entre ellos.

Para evaluar el funcionamiento de la herramienta y la precisión de cada uno de los métodos, se llevan a cabo una serie de pruebas a partir de las cuales se han obtenido unos resultados en forma de tablas. El sistema de reconocimiento se ha probado con diferentes bases de datos, para poder evaluar cómo afecta a dicho sistema características como la iluminación, la calidad de la imagen y el número de imágenes utilizadas. Tras la realización de estas pruebas se han extraído varias conclusiones que se comentan a continuación.

La herramienta de la que se dispone permite un manejo fácil y rápido de la misma y da la opción de poder comparar los resultados que se obtienen con diferentes métodos, ya sea de clasificación o de extracción, y bases de datos. Viendo los resultados del reconocimiento en la propia interfaz.

A la hora de seleccionar una base de datos para el reconocimiento de emociones es muy importante tener en cuenta el número de imágenes de la misma y la calidad de las imágenes. Cuantas más imágenes se tengan mejores resultados se obtienen pero el rendimiento del sistema empeora. También es importante que número de imágenes para cada emoción sea el

mismo ya que esto puede hacer que los porcentajes de acierto sean peores en las emociones de las que se tiene un menor número de imágenes como es el caso de *sadness*. Para la detección de las AUs también es importante tener en cuenta la base de datos que se utiliza, debido a la raza de los sujetos que presentes en la misma. Como se mostró en el apartado 5, los sujetos occidentales con respecto a los asiáticos tienen diferentes formas de expresar las emociones y este es un hecho que influye en las AUs, desde las AUs o CAUs que se eligen como más representativas hasta la forma en la que se pueden detectar.

En cuanto a los métodos de extracción de características de los que se dispone se ha podido comprobar que el método de Thres realiza una mejor localización de los puntos característicos que el método Canny, y esto influye a la hora de obtener los resultados.

Por otro lado el algoritmo de detección de AUs y CAUs no es lo suficientemente preciso y si a esto se le suma el hecho de que la localización de los puntos no termina de ser buena, la detección de las AUs que están presentes en las imágenes no es del todo acertada.

Para el método de clasificación se determina que es necesario el uso de una matriz que permita entrenar al sistema sin llegar a provocar un aprendizaje excesivo del Árbol de Decisión o demasiado escaso es decir, una matriz que proporcione un equilibrio y con la que se pueda crear un árbol ni demasiado grande ni demasiado pequeño.

En general la emoción que mejor se ha conseguido detectar es *anger* para todas las pruebas realizadas, seguido de *surprise* para la CK+ y de *happy* para la JAFFE. Obteniendo resultados de acierto muy bajos para las emociones de *disgust* y *sadness* en todas las pruebas realizadas.

7. Trabajo futuro

Teniendo en cuenta los progresos realizados hasta el momento en este proyecto se podrían seguir realizando algunas mejoras con el uso de la herramienta de reconocimiento de la que se dispone. Se podrían implementar nuevos métodos de extracción de características y de clasificación, como por ejemplo métodos basados en mallas como es el caso del método AAM (*Active Appearance Model*) y las Redes Neuronales como método de clasificación. Así se podrían comparar los resultados con lo implementado en este proyecto.

Ampliar la base de datos CK+ para obtener la misma cantidad de imágenes para cada una de las emociones que se pretenden detectar. Otra propuesta es crear o buscar una nueva base de datos que contenga numerosas imágenes con diferentes características, esto ayudaría al sistema a reconocer mejor las emociones presentes en las imágenes.

A pesar de haber mejorado la localización de los puntos característicos de la cara y teniendo en cuenta que la incorrecta detección de los puntos depende en muchos casos de la presencia de arrugas, sombras, barbas etc., se propone seguir mejorando la localización de los puntos y su precisión (sobre todo mejorar la detección de los puntos de la boca) mejorando la segmentación que se realiza en cada una de las regiones de la cara. Así como la detección de un mayor número de puntos en la cara que ayuden a definir mejor las AUs y CAUs que se quieren detectar.

Por otro lado se puede seguir mejorando y precisando las reglas que se han establecido para la detección de las AUs y CAUs, por ejemplo introduciendo otros nuevos descriptores o usando varios a la vez. Cuanto más precisas sean mejor será la clasificación de la emoción por parte de Árbol de Decisión.

En este proyecto tanto como en el de [3] los métodos que se implementan se centran en el análisis de imágenes estáticas, por lo que se propone ampliar la herramienta para el análisis de

secuencias de video con los correspondientes métodos que lo permitan. También sería necesaria la creación o el uso de una base de datos con secuencias de vídeos.

Por último se propone como ya se ha mencionado anteriormente la mejora de la matriz de entrenamiento o la creación de una nueva para que el Árbol de Decisión pueda obtener mejores resultados a la hora de clasificar las emociones.

8. Referencias

- [1] D. Z. A. Gil, "Reconocimiento facial de emociones," *Mem. Final Prácticas externas en CITSEM*, p. págs. 1–24, 2015.
- [2] D. Zapatero, "Herramienta de reconocimiento facial de emociones en Android," Escuela Técnica Superior de Ingeniería y Sistemas de Telecomunicación, 2016.
- [3] S. González, "Desarrollo de una herramienta para analizar métodos de reconocimiento de emociones," Escuela Técnica Superior de Ingeniería y Sistemas de Telecomunicación, 2015.
- [4] P. Viola y M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. 2001 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognition*, 2001, pp. CVPR 2001, vol. 1, págs. I–511–I–518.
- [5] P. Viola y M. Jones, "Active appearance models," in *IEEE Trans. pattern analysis and machine intelligence*, 2001, p. vol. 23, no. 6, págs. 681–685.
- [6] M. Potzsch, "Internet publication of the University of Bochum," 1996.
- [7] Wikipedia-Enciclopedia libre, "Transformada de Haar," 2014. [Online]. Available: https://pt.wikipedia.org/wiki/Transformada_de_Haar.
- [8] C. Truesdell, "Acknowledgments," in *Pure Appl. Math*, 1980, p. pág. xxv.
- [9] S. Moore y R. Bowden, "Local binary patterns for multi-view facial expression recognition," in *Comput. Vis. Image Underst.*, 2011, p. vol. 115, no. 4, págs. 541–558.
- [10] Wikipedia-Enciclopedia libre, "Red neuronal artificial," 2016. [Online]. Available: http://es.wikipedia.org/wiki/Red_neuronal_artificial.
- [11] P. E. Group, "FACS Archives - Paul Ekman Group, LLC.," 2016. [Online]. Available: <http://www.paulekman.com/product-category/facs/>.
- [12] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, I. Matthews, and F. Aue, "The Extended Cohn-Kanade Dataset (CK +): A complete dataset for action unit and emotion-specified expression," . *Vis. Pattern Recognit. Work.*, no. July, pp. 94–101, 2010.
- [13] N. U. Khan, "A Comparative Analysis of Facial Expression Recognition Techniques," *Adv. Comput. Conf. (IACC), 2013 IEEE 3rd Int.*, pp. 1262–1268, 2012.

- [14] and D. B. S. J. McKenna, S. Gong, R. P. Würtz, J. Tanner, "Tracking facial feature points with Gabor wavelets and shape models," in *Proc. Int. Conf. Audio- Video-Based Biometric Person Authent.*, p. vol. 1206, pp. 35–42.
- [15] W. Friesen and P. Ekman, "Emfacs-7: emotional facial action coding system," 1983.
- [16] "The CMU Multi-PIE Face Database," *Carnegie Mellon University*, 2009. [Online]. Available: <http://www.multipie.org/>
- [17] Y. Li, S. Wang, Y. Zhao, and Q. Ji, "Expression Recognition," vol. 22, no. 7, pp. 2559–2573, 2013.
- [18] and J. G. T. F. Cootes, C. J. Taylor, D. H. Cooper, "Active shape models-their training and application," in *Comput. Vis. Image Understand*, 1995, p. vol. 61, no. 1, pp. 38–59.
- [19] and Q. S. C. X. W. Hou, S. Z. Li, H. J. Zhang, "Direct appearance models," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2001, p. vol.1.pp. 828–833.
- [20] M. Taner Eskil y K. S. Benli, "Facial expression recognition based on anatomy," in *Comput. Vis. Image Underst.*, 2014, p. vol. 119, págs. 1–14.
- [21] T. R. Friedman J, Hastie T, "Additive logistic regression: a statistical view of boosting," *Ann Stat*, 1998.
- [22] "MMI Database," *Intelligent Behaviour Understanding Group (iBUG), Department of Computing, Imperial College London*, 2002. [Online]. Available: <http://ibug.doc.ic.ac.uk/research/mmi-database/>.
- [23] C. G. Kohler, T. Turner, N. M. Stolar, W. B. Bilker, C. M. Brensinger, R. E. Gur, and R. C. Gur, "Differences in facial expressions of four universal emotions," *Psychiatry Res.*, vol. 128, no. 3, pp. 235–244, 2004.
- [24] D. Capino, "Facial Action Coding System Inspired Tutorial," 2009. [Online]. Available: <http://es.scribd.com/doc/18649644/Facial-Action-Coding-System-Khappucino-s-Tutorial>.
- [25] J. Iglesias, A. Loeches, and J. Serrano, "Expresión facial y reconocimiento de emociones en lactantes," in *Infancia y Aprendizaje*, Madrid, 1989, pp. 93–113.
- [26] C. Darwin, "The expression of emotions in man and animals," *Londres, J. Murray. Traducción castellana "La expresión las Emoc. en los Anim. y en el hombre"*, Madrid, Alianza, 1984.
- [27] L. Zhang, M. Jiang, D. Farid, and M. A. Hossain, "Intelligent facial emotion recognition and semantic-based topic detection for a humanoid robot," *Expert Syst. Appl.*, vol. 40, no. 13, pp. 5160–5168, 2013.
- [28] M. Gacitúa Rocha, "Ecología de la expresión facial: codificación y decodificación de emociones," *Universidad Autónoma de Madrid*, 2007.
- [29] J. Hamm, C. G. Kohler, R. C. Gur, and R. Verma, "Automated Facial Action Coding

System for dynamic analysis of facial expressions in neuropsychiatric disorders,” *J. Neurosci. Methods*, vol. 200, no. 2, pp. 237–256, 2011.

- [30] M. F. Valstar and M. Pantic, “BIOLOGICALLY VS . LOGIC INSPIRED ENCODING OF FACIAL ACTIONS AND EMOTIONS IN VIDEO.”
- [31] “JAFFE Database.” [Online]. Available: <http://www.kasrl.org/jaffe.html>.
- [32] “Imágenes de caras con emociones,” *Google*, 2016. [Online]. Available: https://www.google.es/search?hl=es&site=img&tbm=isch&source=hp&biw=1366&bih=667&q=face+recognition&oq=face+recognition&gs_l=img.3..0l4j0i30l6.1272.5816.0.6002.24.12.3.9.8.0.196.1235.0j9.9.0....0...1ac.1.64.img..4.14.869...0i10.Qa_Yy_TCMag#hl=es&tbm=isch&q=face+expressions.
- [33] C. Silva, L. Schnitman, and L. Oliveira, “Detection of Facial Landmarks Using Local-Based Information,” *19th Ed. Brazilian Conf. Autom. - CBA 2012, Camp. Gd. PB, Brazil (oral Present.*, no. c, 2012.
- [34] C. Silva, “Facial Landmarks,” 2015. [Online]. Available: <http://www.mathworks.com/matlabcentral/fileexchange/?term=id%3A47713>.
- [35] y B. A. B. Seddik, H. Maâmatou, S. Gazzah, T. Chateau, N. Essoukri, “Unsupervised Facial Expressions Recognition and Avatar Reconstruction from Kinect,” in *10th International Multi- Conference on Systems, Signals & Devices (SSD)*, 2013, p. págs. 1–6.
- [36] C. P. and A. D. N. Aifanti, “The MUG Facial Expression Database,” *Proc. 11th Int. Work. Image Anal. Multimed. Interact. Serv. (WIAMIS)*, Desenzano, Italy, 2010.
- [37] Tom M. Mitchell, “Machine Learning,” in *McGraw-Hill Science*, 1997, pp. pp. 52–78.
- [38] S. G. y Á. Martínez, “Interfaces Naturales para Realidad Aumentada,” *Inf. Final prácticas, CITSEM*, p. págs. 1–16, 2014.
- [39] J. F. M. M. Eckert, A. Gil, D. Zapatero, J. Meneses, “Fast facial expression recognition for emotion awareness disposal,” *Res. Cent. Softw. Technol. Multimed. Syst. Sustain.*