ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA Y SISTEMAS DE TELECOMUNICACIÓN

PROYECTO FIN DE GRADO

TÍTULO: Desarrollo de una herramienta para analizar métodos de reconocimiento de emociones

AUTOR: Sara González Martín

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: Amador Miguel González Crespo

VOCAL: Martina Eckert

SECRETARIO: Enrique Rendón Angulo

Fecha de lectura: 26 de Mayo de 2015

Calificación:

El Secretario,

Agradecimientos

En primer lugar quiero agradecer a mi tutora Martina Eckert la oportunidad brindada al poder colaborar en el centro de investigación CITSEM y su apoyo al realizar mi Proyecto Fin de Grado.

Igualmente estoy muy agradecida a los miembros del CITSEM que han compartido conmigo el trabajo y esfuerzo realizado, por darme la oportunidad de iniciarme en el mundo de la investigación y por impulsarme a continuar ampliando mis conocimientos. En especial quiero agradecer a mis compañeros Irene Sánchez, Álvaro Martínez y Almudena Gil por su colaboración y por haberme permitido disfrutar realizando este proyecto.

También quiero agradecer a mis amigos y compañeros conocidos durante mi carrera universitaria, los grandes momentos que hemos compartido y que han hecho de esta una experiencia maravillosa.

Por último, quiero agradecer a mi familia por el apoyo mostrado a lo largo de mi vida y a mis amigos por esos momentos de diversión y distracción del ámbito de estudio.

Gracias a todos por estos años que nunca olvidaré.

Sara González
Mayo 2015

Resumen

Desde hace más de 20 años, muchos grupos de investigación trabajan en el estudio de técnicas de reconocimiento automático de expresiones faciales. En los últimos años, gracias al avance de las metodologías, ha habido numerosos avances que hacen posible una rápida detección de las caras presentes en una imagen y proporcionan algoritmos de clasificación de expresiones.

En este proyecto se realiza un estudio sobre el estado del arte en reconocimiento automático de emociones, para conocer los diversos métodos que existen en el análisis facial y en el reconocimiento de la emoción. Con el fin de poder comparar estos métodos y otros futuros, se implementa una herramienta modular y ampliable y que además integra un método de extracción de características que consiste en la obtención de puntos de interés en la cara y dos métodos para clasificar la expresión, uno mediante comparación de desplazamientos de los puntos faciales, y otro mediante detección de movimientos específicos llamados unidades de acción. Para el entrenamiento del sistema y la posterior evaluación del mismo, se emplean las bases de datos *Cohn-Kanade+* y JAFFE, de libre acceso a la comunidad científica.

Después, una evaluación de estos métodos es llevada a cabo usando diferentes parámetros, bases de datos y variando el número de emociones.

Finalmente, se extraen conclusiones del trabajo y su evaluación, proponiendo las mejoras necesarias e investigación futura.

Abstract

Currently, many research teams focus on the study of techniques for automatic facial expression recognition. Due to the appearance of digital image processing, in recent years there have been many advances in the field of face detection, feature extraction and expression classification.

In this project, a study of the state of the art on automatic emotion recognition is performed to know the different methods existing in facial feature extraction and emotion recognition. To compare these methods, a user friendly tool is implemented. Besides, a feature extraction method is developed which consists in obtaining 19 facial feature points. Those are passed to two expression classifier methods, one based on point displacements, and one based on the recognition of facial Action Units. Cohn-Kanade+ and JAFFE databases, both freely available to the scientific community, are used for system training and evaluation.

Then, an evaluation of the methods is performed with different parameters, databases and varying the number of emotions.

Finally, conclusions of the work and its evaluation are extracted, proposing some necessary improvements and future research.

Índice

Lista de acrónimos

HCI: Human Computer Interaction (Interacción hombre-máquina)

PCA: Principal Component Analysis (Análisis de componentes principales)

LDA: Linear Discriminant Analysis (Análisis de discriminación lineal)

AAM: Active Appearance Model (Modelo de apariencia activa)

EBGM: Elastic Bunch Graph Matching (Adaptación elástica de un conjunto de grafos)

FACS: Facial Action Coding System (Sistema de codificación de acciones faciales)

AU: Action Unit (Unidad de acción)

LBP: Local Binary Patterns (Patrones binarios locales)

SIFT: Scale Invariant Feature Transform (Transformada de características invariante en escala)

ANN: Artificial Neural Network (Redes neuronales artificiales)

SVM: Support Vector Machine (Máquina de soporte vectorial)

k-NN: k-Nearest Neighbour (k-vecinos más próximos)

CK+: Extended Cohn-Kanade (Cohn-Kanade extendida)

JAFFE: Japanese Female Facial Expression (Expresión facial de mujeres japonesas)

MIT: Massachusetts Institute of Technology (Instituto tecnológico de Massachusetts)

LGBP: Local Gabor Binary Patterns (Patrones binarios locales con Gabor)

Índice de figuras

Índice de tablas

1. Introducción

Durante los últimos años, el reconocimiento facial se ha convertido en una de las áreas más estudiadas en el ámbito de la investigación. Una de las razones es la necesidad de desarrollar aplicaciones de seguridad y vigilancia (por ejemplo la seguridad en los automóviles o la videovigilancia), y mejorar la comunicación entre las máquinas y las personas, HCI (*Human Computer Interaction*). Un ejemplo de ello es su aplicación en robótica, pero existen otros ámbitos de aplicación como: la biometría, seguridad de la información, autenticación de usuarios, videojuegos, marketing, psicología, medicina y rehabilitación.

Una de las ventajas del reconocimiento automático de expresiones faciales es que es no intrusivo, pero sigue siendo necesario mejorar la detección e identificación de la emoción en imágenes adquiridas en condiciones de iluminación insuficiente, con oclusión parcial de la cara, con inclinación del rosto y en personas con vello facial o signos del envejecimiento (como arrugas).

Por otro lado, el desarrollo en procesado digital de la imagen y en los gráficos por ordenador, han permitido realizar numerosos avances en el reconocimiento automático de emociones. Debido a la gran cantidad de métodos existentes, es muy difícil poder desarrollar uno nuevo que aporte mejoras, sin aumentar el gasto computacional o ralentizar el sistema (lo que repercutiría de forma directa en el reconocimiento de emociones en tiempo real). Aún más difícil es comparar resultados de forma directa, evaluando qué métodos son mejores para cada condición, realizando pruebas sobre una misma base de datos o conocer cómo afecta la modificación de ciertos parámetros en la clasificación de las emociones.

Uno de los objetivos principales de este proyecto es dar una supervisión sobre los métodos implementados hasta la actualidad y sus aplicaciones. Se propone una herramienta que parte de una versión muy básica creada en el ámbito de prácticas de empresas en el centro CITSEM durante el semestre de primavera de 2013/14 [37], con la cual se pretende ofrecer una solución intuitiva y accesible que permita el análisis y la comparación sobre distintos materiales (imágenes estáticas, vídeo…), tanto en métodos de extracción de características y clasificación de la emoción, como en combinaciones de ellos.

Entre las funcionalidades que este software incluye están:

- Obtención de la fuente a analizar (cara). Puede ser una imagen estática o un vídeo que provenga de una base de datos o de una webcam.
- Elección del método de extracción de características. Consiste en localizar puntos o características relevantes en la cara, para realizar un análisis posterior e identificar las diferentes emociones.
- Elección del método de clasificación de la emoción y funcionalidad de aprendizaje. Según la fuente a analizar, se ejecuta un algoritmo de aprendizaje automatizado (por ejemplo árbol de decisión).
- Ventana de pruebas. Mediante la creación de una interfaz gráfica, cuando la herramienta reciba material desconocido para analizar, ejecutará el algoritmo seleccionado de reconocimiento y clasificación de la emoción, que habrá sido anteriormente entrenado.

Para poder obtener resultados sobre métodos concretos, se ha desarrollado un sistema de reconocimiento automático de emociones basado en la localización de puntos característicos de la cara, de manera que se puedan detectar expresiones faciales mediante un sencillo método clasificador de la emoción. Antes de la extracción de características, se realiza un pre-procesado de la imagen para eliminar ruido o características redundantes y mejorar los resultados de la posterior clasificación de la emoción.

Este proyecto está organizado de la siguiente manera. El capítulo dos presenta el estado del arte en el ámbito de la detección de la cara, extracción de características y clasificación de la emoción mostrando algunas ventajas o diferencias entre ellos. Además, contiene información sobre las bases de datos utilizadas. En el capítulo tres se describe la herramienta creada en este proyecto, cuya función es la comparación de diferentes métodos, y se explican las funcionalidades implementadas en los distintos

módulos que la componen. A continuación, en el capítulo cuatro se describen los métodos propuestos para el reconocimiento automático de emociones, tanto el método de extracción de características como los métodos de clasificación de la emoción, y en el capítulo cinco se muestran los resultados obtenidos utilizando distintas combinaciones de los métodos propuestos y bases de datos. Por último, en el capítulo seis se extraen conclusiones y se proponen futuras líneas de investigación.

2. Antecedentes

2.1. Introducción

P. Ekman and W. V. Friesen [1] definieron seis expresiones básicas (*anger*, *disgust*, *fear*, *happiness, surprise* y *sadness*), en base a estudios psicológicos realizados sobre las diferencias entre ellas y la capacidad de distinguirlas. Los sistemas de reconocimiento de emociones que se han desarrollado hasta la actualidad buscan clasificar estas expresiones, y el nuevo reto de los últimos años es reconocer expresiones menos expresivas y más sutiles que estas seis básicas (*non-basic expressions*).

Existen diferencias relevantes entre imágenes que contienen sujetos expresando emociones de forma espontánea o si se les ha pedido que posen en un laboratorio (*posed expressions*). Es usual que las expresiones de las imágenes tomadas en un laboratorio sean artificiales y normalmente exageradas, pero es difícil crear una base de datos que contenga imágenes o vídeos de sujetos expresando emociones de forma espontánea.

Por otro lado, existen otro tipo de expresiones como las microexpresiones o las llamadas 'expresiones silenciadas' [2]. Las microexpresiones son aquellas que ocurren en un periodo de tiempo inferior a un segundo y que, por ese motivo, son muy difíciles de capturar y analizar. Las expresiones silenciadas son aquellas que han comenzado a mostrarse, pero que el sujeto, por el motivo que sea, ha finalizado de forma abrupta.

Es importante tener en cuenta la dinámica temporal cuando se expresa una emoción. Hay una fase de aparición de la expresión (*onset*), de máximo grado (*apex*) y de desaparición (*offset*) [2]. Esta característica adquiere relevancia cuando se trata de reconocimiento de emociones en tiempo real o sobre archivos de vídeo. Además, las cuatro regiones más significativas cuando se muestran emociones son, por orden de importancia: la boca, las cejas y los ojos [2].

En general, los sistemas automáticos de reconocimiento de expresiones faciales se dividen en cuatro etapas: detección de la cara, extracción de características, aprendizaje del sistema y clasificación de la expresión [3]. Se puede observar en el diagrama de bloques de la Figura 1.

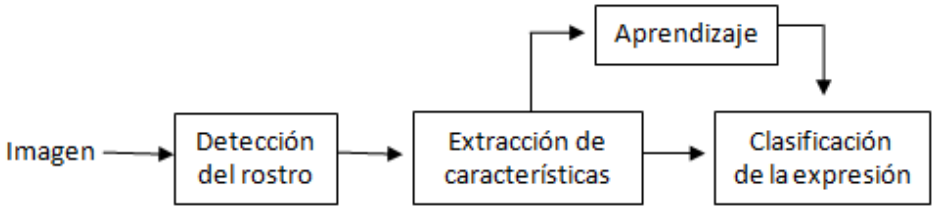


Figura 1. Diagrama de bloques para reconocimiento de emociones.

2.2. Detección y seguimiento de la cara

A principios de los años 90, B. D. Lucas y T. Kanade [4], propusieron un método para **la detección y el seguimiento de la cara** que fue después mejorado por C. Tomasi y T. Kanade [5]. Se trata de un acercamiento a la extracción de características, y ofrece una mayor rapidez al detectar posibles coincidencias entre imágenes que las técnicas que existían anteriormente. En 1998, T. Kanade y H. Schneiderman [5] desarrollaron un algoritmo de detección de objetos usando métodos estadísticos, que permitía la detección de la cara independientemente de su posición (de frente, hacia la izquierda o hacia la derecha). En 2001, P. Viola y M. Jones [6] implementaron un método para detectar objetos en tiempo real con una gran rapidez. Sus principales características son el uso del algoritmo de aprendizaje basado en AdaBoost, y la combinación de diferentes clasificadores por orden de complejidad (en forma de "cascada"), lo que permite descartar las regiones con baja probabilidad de que sean el objeto de interés, es decir, permite descartar las regiones del fondo (Figura 2).

AdaBoost (*Adaptative Boost*) es un algoritmo de aprendizaje que se caracteriza por su bajo índice de error y que consiste en la suma ponderada de diferentes algoritmos débiles de aprendizaje [7]. En el área de detección de rostros, el algoritmo busca las características que más información aportan sobre una cara, entre todas las previamente obtenidas en una imagen [6]. Actualmente es el método de detección automática de rostros que más se utiliza, y existen versiones que introducen cambios como la modificación de AdaBoost por GentleBoost. El uso de GentleBoost permite aumentar la efectividad de cómputo y el rendimiento del clasificador, ya que reduce el número de características a tener en cuenta, y proporciona una mayor estabilidad que AdaBoost.

Figura 2. Localización de la cara usando el método implementado por Viola & Jones.

Por otro lado, se están desarrollando técnicas de reconocimiento y seguimiento de rostros basados en modelos parametrizados, de los cuales uno de los más populares es *Candide-3*. Se trata de una máscara basada en modelos de rostros humanos y que usa un conjunto de polígonos (vértices y superficies) como se muestra en la Figura 3. Esta malla se superpone a las caras y se adapta a ellas.

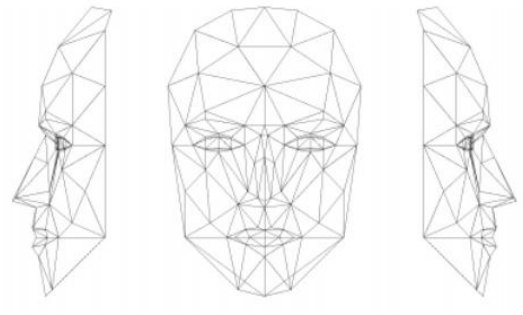


Figura 3. Modelo facial usando método Candide-3 [8].

Existen diferentes versiones para *Candide,* la más extendida es *Candide-1* que trabaja con 79 vértices, 108 superficies y 11 unidades de acción, pero la más reciente es *Candide-3*, que fue desarrollada para poder adaptarse al nuevo estándar de animación MPEG-4 y está definido por 113 vértices, 168 superficies y seis unidades de acción [8]. Las unidades de acción, que se explicarán en profundidad en el apartado 2.3, están relacionadas con la acción de un único músculo o un conjunto de ellos. Concretamente para el modelo *Candide*, se definen unidades de acción globales, que corresponden a acciones de rotación alrededor de los ejes realizadas por las diferentes regiones del rostro. Cada unidad de acción está asociada a los vértices que la provocan, y cada vértice tiene su correspondiente vector de desplazamiento.

A partir de imágenes del rostro de un sujeto (de frente y de perfil) se extraen características que correspondan a los vértices utilizados en el modelo *Candide-3*, y de esta forma se puede adaptar dicho modelo al rostro que se quiere reconocer.

En este proyecto se usa el algoritmo de *Viola & Jones* debido a su baja tasa de error y su buen rendimiento, con la adicional ventaja de que está incluido en la *Computer Vision Toolbox* de MATLAB, programa utilizado para el desarrollo del proyecto.

2.3. Extracción de características

Una vez que la cara está detectada, el siguiente paso es la **extracción de características**, es decir, la extracción de información relevante en el rostro para identificar y clasificar las distintas emociones.

Los métodos de extracción de características se pueden dividir en dos grupos:

- **Extracción de características geométricas (*Geometric feature extraction).*** Contienen información sobre la forma y la localización espacial de las características de una imagen facial, por ejemplo ojos y boca. El problema de este tipo de métodos es su dependencia de la correcta detección y seguimiento de la cara, y de la selección de los puntos característicos (si no son marcados manualmente, los posibles errores afectan directamente al reconocimiento de emociones). Por otro lado suelen ser más sensibles al ruido. Entre los métodos basados en características geométricas se encuentra EBGM (*Elastic Bunch Graph Matching*).

- **Extracción de características de apariencia (*Appearance features methods*).** En ellos son examinados los cambios globales en la cara, teniendo en cuenta las texturas (diferencias de luminosidad o color, cambios de dirección, variaciones en el tamaño y forma, etc.). Las características se extraen mediante el filtrado de la imagen o de una región de la misma. Por otro lado, la información obtenida es almacenada en un vector cuya dimensión tiene un gran tamaño, lo que hace necesario utilizar técnicas de reducción como PCA (*Principal Component Analysis*) o LDA (*Linear Discriminant Analysis*). Entre los métodos basados en características de apariencia se encuentra AAM (*Active Appearance Model*), FACS (*Facial Action Coding System*), LBP (*Local Binary Patterns*) y SIFT (*Scale Invariant Features Transform*).

El método **AAM** fue introducido por Cootes, Edwards y Taylor en 1998 [9]. Se basa en adaptar un modelo estadístico de apariencia previamente creado a partir de puntos de referencia de un conjunto de imágenes, con la imagen que se pretende analizar, utilizando el algoritmo maximizador de la esperanza [9].

El modelo estadístico es creado en una fase de entrenamiento, donde adquiere la forma y apariencia del objeto de interés. En el área de reconocimiento de emociones, para cada imagen de entrenamiento, se crea una malla triangular en 2D a partir de un conjunto de 68 puntos característicos de la cara que representan la forma del rostro (se puede ver un ejemplo en la Figura 4). Para todas ellas, se realiza la normalización, el análisis de componentes principales (PCA) para obtener las variaciones y se obtiene la malla modelo, que consiste en una media de las mallas de entrenamiento y un subespacio vectorial con las variaciones principales [10], [11].

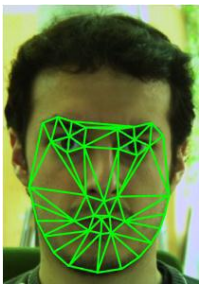


Figura 4. Ejemplo de malla usando el método AAM [11]

Una de sus ventajas es que se beneficia tanto de las características de la forma, como de las características de textura presentes en una cara para crear un modelo eficiente y que utiliza tanto puntos de los bordes como puntos interiores. La desventaja es, que necesita como entrada además de la imagen, los puntos característicos, y su rendimiento es afectado negativamente.

Basado en la teoría de grafos se encuentra la técnica de **EBGM.** En primer lugar, se normalizan las imágenes y se buscan las coordenadas correspondientes a los ojos para ubicarlos en una posición predeterminada. En segundo lugar, mediante un modelo estadístico y la operación de convolución, se

ajusta un grafo a los puntos faciales más importantes del individuo analizado. Dependiendo de la distancia del grafo ajustado, con respecto al grafo modelo, se detecta el rostro del individuo o no. Se puede ver un ejemplo en la Figura 5.

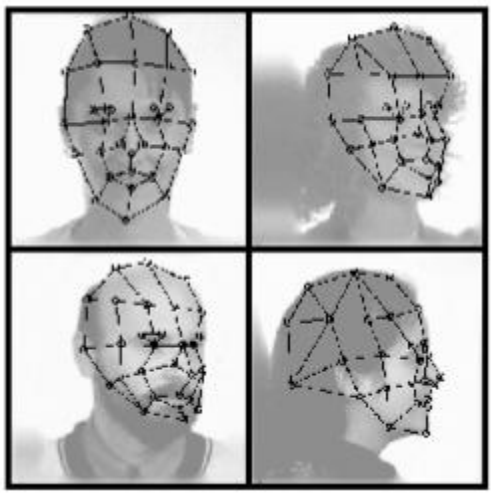


Figura 5. Ejemplo de grafo resultado de la aplicación de la técnica EBGM [12]

Para implementar este método, concretamente para la representación de las características locales, se utilizan filtros Gabor. Se trata de unos filtros casi paso-banda espaciales que permite obtener información frecuencial en una región de la imagen. La transformada de la respuesta al impulso de Gabor consiste en la convolución de la transformada de Fourier de la función gaussiana y de la función sinusoidal. Los filtros Gabor son funciones que se puede definir como un banco de filtros, cada uno con diferentes orientaciones y frecuencias, de manera que cada punto del grafo se determina por sus coordenadas y por la respuesta ante un determinado banco de filtros, como se puede apreciar en la Figura 6 [12].

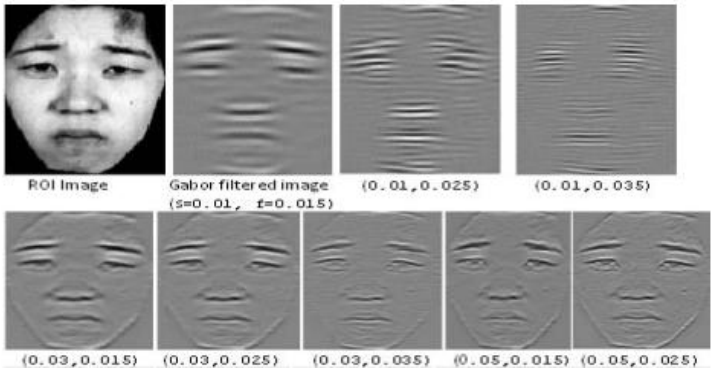


Figura 6. Filtrado Gabor utilizando distintos parámetros de orientación y frecuencia[13]

Los músculos de una cara humana no varían entre individuos y razas. En este hecho se basa el método de codificación de movimiento facial **FACS** desarrollado por Ekman y Friesen en 1978 [14]. Este método consiste en asociar los cambios que se producen en la cara al expresar una emoción con las acciones de los músculos que los provocan. Estas acciones se denominan AUs (*Action Units*), y se definieron 46 basándose en estudios psicológicos. Una AU se refiere a la acción de un único músculo o al conjunto de ellos que provocan una acción visual característica. En la Tabla 1 se muestra las AUs que son consideradas las más representativas para el reconocimiento de emociones, y en la Tabla 2 se pueden observar algunos ejemplos sobre posibles combinaciones de AUs. En la Tabla 3 se muestran las seis expresiones básicas relacionadas con las *Action Units* que estarían en acción. Además, FACS ofrece cinco grados de intensidad para la emoción (Tabla 4), que son principalmente relevantes para las *Action Units* 25 (boca entre abierta), 26 (mandíbula caída), 27 (boca abierta por el labio inferior), 41 (párpado caído), 42 (ojos entrecerrados) y 43 (ojos cerrados).

Tabla 1. Definición de las AUS más representativas para reconocimiento de seis emociones [15]

AU	Imagen representativa	Definición	AU	Imagen representativa	Definición
AU1		Interior de las cejas elevado	**AU12**		Comisuras de los labios tirantes
AU2		Exterior de las cejas elevado	**AU15**		Comisuras de los labios deprimidas (hacia abajo)
AU4		Cejas bajadas	**AU17**		Barbilla elevada
AU5		Parpado superior elevado	**AU23**		Labios tirantes, tensos
AU6		Mejillas elevadas	**AU24**		Labios presionados
AU7		Parpados estrechados	**AU25**		Labios separados
AU9		Nariz arrugada	**AU26**		Boca entreabierta
AU10		Labio superior elevado	**AU27**		Boca abierta

T*abla 2. Ejemplos de AUs y su combinación según FACS [8]*

AU	Nombre	Ejemplo
1	Interior de las cejas elevado	
2	Exterior de las cejas elevado	
1+2	Interior y exterior de las cejas elevado	

Tabla 3. Descripción de emociones basada en AUs [16] [17]

Emoción	Descripción	AU en acción
Anger	Interior de las cejas junto y hacia abajo, ojos abiertos. Labios presionados o entreabiertos enseñando los dientes.	AU4, AU23, AU24
Disgust	Cejas y ojos relajados. Labio superior elevado y curvado frecuentemente de forma asimétrica. Nariz arrugada	AU9, AU10
Fear	Cejas elevadas y juntas. Ojos tensos y alerta. Frecuentemente boca entreabierta.	AU1+AU2, AU5, AU20
Happiness	Cejas relajadas, boca abierta con las comisuras estiradas hacia las orejas.	AU6, AU12, (AU25)
Sadness	Interior de las cejas hacia arriba, ojos entrecerrados y comisuras de la boca hacia abajo.	AU1, AU7, AU15, (AU16)
Surprise	Cejas elevadas. Ojos muy abiertos con el párpado superior elevado. Boca muy abierta.	AU1+AU2, AU5, AU27, (AU25, AU26)

Tabla 4. Grados de intensidad definidos en FACS [18]

Grado de intensidad	Denominación
Trace	A
Slight	B
Marked/Pronunced	C
Severe/Extreme	D
Maximum	E

La principal desventaja que presenta el **FACS** es la dificultad de identificar AUs cuando la cara esta inclinada o parcialmente tapada. Además, es importante elegir una cantidad adecuada de puntos característicos de la cara, a partir de los cuales se obtienen las diferentes AUs, ya que un número insuficiente de ellos no daría una buena precisión [18], [19]. También es importante que los puntos que se detectan sean realmente característicos, que aporten información relevante. Por ejemplo, un punto situado en medio de la frente no varía (o muy poco) al expresar una u otra emoción, es decir, apenas aporta información. Sin embargo, un punto situado en uno de los extremos de la boca, como por ejemplo en el labio inferior, es muy significativo ya que la boca adopta diferentes aperturas dependiendo de la emoción que se exprese.

Otro método que ha tenido mucho éxito y cada vez es más usado por los investigadores en el área de reconocimiento facial, es **LBP** (*Local Binary Patterns*) [20]. Se trata de un método introducido por T. Ojala en 2002 [21], basado en texturas (bordes, puntos, esquinas...) y es capaz de detectar movimiento. Sus principales ventajas son su simplicidad de cómputo y su tolerancia a cambios de iluminación.

Su funcionamiento consiste en realizar la derivada circular de orden uno sobre un píxel y su vecindario, normalmente de tamaño 3x3 (pero puede ser mayor). En función del pixel central, se evalúan los pixeles del vecindario mediante una umbralización en escala de grises, como muestra la siguiente función:

$$S(x)\begin{cases}1, & x \geq 0\\ 0, & x < 0\end{cases} \tag{1}$$

y se obtiene un patrón de textura resultado de la concatenación de la dirección del gradiente binario:

$$LBP_{P,R} = \sum_{P=0}^{7} S(f_p - f_c)2^p \tag{2}$$

donde *P* hace referencia al número de puntos del vecindario, *R* al radio del círculo y 2^p es un factor binomial. Cuando el pixel vecino es mayor que el pixel central se obtiene el valor 1, en caso contrario, el 0. En la Figura 7 se pueden ver varios ejemplos en función del valor de *P* y *R* [21].

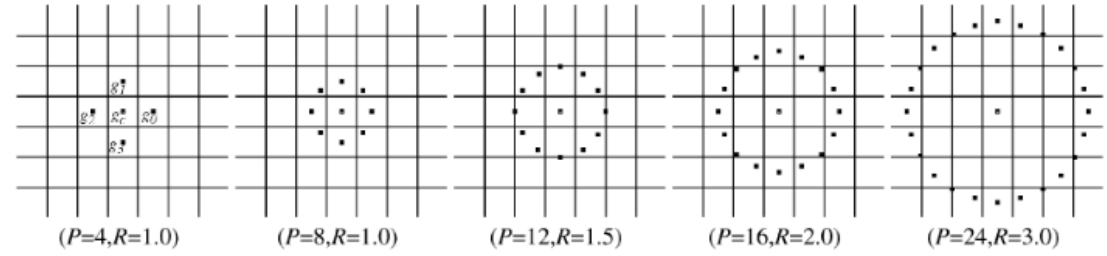


Figura 7. Ejemplo de la derivada circular en función del tamaño del vecindario y del radio del círculo.

Se muestra un ejemplo en la Figura 8, donde se obtiene un patrón binario de 8 dígitos (*P*=8). En la primera celda (arriba a la izquierda), que corresponde a un píxel vecino, su valor es mayor que el del pixel central (165> 149), por lo que el valor es 1.

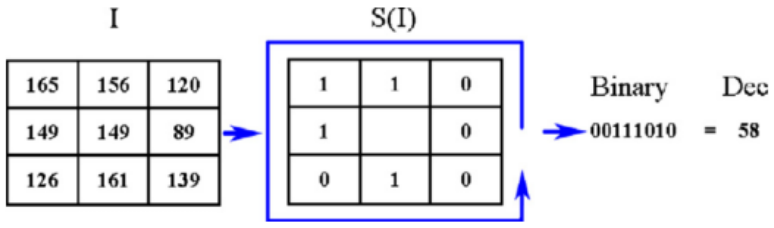


Figura 8. Ejemplo del uso del operador LBP [22].

A partir de los patrones de textura obtenidos con las ecuaciones descritas anteriormente es creado el histograma LBP, que contiene información sobre bordes, áreas lisas o manchas. Las regiones de radio R son necesarias para aportar información espacial y representar de forma eficiente el rostro.

Otro método de reconocimiento de objetos que se basa en características de texturas, es **SIFT** (*Scale Invariant Feature Transform*). Fue completamente descrito en 2004 por David Lowe [23]. Al igual que AAM, parte de las coordenadas de una serie de puntos de interés, y crea un vector de características utilizando la magnitud del gradiente de cada uno de los puntos.

El proceso completo es el siguiente. Primero se calculan las magnitudes y orientaciones de los gradientes para cada punto, después estas magnitudes son ponderadas usando la función de Gauss, y por último, las magnitudes ya ponderadas se representan en un histograma de orientación, en el cuál los picos se consideran las orientaciones más relevantes del punto evaluado. El descriptor final se obtiene concatenando todos los histogramas [10].

Tanto LBP como SIFT son métodos basados en la captura de patrones de micro textura y utilizan un histograma para representar la distribución de dichos patrones. El uso de un histograma hace que estos métodos sean robustos frente al ruido. Pero con el fin de optimizar el tiempo de procesado, se establecen unos límites y cierta información espacial es descartada.

Es importante tener en cuenta que un vector de textura tiene un tamaño mayor que un vector de características geométricas y los métodos que usan filtros Gabor son más sensibles a los cambios de tamaño de la región de la cara que los métodos SIFT o LBP [24]. Además, en [24] se concluye que la combinación de métodos basados en características geométricas y métodos basados en texturas, proporcionan mejores resultados que si se usan dichos métodos por separado.

2.4. Clasificación de la expresión

Una vez que las principales características han sido extraídas, se procede a **la clasificación de la expresión** facial, es decir, a la identificación de una de las seis expresiones básicas anteriormente mencionadas, que a veces son siete (cuando se incluya por ejemplo *Contempt*) y que son las que normalmente se detectan en los métodos de reconocimiento de emociones.

Para poder clasificar una expresión facial es necesario un previo **aprendizaje** del sistema en función del método de clasificación que se emplea. Se parte de una base de datos que contenga suficientes imágenes de sujetos expresando las emociones que se desea que el sistema clasifique, y se aplican los correspondientes métodos de extracción de características y de clasificación de la emoción sobre cada una de las imágenes. Dichas imágenes están etiquetadas con el nombre de la expresión que representan, de forma que el sistema aprende las características propias de cada expresión y permite la posterior diferenciación y clasificación de cada una de ellas.

Uno de los métodos de clasificación consiste en la creación de **árboles de decisión**, que se basan en los atributos de los objetos de estudio para su clasificación. Un árbol de decisión está formado por nodos de decisión y por nodos finales (hoja). Cada nodo de decisión está asociado a un atributo y a sus posibles valores, y los nodos-hoja determinan el objetivo del árbol de decisión, es decir, son el resultado de cada rama de dicho árbol. Se puede ver un ejemplo de árboles de decisión y los tipos de nodo en la Figura 9.

Un algoritmo que se emplea normalmente para la creación de árboles de decisión es ID3, que fue creado por J. Ross en 1979 [25]. Se basa en la fórmula de la entropía E para calcular la incertidumbre:

$$\mathrm{E(s)} = \sum_{\mathrm{i=1}}^{\mathrm{c}} -\mathrm{p_i log_2 p_i}, \tag{3}$$

donde *s* es el número total de atributos y p_i es la proporción de las veces que se cumple un atributo entre el total (*s*), pero es modificada según:

$$\mathrm{E(s)} = -\mathrm{P}\log_2(\mathrm{P}) - \mathrm{N}\log_2(\mathrm{N}) \tag{4}$$

donde *P* son los casos positivos y *N* los negativos.

El árbol se define mediante una serie de atributos. Para cada elemento de entrada se determina si cumple o no cada uno de ellos, es decir, se le asignan valores de 0 o 1, y a continuación se calcula la Entropía (2). Si la entropía es mínima, dicho atributo se convierte en nodo-hoja, es decir, se obtiene un resultado. En caso contrario, la rama analizada se sigue dividiendo hasta que se llegue a nodos-hoja.

Para el reconocimiento de expresiones, los atributos del árbol de decisión pueden ser las AUs (explicados en el apartado anterior) asociadas a cada una de las expresiones que se quiere identificar. De esta manera, el cumplimiento o no de una AU en una imagen de entrada definiría cada nodo de decisión que forma el árbol. Se puede ver un ejemplo en la Figura 10.

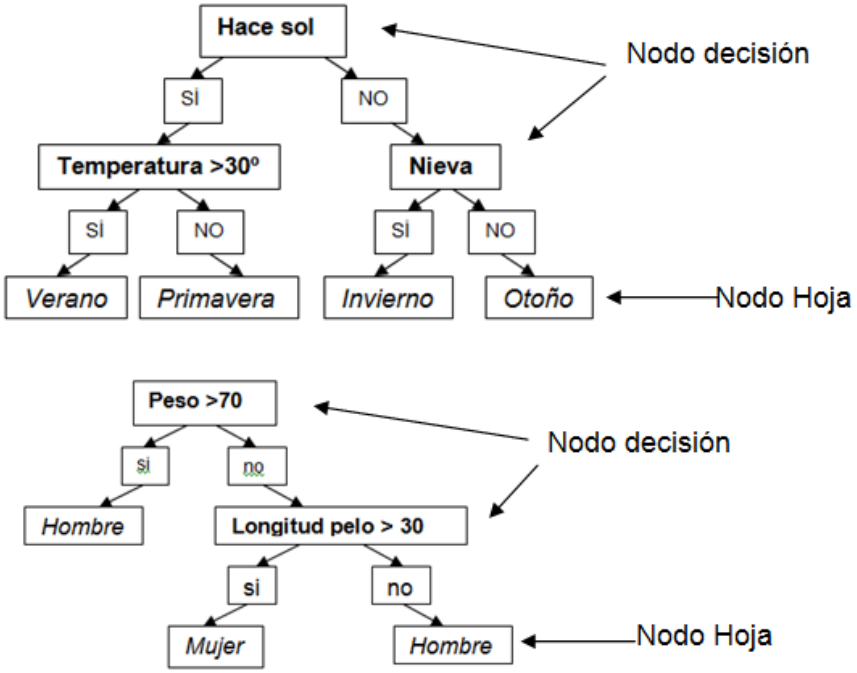


Figura 9. Ejemplo de árboles de decisión especificando algunos nodos de decisión y nodos hoja.

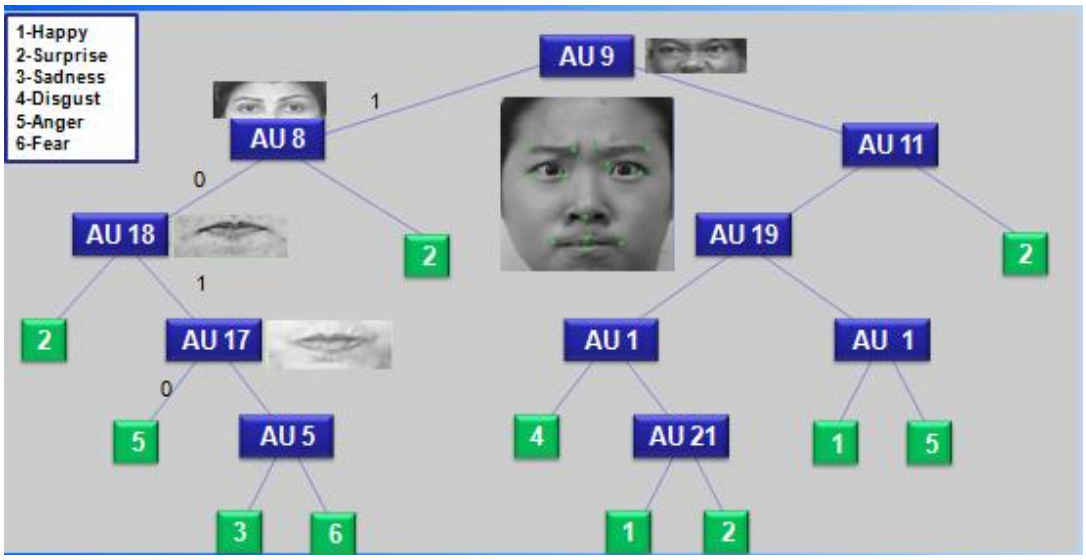


Figura 10. Ejemplo de árbol de decisión para reconocimiento de emociones [26].

También es usual combinar árboles de decisión con **redes neuronales**. El método de redes neuronales (*Neural Networks*) desarrollado por W. McCulloch y W. Pitts, y que ha evolucionado a **ANN** (*Artificial Neural Networks*), está basado en el sistema neuronal de los humanos. Su ventaja es la rapidez y la capacidad de auto aprendizaje [3].

Una red neuronal está formada por un conjunto de neuronas y por una capa de entrada, una o más capas ocultas y una capa de salida, como muestra la Figura 11. Cada neurona consta de una o varias entradas y una o varias salidas, cuyo valor depende de su estado de activación. Dicho estado de activación (con valor 1 o 0) es determinado mediante límites o condiciones.

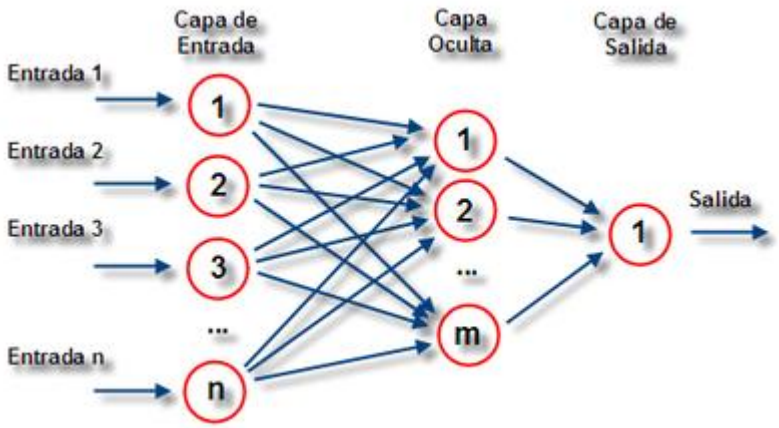


Figura 11. Ejemplo de Neural Network [27].

Para los sistemas de reconocimiento de emociones faciales el parámetro de entrada de las neuronas puede ser el nivel de gris, puntos de referencia o AUs entre otros, y lo usual es usar seis nodos de salida, uno para cada expresión.

Un método relacionado con las redes neuronales y el más utilizado actualmente es **SVM** (*Support Vector Machine*). Se trata de un modelo basado en aprendizaje de máquinas (*Machine Learning*) que fue por primera vez propuesto en 1992. Consiste en un clasificador binario lineal que permite determinar si un vector de entrada pertenece o no a un grupo definido previamente (se dispone solo dos posibles grupos y cada vector de entrada pertenece a uno u otro).

SVM busca un plano que separe adecuadamente los puntos pertenecientes a cada una de las dos categorías. Se puede ver en la Figura 12, donde el plano de separación es representado por una línea continua, y los puntos más cercanos al plano de separación pertenecen a la línea discontinua y forman el llamado vector de soporte.

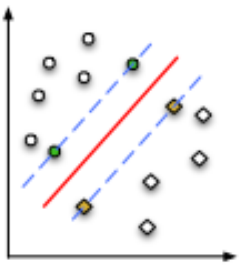


Figura 12. Separación de los puntos de entrada según la categoría mediante la línea continua [16].

La fórmula empleada al implementar SVM es

$$f(x) = \sum_{i=1}^{N} \propto_i y_i K(x_i, x) + b, \tag{5}$$

donde *K* hace referencia a las distintas funciones *Kernel* que se pueden usar [28].

Las funciones *Kernel* permiten la representación de la información en el espacio minimizando la carga computacional de las máquinas de aprendizaje. Consiste en una conversión de un espacio de entrada cuya clasificación es no lineal, a un espacio de mayor dimensión que permita clasificación lineal.

La principal ventaja de SVM frente a ANN es que el entrenamiento del sistema es muy eficiente y mucho menos costoso, dando ambos métodos una buena precisión en la clasificación. Además, SVM ofrece una mayor robustez.

Un método basado en probabilidad, también llamado método de clustering, es **k-NN** (*k-Nearest Neighbour*). Este algoritmo consiste en clasificar un objeto evaluando los *k* objetos vecinos más próximos [29].

En primer lugar, el algoritmo es entrenado con un conjunto de casos de ejemplo del cual se conoce el resultado. Ante un nuevo caso, se pretende estimar el resultado teniendo como base los ejemplos anteriormente mencionados y los k objetos más próximos. Para medir la distancia entre el nuevo objeto y los casos de ejemplo existen varias opciones, pero la más conocida es la Distancia Euclidiana cuya fórmula es:

$$D(x, m) = \sqrt{(x - m)^2}\,, \tag{6}$$

donde *x* corresponde con el valor del objeto analizado y *m* con el de ejemplo.

El valor de k es muy importante en la creación del modelo k-NN ya que influye directamente en la calidad de las estimaciones. Si k es demasiado grande, puede dificultar la diferenciación entre casos parecidos, pero si es muy pequeño aumenta la probabilidad de errores de clasificación. Para realizar una estimación de k es común usar el algoritmo de validación cruzada [29], que principalmente consiste en dividir los ejemplos disponibles en varios grupos, aplicar el modelo k-NN a cada uno de estos grupos utilizando diferentes valores de k y en función de los errores y aciertos obtenidos en cada caso, estimar el mejor valor de k.

k-NN asume que los vecinos más cercanos proporcionan la mejor clasificación, pero si el objeto tiene muchos atributos, es conveniente identificar aquellos que tienen mayor relevancia (por ejemplo, asignando pesos entre los diversos atributos).

En la Figura 13 se muestra un ejemplo de este método para distintos valores de k. El círculo verde es el objeto y los cuadrados y triángulos son los diferentes atributos a clasificar, y según el vecino más cercano se identificara el objeto como uno u otro.

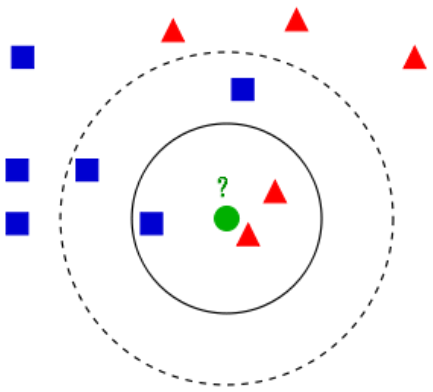


Figura 13. Ejemplo del resultado de aplicar un modelo K-NN a distintos valores de K [38]

Según el valor de k, el resultado del modelo k-NN en la Figura 13 es:

1-NN→ Triángulo. Solo hay un atributo que evaluar (el más cercano).

3-NN→ Triángulo. Entre los tres atributos más cercanos hay mayoría triángulos.

5-NN→ Cuadrado. Hay mayoría de cuadrados (tres cuadrados y dos triángulos).

Otro método usado en clasificación de la expresión es el basado en ***Eigenfaces*** [30]. Se trata de un método estadístico basado en el análisis de componentes principales (PCA) para el reconocimiento de caras (normalmente usado para reconocer personas). La idea es crear desde un conjunto de caras normalizadas, una combinación lineal que se aproxime mejor a la cara buscada.

Cada imagen facial es interpretada en escala de grises como un conjunto bidimensional de patrones claros y oscuros. Con un conjunto de imágenes, estos patrones son representados de forma vectorial (*Eigenvector*) formando una base de vectores (*Eigenface*) capaz de representar diferentes caras que poseen características comunes. En las Figura 14 y Figura 15 se muestra un ejemplo de la formación de *Eigenfaces*, en la Figura 14 se muestran las imágenes de entrenamiento y la imagen facial media de

dichas imágenes, y en la Figura 15 se muestran los *Eigenfaces* con los mayores *Eigenvalues* calculados en las imágenes de entrenamiento de la Figura 14 [30].

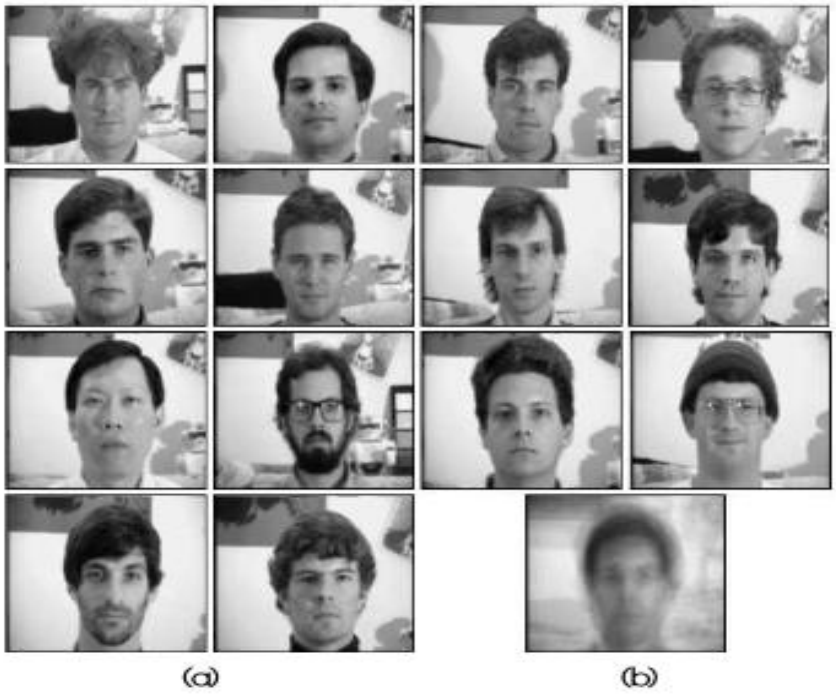


Figura 14. (a) Imágenes de entrenamiento. (b) imagen facial media [30].

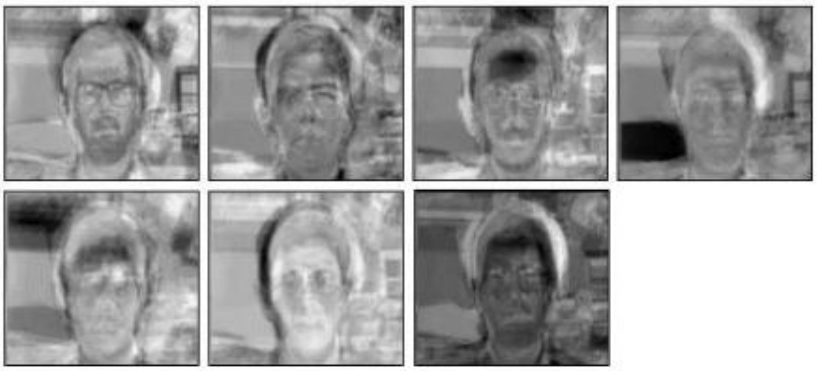


Figura 15. Eigenfaces calculados en las imágenes de entrenamiento de la Figura 14 [30].

En reconocimiento de emociones, cada emoción se obtiene eligiendo el *Eigenface* que más se aproxime a la imagen facial que se pretende analizar, ya sea una persona o una expresión concreta. Se trata de un método que no necesita una previa extracción de características. Debido a ello, es más sencillo y más rápido.

De entre los métodos mencionados , el más utilizado hasta el momento para la extracción de características ha sido FACS (que define Unidades de Acción) y en la actualidad el método de clasificación de la emoción más implementado es SVM, ya que su eficiencia en la clasificación es muy elevada y, como se comentó anteriormente, aunque es similar a la obtenida con redes neuronales (AAN) es algo más sencillo de implementar y más robusto.

Para el desarrollo de este proyecto se han elegido métodos sencillos tanto conceptualmente como de desarrollo, concretamente se ha basado en la extracción de puntos faciales que facilitan la implementación de un método de clasificación de la emoción propio de desplazamiento de puntos, y la introducción a la decisión por árboles utilizando Unidades de Acción.

Para poder evaluar la precisión de los métodos mencionados en los apartados 2.2, 2.3 y 2.4 es necesario disponer de bases de datos correctamente etiquetadas según el objetivo al que se quiere llegar, en este

caso, expresiones faciales. En el apartado siguiente se comentan algunas de las bases de datos actualmente disponibles para reconocimiento de emociones.

2.5. Bases de datos

Durante los últimos años, diferentes investigadores y universidades han recopilado bases de datos para el reconocimiento facial y/o de expresiones faciales. Algunas de estas bases de datos están formadas solo por imágenes, otras por vídeo y otras por audio y vídeo. Por otro lado, algunas presentan condiciones de iluminación no homogéneas, oclusiones (por gafas o vello facial) o vista frontal y lateral. En todas las bases de datos que se comentan a continuación, están identificados los tipos de expresión en cada archivo, y han sido recopiladas en un laboratorio, es decir, se ha pedido a los diferentes sujetos que posen expresando diferentes emociones.

La base de datos que ha sido mayormente utilizada por los investigadores en el área de reconocimiento automático de emociones es la que fue creada en el año 2000 y después mejorada por J. Cohn y T. Kanade, que se denomina *Extended Cohn-Kanade* (CK+) *database*. Consta de 210 sujetos entre hombres y mujeres adultos de diferentes nacionalidades. Para cada uno de ellos, se tomaron varios fotogramas en las que están representadas la expresión *neutral* como inicio y una expresión de las seis básicas (*anger*, *disgust*, *fear*, *happiness, sadness y surprise*) más otra expresión denominada en [16] como *contempt* (desprecio), con máxima intensidad al final. Para algunos sujetos existen imágenes para cada expresión, sin embargo, para otros solo se tienen clasificadas algunas de estas seis o siete expresiones, sin tener en cuenta la expresión *neutral*, que se encuentra para todos los sujetos.

Para cada sujeto, cada secuencia de fotogramas está identificada con una emoción concreta y etiquetada utilizando el método de extracción de características FACS. Como se explica en el capítulo 2.2, consiste en la definición de una serie de Unidades de Acción que permiten identificar la emoción correspondiente.

Las imágenes de esta base de datos presentan distintas condiciones de iluminación, pero sin oclusiones y con apenas diferencias en la pose. Se dispone a su vez de imágenes en escala de grises y en color dependiendo del sujeto. En la Figura 16 se puede observar un ejemplo para cada una de las expresiones, de izquierda a derecha y de arriba abajo las expresiones son: *disgust*, *happiness, surprise*, *fear*, *anger*, *contempt, sadness* y *neutral* [16].

Figura 16. Ejemplo de la base de datos CK+.

Otra base de datos también bastante utilizada es JAFFE (*Japanese Female Facial Expression*) recopilada por M. Lyons, M. Kamachi y J. Gyoba [31]. En este caso, las imágenes corresponden a 10 mujeres japonesas. Cada una de ellas representa las seis emociones básicas entre uno y cuatro grados de intensidad, por lo que de cada sujeto se pueden tener hasta 24 imágenes más la correspondiente neutral. En esta base de datos las condiciones de iluminación son homogéneas para todas las imágenes y no hay oclusiones ni cambios en la pose. En la Figura 17 se muestra un ejemplo de cada una de las expresiones para un sujeto (de izquierda a derecha y de arriba abajo, las expresiones mostradas son: *neutral*, *surprise*, *sadness*, *happiness*, *anger*, *disgust* y *fear),* y en la Figura 18 se puede observar un ejemplo de la expresión *anger* expresada en tres niveles distintos (o de tres formas distintas).

Figura 17. Ejemplo para un mismo sujeto.

Figura 18. Ejemplo de tres niveles de intensidad de la expresión anger para un mismo sujeto.

La base de datos **MMI**-*Facial Expression Database* es también utilizada normalmente en el área de reconocimiento de expresiones faciales. Fue creada por M. Pantic y M. F. Valstar en el año 2002 [32] y de libre acceso a la comunidad científica. Contiene 2900 vídeos de alta resolución correspondientes a 75 sujetos, capturados tanto en una pose frontal como lateral. En la Figura 19 se muestra como ejemplo seis sujetos expresando las seis emociones básicas (de izquierda a derecha: *happiness, sadness, fear, disgust, surprise* y *anger*), y en la Figura 20 se muestra un ejemplo para un sujeto con pose frontal y lateral.

El objetivo de recopilar esta base de datos fue suplir las deficiencias que tenían otras anteriormente creadas. La principal mejora fue introducir la secuencia de imágenes en modo *onset-apex-offset*, desde que un sujeto está representando la expresión neutral (*onset*), hasta que expresa una emoción (*apex*) y la posterior reconversión a la cara neutral (*offset*).

Al igual que la base de datos CK+ anteriormente comentada, utiliza el método FACS para etiquetar cada imagen. La diferencia con CK+ es que no contiene solo las seis expresiones básicas, sino que también contiene imágenes que solo representan una Unidad de Acción específica, sin pertenecer a una expresión determinada.

Figura 19. Ejemplo base de datos MMI para seis sujetos distintos

Figura 20. Ejemplo base de datos MMI para un sujeto con pose frontal y lateral.

Una base de datos de vídeo y audio es INTERFACE EMOTION Database, también conocida como eNTERFACES. Fue recopilada en 2005 por miembros del MIT (*Massachusetts Institute of Technology*) [33] y es usada para reconocimiento de emociones a partir de audio, vídeo o ambos juntos. Esta base de datos está formada por 42 sujetos entre hombres y mujeres de 14 nacionalidades. Algunos sujetos presentan oclusiones (gafas o barba), pero las condiciones de iluminación son homogéneas. En la Figura 21 se muestran ejemplos de instantes de los vídeos de un sujeto expresando seis emociones básicas (*fear, anger, disgust*, *happiness*, *sadness y surprise*).

Figura 21. Ejemplo de la base de datos eNTERFACES para un mismo sujeto [33].

La Tabla 5 contiene un resumen de las principales características de las bases de datos mencionadas.

Tabla 5. Resumen de las principales características de diferentes bases de datos

Base de datos	Nº Sujetos	Edad (años)	Sexo	Variedad étnica/oclusiones	Número expresiones
Cohn-Kanade+	210	18-50	69% mujeres 31% hombres	81% Europeos-americanos 13% Afro-Americanos 6% Otros grupos (asiáticos…) Sin oclusiones	7 + neutral
JAFFE	10*	--	100% mujeres	100% Japonés Sin oclusiones	6 + neutral
MMI	42	--	Más hombres que mujeres	Diversas nacionalidades Sin oclusiones	6 + neutral Otras **
eNTERFACE	42	--	19% mujeres 81% hombres	14 nacionalidades 31% gafas 17% barba	6

**4 intensidades por sujeto y expresión.*

*** Expresiones sólo caracterizadas por una AU.*

3. Descripción de la herramienta propuesta

En este proyecto se describe el diseño y la creación de una herramienta modular y ampliable, que posibilite la comparación de la efectividad de diferentes métodos de reconocimiento automático de emociones, pudiendo a la vez cambiar la combinación entre técnicas de extracción de características y algoritmos de clasificación. En el alcance de este proyecto se crean las bases de esta herramienta con dos métodos de extracción de características faciales geométricas (ambos basados en puntos marcadores) más sencillas que las mencionadas en el estado del arte y dos algoritmos de clasificación (basados en la comparación de desplazamientos de los puntos y en la detección e interpretación de unidades de acción).

Para facilitar el manejo de la herramienta se ha creado una interfaz gráfica en Matlab empleando la función GUIDE de Matlab. Esta función permite crear una ventana seleccionando los controles necesarios para realizar cada una de las funciones del sistema de la forma más intuitiva posible, así como para poder visualizar los resultados de forma unificada.

Para el desarrollo de este proyecto, se parte de una versión muy básica creada en el ámbito de prácticas de empresas en el centro CITSEM durante el semestre de primavera de 2013/14 [37]. Esta versión funcionó solamente con una configuración por defecto, que permitió evaluar un único método de extracción de características y un clasificador de la expresión, sin modulo propio de aprendizaje. Los objetivos de mejora realizados en este proyecto fueron:

- Reestructuración modular del código para posibilitar su ampliación con diferentes algoritmos de extracción de características y clasificación en el futuro
- Separación de un módulo de aprendizaje que se ejecuta independientemente de las pruebas a realizar
- Mejora del módulo de pruebas para poder separar las bases de datos en diferentes cantidades de imágenes para prueba y aprendizaje (*folds*) de forma automatizada
- Almacenamiento automático de los resultados en tablas de formato .xlsx
- Modificación de la interfaz de usuario para que sea más intuitiva y que incluya todas las etapas en las que se dividen los sistemas de reconocimiento automático de expresiones, como se comentó en el apartado 1 y que se puede observar en el diagrama de bloques de la Figura 22.

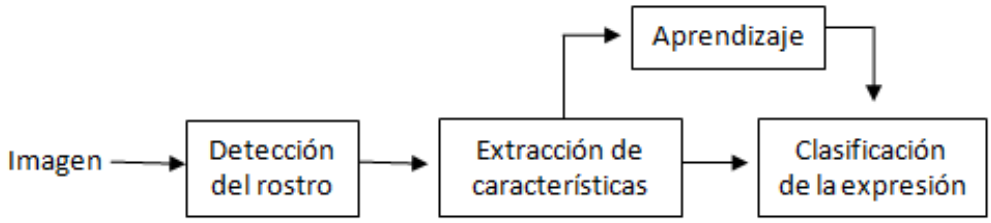


Figura 22. Diagrama de bloques para reconocimiento de emociones.

Para mejorar el aspecto general de la interfaz, se ha modificado la apariencia, tanto el color (para ser más agradable a la vista) como la distribución de las diferentes ventanas que lo componen, quedando una división en cuatro módulos como se muestra en la Figura 23.

La modificación más importante ha sido añadir un módulo (S*ettings*) que incluye las etapas del diagrama de bloques mostrado en la Figura 22. Permite elegir tanto el método de extracción de características como el método de clasificación de la emoción, y el cambio de uno a otro de forma rápida e intuitiva, y también permite realizar el entrenamiento del sistema en base a los métodos elegidos. Esta nueva característica es la base para poder comparar diversos métodos utilizando la herramienta.

Para poder implementar el carácter modular en la herramienta, se ha hecho una limpieza del código para que sea más eficiente y rápido. Se han eliminado elementos innecesarios, se han ordenado cada una de las partes que lo componen, y se ha implementado cada método (tanto de análisis como de

entrenamiento) en funciones independientes, adaptando todos los parámetros necesarios. A su vez, se ha unificado todo el código al idioma inglés (en la versión básica había mezcla entre español e inglés) y se ha comentado todo el código para que sea de fácil uso a futuros investigadores, y por lo tanto que se pueda ampliar. Se visualiza la interfaz completa de la herramienta en la Figura 23.

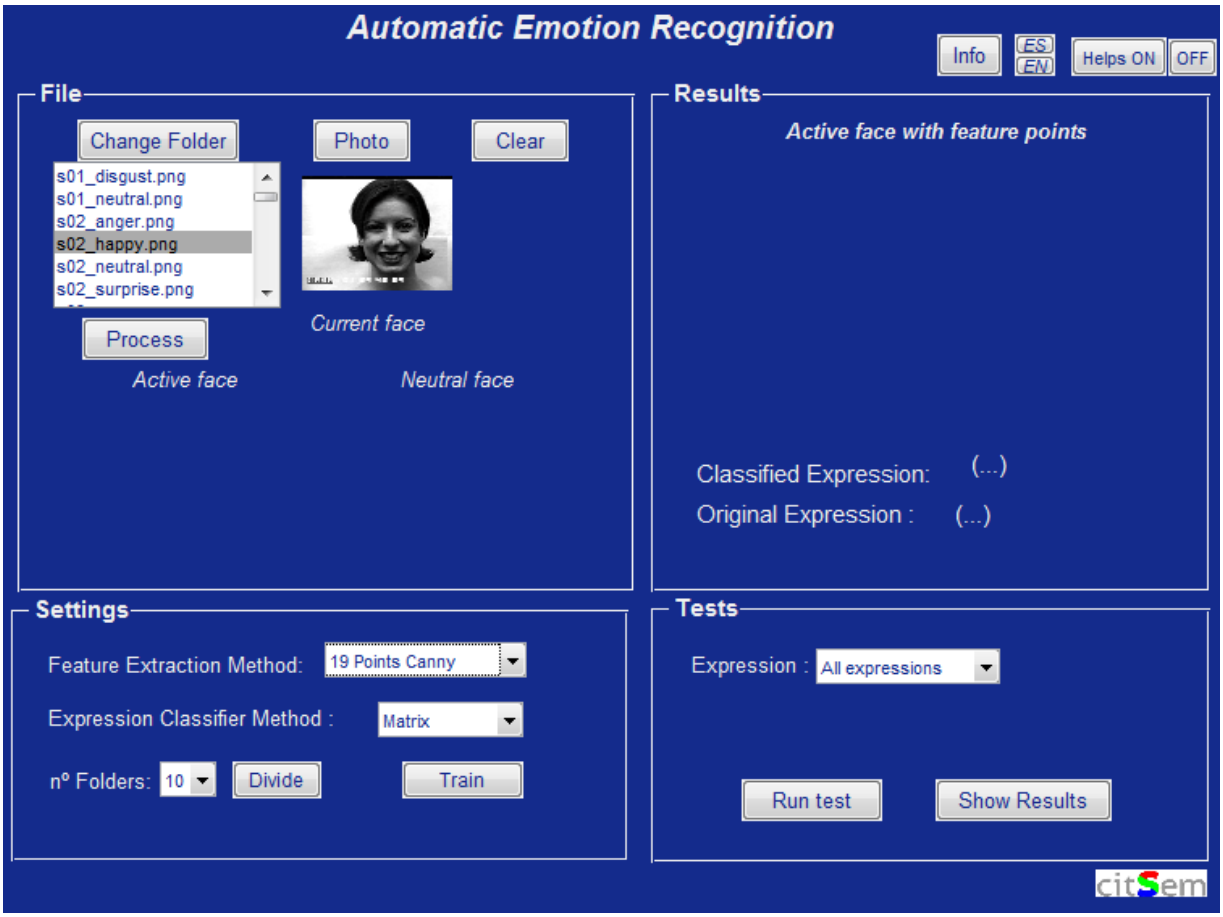


Figura 23.Interfaz gráfica de la herramienta para comparar métodos de reconocimiento de emociones

La interfaz está dividida en cuatro ventanas que contienen los módulos de la herramienta y que se describen en detalle a continuación: ventana **FILE** (obtención de la imagen a clasificar), ventana **SETTINGS** (configuración de los métodos de extracción de características y de clasificación), ventana **RESULTS** (presentación del resultado) y ventana **TESTS** (conducción de pruebas automatizadas).

- **FILE**: en esta ventana se obtiene la fuente a analizar. Para ello hay dos opciones: cargar los archivos correspondientes a una base de datos (opción por defecto), o tomar una fotografía del sujeto a analizar mediante la webcam del ordenador (botón *Photo*). Si es necesario cambiar la base de datos de estudio, se dispone de un botón que permite cambiar el directorio que contiene las imágenes (*Change folder*).
  Las imágenes de las bases de datos que se utilizan para esta herramienta están nombradas identificando primero al sujeto mediante un número y después la expresión que contiene dicha imagen, por ejemplo *s01_happy* o *s20_neutral*. De forma que es necesario ordenar y renombrar las bases de datos utilizadas en base a este formato.

  Se pueden previsualizar las imágenes de la base de datos (*Current face*) desplazándose por la lista que contiene este módulo mediante el ratón o con las teclas flecha del teclado. Una vez seleccionada la imagen que se quiere procesar, se ejecuta el reconocimiento mediante el botón *Process* de la interfaz o con el botón *Enter* del teclado. El programa busca la correspondiente neutral y muestra ambas imágenes como se puede ver en la Figura 24 (*Active*

face es la imagen mostrando alguna emoción y *Neutral face* es la imagen neutral correspondiente). La forma que tiene la herramienta de encontrar la imagen neutral es buscando primero al sujeto (s01, s20…) y de entre todas sus imágenes, buscar la denominada *neutral*.

Para el caso de elegir la opción de obtener la imagen desde la webcam, el programa pide al usuario obtener dos fotografías, una con expresión y otra mostrando su cara neutral. También pide al usuario que indique una base de datos para utilizar el aprendizaje realizado con ella y poder clasificar la emoción.

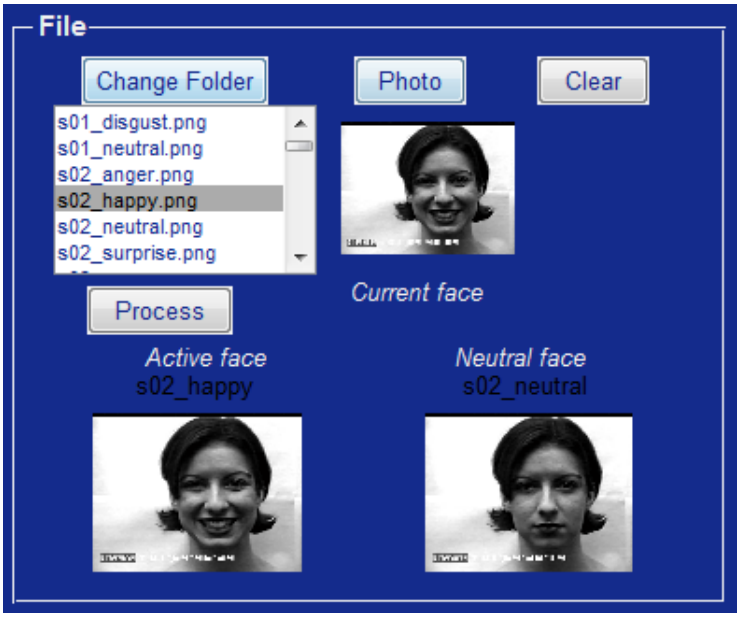


Figura 24. Módulo FILE cuando se selecciona una imagen a analizar.

Una vez procesadas dichas imágenes, en la parte inferior de la interfaz aparece el tiempo total que ha tardado en clasificarse la emoción, desde que se detectó la cara en la imagen. Además, se pueden eliminar todos los datos referentes a las imágenes y a la detección de la expresión mediante el botón *Clear*, y la interfaz recuperaría la misma apariencia que en la Figura 22.

- **SETTINGS**: Esta ventana permite seleccionar los métodos que se van a utilizar para reconocer la emoción que es expresada por un sujeto. Permite tanto elegir el método de extracción de características como el método de clasificación de la expresión y realizar el enteramiento previo necesario para cada uno de ellos. Gracias al módulo **TESTS**, que se explica más adelante en este capítulo, se pueden comparar los resultados de cada uno de ellos y de su combinación.
  Mediante dos menús desplegables se eligen los métodos que se aplican a la imagen seleccionada en el módulo **FILE**. Se pueden incluir tantos métodos adicionales como se desee de forma independiente, pudiendo eliminarse o modificarse en cualquier momento, y sin tener influencia en el resto de la herramienta. Es importante adaptar los métodos de clasificación que se añaden a los datos obtenidos en la extracción de características, es decir, si un método de clasificación necesita características diferentes a las que se extraen con el método de extracción de características seleccionado, se indica con un mensaje de aviso, de manera que se pueda proceder a seleccionar uno diferente, o a implementarlo en la herramienta.

  Para los métodos de clasificación de la emoción que requieren un previo aprendizaje de la herramienta, cuando son seleccionados aparece la opción (botón *Train*) para ejecutar dicho entrenamiento. Esta funcionalidad, que no estaba integrada en la interfaz básica, es necesaria en los sistemas de reconocimiento automático de emociones y facilita la utilización a usuarios no muy familiarizados con la herramienta y el proceso.

  Por otro lado, para que los resultados obtenidos sean válidos (se explicará en el apartado 5) es necesario dividir la base de datos en diversos grupos de forma que unos sirvan para entrenar al

sistema y otros para probarlo, concretamente se entrena el sistema con el 90% de las imágenes de la base de datos y se prueba con el 10% restante. Para ello se ha introducido en este módulo la opción de dividir la base de datos en un número de grupos (*folds*) elegido por el usuario, es decir, repartir las imágenes de la base de datos en dichos grupos. Por ejemplo, si se divide la base de datos en diez *folds*, a través de la herramienta se utilizan nueve *folds* para entrenamiento, y una *fold* para prueba. Para dichos grupos se ha establecido un formato de escritura/lectura. Se almacenan las imágenes de cada grupo en un directorio (creado de forma automática) identificando el numero del grupo junto al total de grupos, por ejemplo *Folder_1de10* o *Folder_3de5*. Al incluir esta funcionalidad, ha sido además necesario adaptar las fases de aprendizaje y de obtención de resultados para su correcta ejecución.

Para los datos obtenidos en la fase de aprendizaje también se establece un formato de escritura/lectura. Se crea un directorio con el nombre del método de clasificación de la emoción (o un nombre adecuado para su identificación) y se nombra al archivo obtenido con el nombre de la base de datos utilizada, el método de extracción de características y si procede, con el número del *fold* al que pertenece. Por ejemplo, para uno de los métodos implementados que se explicará en el apartado 4, se crea el directorio "Matrices representativas" en el cual se encuentran archivos .mat con el nombre *ck clasificada_Folder_1de10_19 Points Canny.*

- **RESULTS**: En esta ventana se visualiza el resultado de la extracción de características, (por ejemplo la localización de un determinado número de puntos) y también el resultado de la clasificación.

  Si el archivo de entrada pertenece a una base de datos, y tiene etiquetada la expresión correspondiente, dicha expresión aparecerá en el apartado *Original Expression* y la emoción detectada por el sistema se mostrará como *Classified Expression.* Se puede ver un ejemplo en la Figura 25. Si por el contrario, el archivo de entrada procede de una foto obtenida mediante la webcam, no se mostrará nada en *Original Expression*, ya que no hay un título para esa imagen capturada, sino que el propio usuario podrá comprobar si la expresión que intentaba expresar coincide con la detectada por la herramienta.

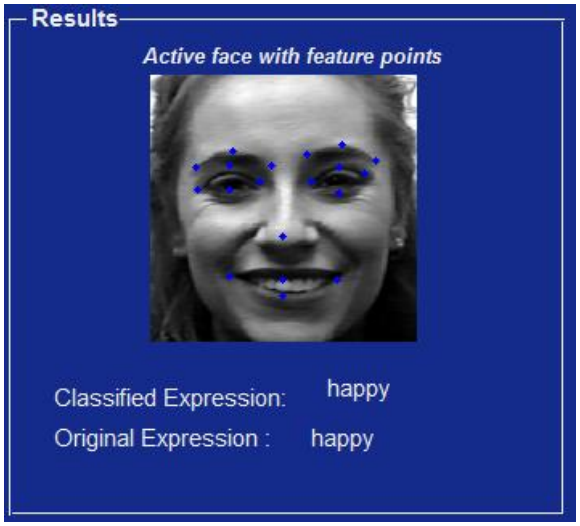


Figura 25. Ejemplo del resultado obtenido en el Módulo RESULTS.

- **TESTS**: Después de ejecutar los diferentes métodos seleccionados en el apartado **SETTINGS**, este módulo permite obtener los resultados de fallo y acierto de todas las imágenes de una base de datos para cada una de las emociones. Los resultados se almacenan en una matriz de confusión y esta se guarda en un documento Excel.
  El módulo **TESTS**, dispone también de un menú desplegable que permite seleccionar la obtención de resultados de las seis expresiones de forma conjunta, o solo evaluar entre cinco

expresiones. Al seleccionar el último caso, aparece otro menú desplegable en el cual se selecciona la expresión que se excluye de la realización de dichas pruebas, como se puede ver en la Figura 26.

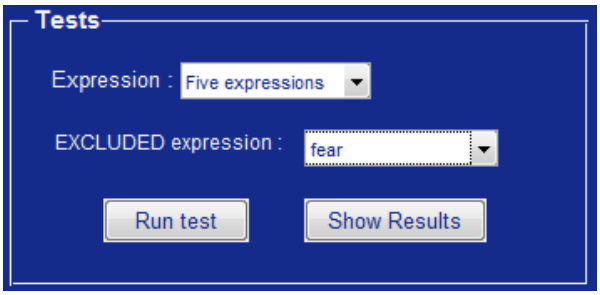
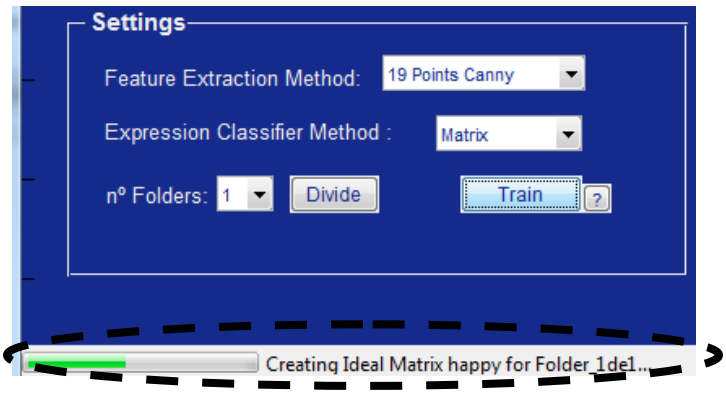


Figura 26. Módulo Test de la interfaz para comparar métodos de reconocimiento de emociones.

Comparado con la herramienta básica, se ha optimizado el proceso de pruebas al incluir la opción de poder evaluar de forma consecutiva y automática las seis expresiones definidas, sin tener que ir seleccionando una a una, y ya que se evaluaban resultados con sólo cinco expresiones, también se ha incluido esa opción. Gracias al botón *Run Test*, comienzan a procesarse cada una de las imágenes de la base de datos seleccionada de forma automática (recordar que las imágenes de la base de datos están nombradas de una forma estándar, lo que permite esta automatización), ejecutándose los métodos seleccionados en el apartado **SETTINGS**. Se procesan las expresiones de forma consecutiva para ir obteniendo los resultados de cada una de ellas y se van almacenando en una tabla Excel externa.

Para facilitar al usuario el análisis de los métodos seleccionados, se ha incluido también la opción de abrir dicha tabla Excel desde la interfaz (Botón *Show Results*). Para ello se ha procedido a nombrar los documentos Excel resultantes de forma similar al caso de los datos del entrenamiento. Se crea un directorio (de forma automática si no existe ya) con el nombre "Results" y se nombra al archivo obtenido con el nombre de la base de datos utilizada, el número de grupos en los que la base de datos está dividida (1 si no lo está), el método de extracción de características y el método de clasificación de la emoción, por ejemplo *Results _ck clasificada_10Folds_19 Points Canny_Matrix.*

Entre las utilidades añadidas a la interfaz se encuentran la implementación de una barra de estado en la parte inferior, que muestra el progreso de las distintas fases como la división de la base de datos, el entrenamiento del sistema o la ejecución de las pruebas y así poder estimar el tiempo que tarda en finalizar. Un ejemplo se muestra en la Figura 27.



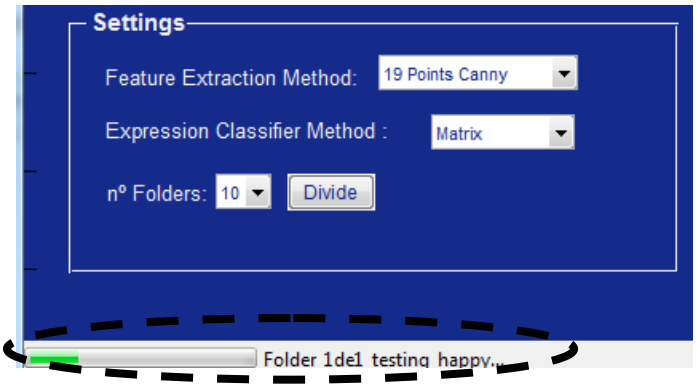


Figura 27. Ejemplos del progreso de una acción en la barra de estado.

Para facilitar el manejo a cualquier tipo de usuario, se han incluido en la herramienta botones de ayuda que explican la función de cada uno de los botones o módulos. Estas ayudas se activan/desactivan mediante la opción *Helps ON/OFF*. Además se puede seleccionar el idioma en el que se presenta la información (español o inglés). Se puede ver un ejemplo en la Figura 28 y Figura 29.

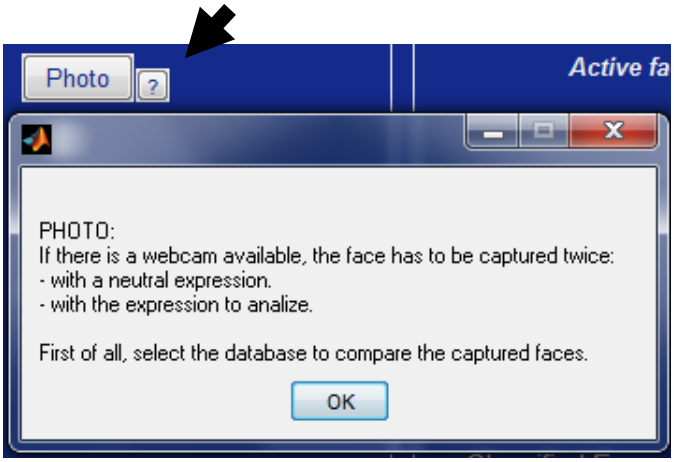


Figura 28. Ejemplo de pulsar la ayuda del botón Photo. Idioma seleccionado: inglés.

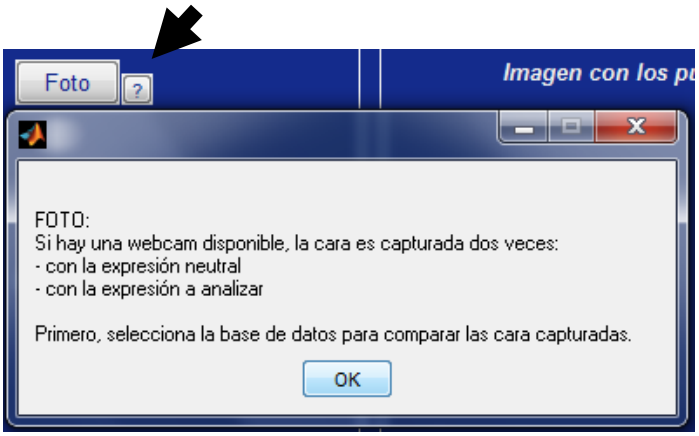


Figura 29. Ejemplo de pulsar la ayuda del botón Foto. Idioma seleccionado: español.

4. Descripción de los métodos implementados

4.1. Introducción

La realización de este proyecto tiene como origen el algoritmo propuesto en el proyecto de fin de máster [34], que consiste en detectar tres emociones: ira, sorpresa o tristeza. La cara y los ojos se localizan con ayuda del algoritmo *Viola & Jones* implementado en la *Computer Vision Toolbox* de Matlab. La boca y los ojos son segmentados mediante un umbral automático que está determinado por las características estadísticas del histograma de la imagen que se analiza. A continuación se realizan operaciones morfológicas para extraer la forma de las regiones boca y cejas. Las expresiones de ira, sorpresa o tristeza son clasificadas estudiando la inclinación de las cejas y la apertura de la boca.

La extracción de información basada sólo en el tamaño o inclinación de las regiones boca y cejas no permite detectar un gran número de expresiones, por lo que este sistema solo detecta tres expresiones, frente a las seis del resto de los sistemas existentes. Debido a esto, el método propuesto en este proyecto propone la detección de puntos de interés en el rostro, al igual que se propone en [35], para ampliar el rango de expresiones clasificadas y la veracidad de éstas.

En el grupo de trabajo de prácticas que realizó la herramienta básica, se implementó un método que localiza 12 puntos faciales. Para ello se realiza una división de la cara en regiones de interés y se aplican operaciones morfológicas a estas ventanas como se explica en el apartado 4.3. Posteriormente, se desarrolló un método para clasificar las seis expresiones consideradas como básicas, que consiste en comparar las variaciones de la posición y de los movimientos desde una expresión hasta su correspondiente cara neutra, con unas matrices representativas correspondientes a dichas seis expresiones (*anger, disgust, fear, happiness, sadness* y *surprise*) como se explica en el apartado 4.3.1.

Estos métodos presentan problemas que se han mejorado o solucionado en este proyecto. Algunas de las regiones en las que se dividía la cara y que se procesaban no estaban bien delimitadas o eran innecesarias, dando lugar a una errónea detección de puntos. El tamaño de la ventana que enmarca la cara requería un estudio en mayor profundidad para elegir el tamaño óptimo. Se carecía de una normalización de las caras evaluadas, que es necesario sobre todo para los casos en los que la cara esta inclinada, ya que permite colocarla en una posición estándar para el sistema.

Para mejorar estos problemas, así como el método de clasificación de la emoción desarrollado, se aumenta el número de puntos de interés detectados en la cara a 19 y se hacen ajustes en algunos parámetros de las operaciones morfológicas, en el tamaño de la ventana que contiene el rostro detectado y en el tamaño y número de regiones en las que se divide dicho rostro.

Por otro lado, resultados obtenidos revelaron que el método de comparación de las variaciones en la posición y movimientos con matrices representativas no es apropiado, principalmente porque depende de una imagen neutral, como se explica en el apartado 4.4.1. Como alternativa, se implementa un segundo método basado en el método FACS. Consiste en la obtención de unidades de acción a partir de los puntos característicos extraídos anteriormente para clasificar la emoción. En la Figura 30 se muestra el diagrama de bloques del sistema incluyendo ambos métodos de clasificación.

Además, se lleva a cabo una normalización de los puntos detectados en el rostro que se analiza mediante transformación afín, con el objetivo de hacer válidos los resultados como se explica al final del apartado 4.2.

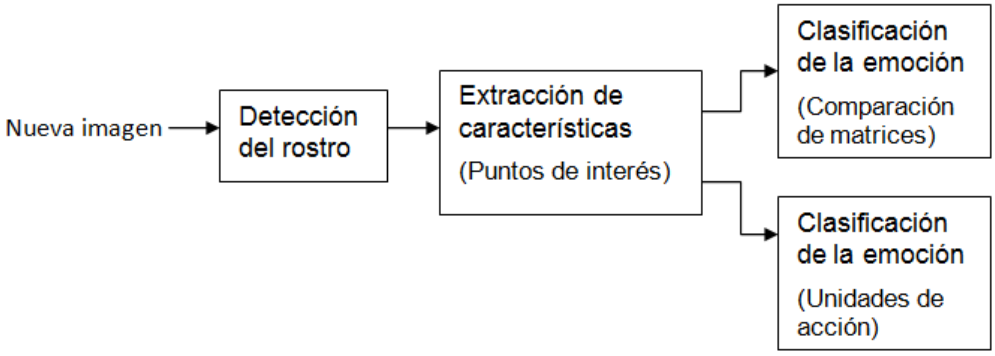


Figura 30. Diagrama de bloques del sistema de reconocimiento automático de emociones.

4.2. Detección de puntos característicos

El primer paso en todo sistema de reconocimiento de expresiones debe incluir un reconocimiento facial. Tras la obtención de una imagen, se localiza la cara (o caras) presente utilizando el algoritmo de *Viola & Jones* que está incluida en la *Computer Vision Toolbox* de MATLAB y cuyo resultado se muestra en la Figura 31. Para independizar el programa del tamaño de las caras encontradas, se recorta la zona de la imagen detectada como cara y a continuación se normaliza con un tamaño de NxN píxeles. Tras realizar pruebas con diferentes tamaños y evaluarlas (se explica este proceso en el apartado 5.1.1), en este proyecto el elegido es de 300x300, ya que con este tamaño se obtiene una buena precisión en la detección de los puntos, manteniendo un tiempo razonable en el procesado de la imagen. Si el tamaño es menor, dificulta la detección de los puntos, y si es mayor, el tiempo de procesado en la fase de extracción de características se incrementa considerablemente [36].

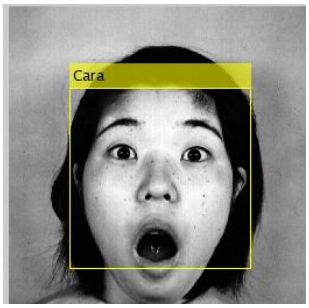


Figura 31. Ejemplo de la localización de la cara usando el algoritmo Viola&Jones.

Una vez detectada la cara del sujeto a analizar, se extrae y se procede a su pre-procesado. La segmentación realizada en el método implementado en el proyecto de fin de máster [34] para las regiones cejas y ojos, produce en los casos en los que la ceja y el pelo están juntos o muy cercanos, una unión de dichas regiones. Cuando posteriormente se eliminan las regiones de la cara que no son ceja, se elimina completamente dicha región como se muestra en la Figura 32, por lo tanto es imposible detectar sus puntos extremos.

Figura 32. Segmentación realizada en el proyecto de fin de máster [34]

Para solucionar esta problema, se divide la cara en regiones como expone [37], en el cual se tienen en cuenta las proporciones geométricas del rostro para localizar unas posiciones *x* e *y* en función de la anchura y altura de la cara, y así obtener las 12 regiones mostradas en la Figura 33. Utilizando las proporciones propuestas en [37], en algunas ocasiones las cejas o la boca se enmarcan incorrectamente en las regiones correspondientes (por ejemplo si las cejas están muy elevadas o la boca muy abierta), por lo que se han realizado determinados ajustes con el fin de obtener mejores resultados. Concretamente, se ha reducido el ancho de la región que contiene cada una de las cejas para limitar los problemas causados por la presencia de pelo y se ha aumentado la altura de la región de la boca para que en el caso de encontrarse abierta, pueda ser segmentada en su totalidad.

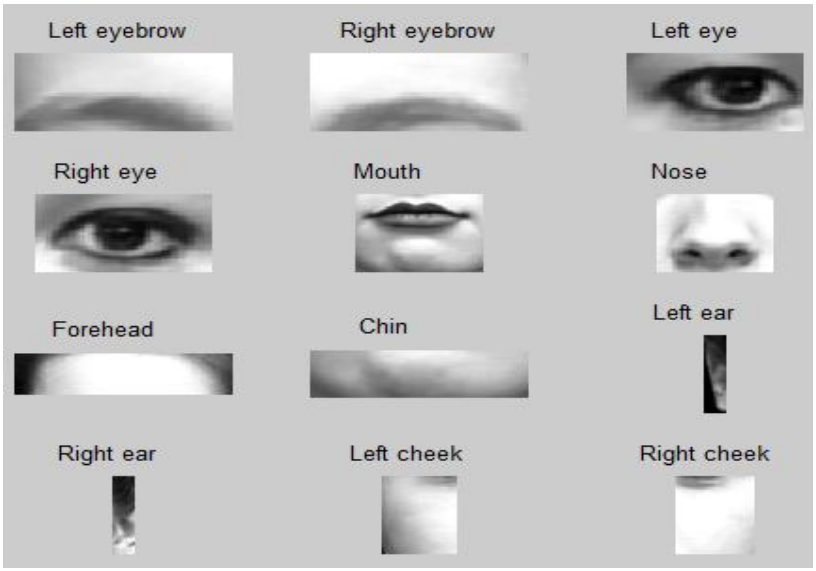


Figura 33.Regiones de la cara para extracción de características.

Debido a que en este proyecto sólo se va a trabajar sobre las regiones ceja, ojo, boca y nariz (para fines de normalización), que son las que contienen más información sobre las expresiones, se ha simplificado el número de regiones en las que se divide el rostro y así se aumenta la efectividad del sistema. En la Figura 34 se muestran las regiones que se van a utilizar en las siguientes fases de este proyecto.

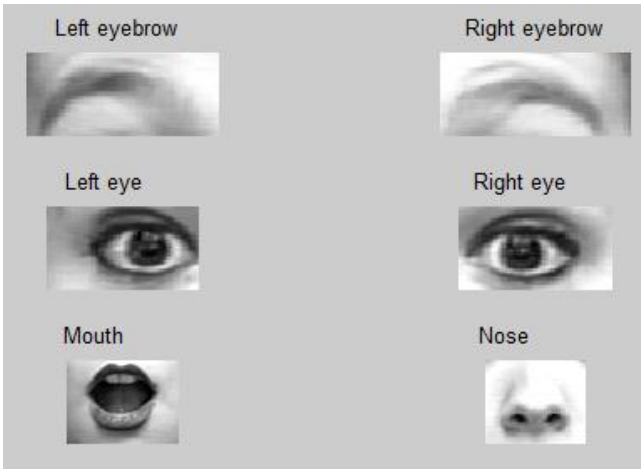
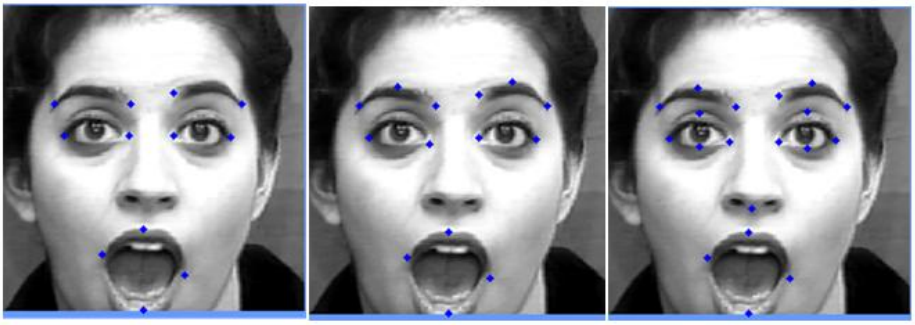


Figura 34. Regiones de la cara usadas para la extracción de características.

En cuanto a la detección de puntos característicos en la imagen, en [35] se eligen un total de 14 puntos distribuidos en las regiones de la cara que aportan más información en la detección de emociones (cejas, ojos y boca). En la versión original de la herramienta, fueron elegidos 12 puntos característicos. El punto omitido es el punto superior en cada una de las cejas, ya que la diferencia en los resultados que se obtienen es mínima y se ahorra tiempo de ejecución.

Debido a que la información aportada por la apertura de los ojos es relevante en la clasificación de emociones se añaden a los 12 puntos ya obtenidos, dos más por cada ojo, correspondientes a la parte superior e inferior del ojo. Debido a la normalización que se realiza posteriormente (explicada al final de este apartado) se requiere un punto característico en la nariz, por lo tanto el número de puntos detectados por la herramienta final aumenta a 19. En la Figura 35 se muestra la distribución de dichos puntos para los tres casos (12, 14 o 19 puntos).



(a) (b) (c)

Figura 35. Extracción de puntos característicos. (a) 12 puntos (b) 14 puntos (c) 19 puntos.

Para localizar cada uno de estos puntos, se han estudiado y probado diversos métodos de segmentación para encontrar el que proporcionase una mayor precisión en su detección. Por ejemplo, en el caso de la extracción de los puntos significativos de las cejas, el principal problema era que debido a la cercanía entre las cejas y el pelo, en numerosas ocasiones era detectado como una sola región. El método finalmente implementado consiste en tres fases. En la primera fase se transforma la imagen a escala de grises y se aumenta el contraste de la imagen, posteriormente se aplica el algoritmo de detección de bordes *Canny*, y por último se emplean diferentes operaciones morfológicas (como erosión o dilatación). El uso de operaciones morfológicas permite eliminar tanto el ruido como objetos que están presentes en la imagen, pero que son innecesarios o irrelevantes para la detección de los puntos de interés, y además se mejora la delimitación de las regiones a analizar (ceja, boca…). Para cada una de las regiones, algunas de las operaciones morfológicas utilizadas son diferentes, o se modifican los parámetros que determinan si una operación se aplica en mayor o menor medida.

Algunos de los métodos probados en [37] pero descartados son: utilización de máscaras elípticas, segmentación por agrupamiento de píxeles, segmentación por texturas y segmentación mediante el rango local de una imagen.

Una vez segmentada la imagen, los puntos significativos son obtenidos hallando los extremos de las regiones encontradas, que son identificados con la función *regionprops()* de Matlab. Concretamente, se detectan cuatro puntos para la boca, tres para cada ceja, cuatro para cada ojo y uno para la nariz. Para los puntos superior e inferior de la boca, se calcula el punto intermedio en el eje *x* entre los extremos laterales (coincidentes con la comisura de los labios) y se recorre el eje *y* hasta encontrar una variación entre dos píxeles consecutivos (ya que la imagen resultante de *Canny* es binaria). Para el caso de los puntos superior e inferior de los ojos y el superior de las cejas, el procedimiento seguido es el mismo. Para el caso de la nariz, se hallan los extremos inferiores de la región nariz y se calcula el punto medio. En la Figura 36 se muestra un ejemplo del proceso completo (excepto el resultado de la localización de los puntos que ya se mostró en la Figura 35).

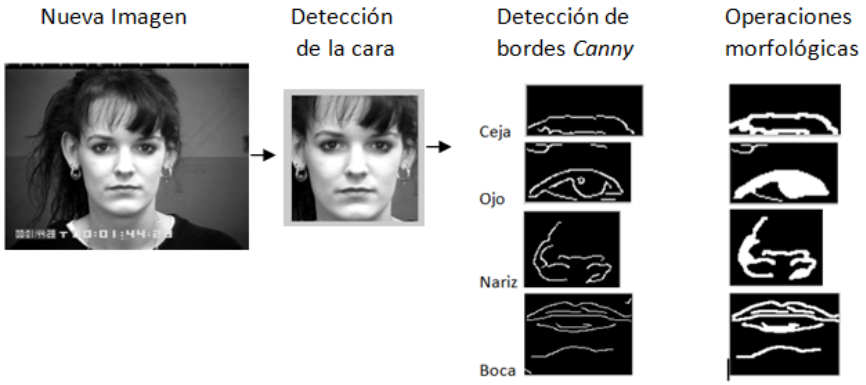


Figura 36.Proceso de extracción de características

Se ha observado que la detección de estos puntos debe ser mejorada, puesto que a veces coinciden con el borde del cuadrado que enmarca la región analizada (ojo, boca...). Para ello se han hecho algunas modificaciones a la hora de segmentar las diferentes regiones del rostro. En vez de dividir la cara en las regiones fijas mostradas en la Figura 34, se adaptan a la anatomía de la cara y a la localización de los ojos obtenida mediante el algoritmo *Viola & Jones* implementado en la *Computer Vision Toolbox* de Matlab que proporciona una región que enmarca los dos ojos. Por lo tanto, para obtener la región análoga a la división que se hacía anteriormente para cada ojo, se divide ese recuadro por la mitad. Se calculan los puntos en cada uno de los ojos y la posición de las regiones que encuadran las cejas, la nariz y la boca se calculan en función de la posición de los puntos obtenidos en los ojos, estableciendo para cada caso las proporciones adecuadas.

Para enmarcar las cejas se establece una región de igual altura que la obtenida para los ojos, y después de detectar los cuatro puntos de cada ojo se posicionan esas regiones de forma que el borde inferior coincide con el punto superior del ojo. Para su situación horizontal, en el caso de la ceja izquierda, se desplaza hacia ese lado 1/6 del punto externo del ojo izquierdo, y lo mismo para la ceja derecha. Para la región que enmarca la nariz, los extremos horizontales coinciden con los puntos medianos de los ojos, y la altura de dicha caja va desde el 50% hasta el 70% de la región que enmarca la cara. Por último, para enmarcar la boca, los extremos en horizontal coinciden con los puntos externos de los ojos, el borde superior coincide con el punto hallado previamente en la nariz y el borde inferior coincide con el borde inferior de la región que enmarca la cara. Se pueden ver estas proporciones en la Figura 37. Si el algoritmo de *Viola & Jones* no es capaz de detectar los ojos, se extraen las regiones que enmarcan los ojos de forma fija como se describió anteriormente, y el resto de zonas siguen el mismo comportamiento en base a la localización de los puntos en los ojos.

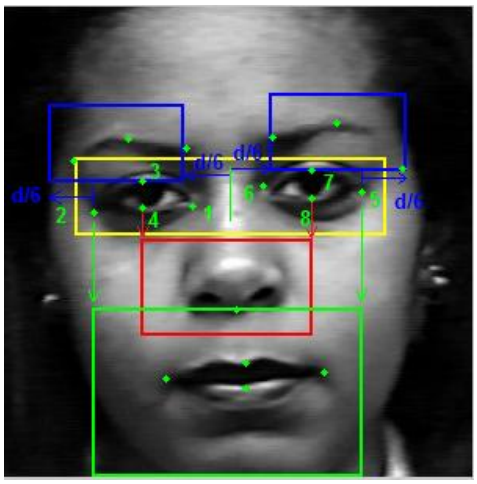


Figura 37. División en regiones de la cara dependientes de la localización de los ojos

Además, la segmentación realizada para este caso se realiza mediante un umbral de valores de grises en vez de utilizando la algoritmo de detección de bordes *Canny*. Por ello se denomina *Canny* a la primera forma descrita y *Thres* a la segunda.

Para evaluar la mejora al modificar la forma de dividir la cara en regiones, se han elegido algunas imágenes de la base de datos *CohnKanade+* y se han posicionado los puntos como serían correctos de forma manual. Después, se ha comparado esta localización con la obtenida con cada una de las formas descritas anteriormente y se muestran los resultados en la Tabla 6. En muchos casos, al emplear segmentación mediante valores de gris, la zona de la boca no es bien segmentada y por lo tanto los puntos en ella no son detectados correctamente, como se puede ver en la Tabla 6.

Tabla 6. Porcentajes de acierto en la detección de puntos

Detección de Puntos	**Canny (%)**	**Thres (%)**	**Detección de Puntos**	**Canny (%)**	**Thres (%)**
Esquina interior de la ceja izquierda	0,25	0,90	Esquina exterior del ojo izquierdo	0,50	0,75
Esquina interior de la ceja derecha	0,55	0,95	Esquina exterior del ojo derecho	0,40	0,75
Centro de la ceja izquierda	0,75	0,80	Esquina interior del ojo izquierdo	0,55	0,55
Centro de la ceja derecha	0,85	0,95	Esquina interior del ojo derecho	0,40	0,60
Esquina exterior de la ceja izquierda	0,70	0,85	Parte inferior del ojo izquierdo	0,50	0,90
Esquina exterior de la ceja derecha	0,60	0,70	Parte inferior del ojo derecho	0,50	0,95
Esquina izquierda de la boca	0,55	0,80	Parte superior del ojo izquierdo	0,35	0,85
Esquina derecha de la boca	0,55	0,55	Parte superior del ojo derecho	0,40	0,95
Parte inferior de la boca	0,40	0,30	Punto de la nariz	0,85	0,85
Parte superior de la boca	0,60	0,60			

Es importante normalizar la posición de dichos puntos para obtener unos resultados correctos en las pruebas que posteriormente se harán. La cara de los sujetos en las imágenes de una base de datos no suelen estar posicionadas de forma análoga, si no que unas están más cerca, otras más lejos e incluso están giradas, por ello es necesario realizar una normalización de los puntos hallados.

La normalización elegida ha sido utilizar la transformación afín que consiste en traslación, rotación y escalado cuya fórmula es

$$y = Ax + B. \tag{7}$$

En la Figura 38 se puede ver un ejemplo de aplicar la transformación afín a un triángulo equilátero, transformando el triángulo *abc* en el triángulo *a'b'c'*.

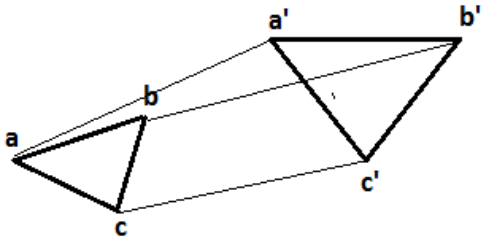


Figura 38. Ejemplo de transformación afín aplicada a un triángulo.

Para aplicar esta normalización al sistema, se ha establecido un triángulo base correspondiente a los puntos exteriores de los ojos y el punto de la nariz, con unos valores fijos. Ante una imagen de entrada, se extraen los 19 puntos característicos y de ellos se seleccionan los puntos exteriores de los ojos y el punto de la nariz, de forma que ya tenemos dos triángulos, el elegido como triángulo base y el triángulo de la imagen que está siendo analizada. A continuación, se aplica la función de transformación afín para determinar los parámetros de la transformación entre los valores de ambos triángulos. Finalmente, se transforman los 16 puntos restantes utilizando dichos parámetros. En la Figura 39 se puede observar un ejemplo de la transformación de los puntos iniciales (rombos) a los puntos normalizados (círculos).

Figura 39. Transformación de los puntos iniciales a los finales tras aplicar la transformación afín.

Finalmente, se enumeran los 19 puntos extraídos de forma intuitiva para el estudio y aplicación de métodos de clasificación de la emoción sencillos. El resultado final de la fase de extracción de características se puede observar en la Figura 40.

Figura 40. Distribución de los 19 puntos extraídos de la región cara.

4.3. Método de clasificación de la emoción

Tras obtener los puntos de interés es necesario crear un algoritmo que permita la clasificación de las emociones a partir de la información obtenida de ellos. En un grupo de trabajo se desarrolló un método que consiste en la comparación de las posiciones y variaciones de los puntos característicos entre la expresión neutra y la expresión analizada. Además de este método, durante el trabajo de este proyecto se añadió un segundo método que consiste en reconocer una serie de AUs y según las que sean identificadas, poder clasificar la emoción. Estos métodos se describen en detalle en los siguientes apartados.

4.3.1. Comparación de posiciones de puntos faciales

Este método consiste principalmente en comparar una matriz creada a partir de la imagen de entrada, de la cual se quiere clasificar la emoción, con las matrices representativas creadas en la fase de entrenamiento de cada una de las seis posibles emociones (*anger, disgust, fear, happiness, sadness* y *surprise*). Se trata de un método sencillo basado en la idea de que todas las imágenes correspondientes a una misma expresión tienen características (en este caso distribución de puntos) comunes y diferenciadas del resto de expresiones, de forma que al hacer un promedio de muchas imágenes, se obtiene una distribución representativa de puntos para cada expresión, guardadas en matrices representativas.

La matriz representativa *P* contiene para cada punto característico facial p_i la media (sobre todas las imágenes de entrenamiento) de la cantidad de movimiento en horizontal y vertical (Δh_i y Δv_i) entre la imagen neutra y la imagen de expresión, así como la media de las posiciones absolutas x_i y y_i, con lo cual su tamaño es de 19x4:

$$P = \begin{bmatrix} \Delta h_1 & \Delta v_1 & x_1 & y_1 \\ \Delta h_2 & \Delta v_2 & x_2 & y_2 \\ \dots & \dots & \dots & \dots \\ \Delta h_n & \Delta v_n & x_n & y_n \end{bmatrix}$$

Para clasificar la emoción de una nueva imagen de entrada en la herramienta, se calcula su matriz *P* y se compara con las matrices representativas de las seis expresiones básicas. Para obtener dichas matrices, es necesario entrenar al sistema. Normalmente, el primer paso del diseño de un sistema de clasificación consiste en utilizar un conjunto de datos, es este caso imágenes de una base de datos, que permita obtener la información necesaria para que el sistema funcione adecuadamente. A este conjunto de datos se le llama “conjunto de entrenamiento”. Posteriormente, se introducen nuevos datos al sistema, para que realice la clasificación y evaluar su efectividad.

Suponiendo una emoción X, para calcular la matriz representativa correspondiente, en primer lugar se obtienen las matrices *P* de todas las imágenes etiquetadas con dicha emoción X en una determinada base de datos de entrenamiento. En segundo lugar, se calcula el promedio de todas las matrices *P* obtenidas para esa emoción X y se obtiene así la matriz representativa buscada.

Una de las mejoras más significativas de este método en cuanto al trabajo realizado con anterioridad en un grupo de trabajo, ha sido optimizar la fase de entrenamiento. El proceso podía tardar más de una hora, en el caso de utilizarse 12 puntos característicos y aún más en el caso de 19 puntos, por lo que se han hecho modificaciones en el código en cuanto a la forma de crear las matrices representativas. El principal problema era que estas matrices se creaban emoción por emoción, de forma que las imágenes de la base de datos eran seis veces (una por cada emoción) leídas y procesadas. Con la nueva optimización, se analizan todas las imágenes correspondientes a un sujeto y se almacenan los resultados correspondientes a cada emoción, y una vez procesadas todas las imágenes de la base de datos (una sola vez) se crean las matrices representativas. El tiempo de la fase de entrenamiento ha sido mejorado hasta un máximo de 10-15 minutos.

El funcionamiento completo de este método de clasificación de la emoción se representa en el diagrama de bloques en la Figura 41. Cuando entra en el sistema una nueva imagen, se busca la imagen neutral correspondiente al sujeto analizado y se obtienen 19 puntos tanto de la imagen neutral como de la imagen a analizar. Una vez que se han detectado los puntos de interés de la región cara, se crea la matriz *P* y se calcula la diferencia (se resta) con cada una de las matrices representativas de cada expresión. Después, se evalúa con cuál de las seis matrices representativas tiene menor diferencia. Para ello se obtiene la media de todos los elementos de la matriz resultado de la resta, y la menor media será la expresión detectada.

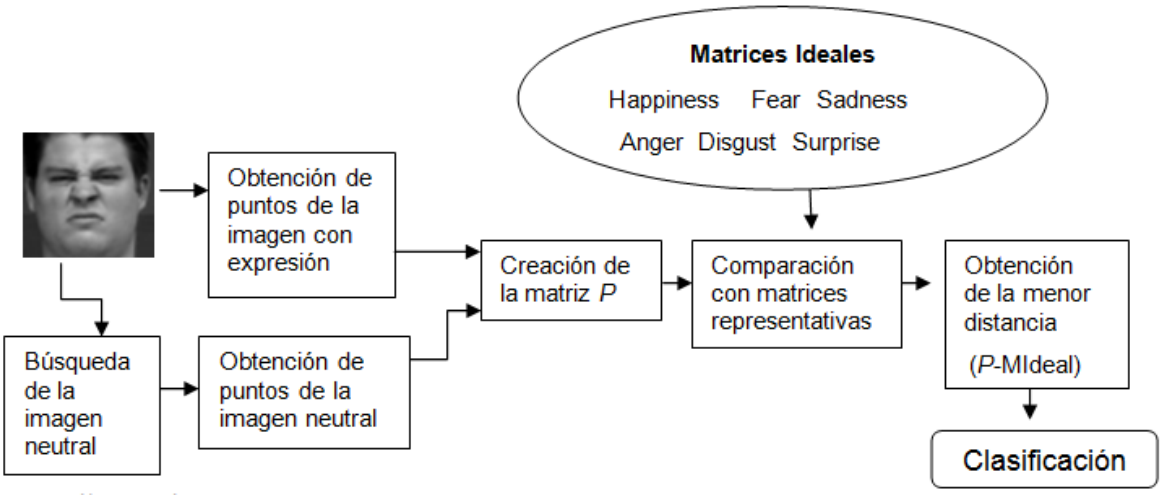


Figura 41. Diagrama de bloques del método comparación de matrices.

Este método de clasificación no sería apto para sistemas en tiempo real debido a la necesidad de tener una imagen neutral. Como mejora futura, el método que se puede implementar es el uso de árboles de decisión, que utilizan *Action Units* creadas a partir de puntos característicos del rostro, pero evitando la necesidad de utilizar la imagen neutral.

La forma de implementar el modelo de aprendizaje de árboles de decisión usando AUs consiste crear una matriz (de ceros y unos) con todas las imágenes presentes en la base de datos, indicando si una AU está presente o no ("1" o "0" respectivamente) en cada imagen y un vector de resultados que relacione cada imagen de la base de datos con la emoción a la que corresponde (cada emoción tendrá un número asociado). A partir de esta matriz se crea el árbol de decisión como se explica en el trabajo realizado en [26] .

En el siguiente apartado se explica un método que utiliza AUs (manteniendo el uso de la imagen neutral), con el objetivo de acercarse de una forma sencilla al uso de árboles de decisión.

4.3.2. Clasificación basada en Unidades de Acción

Este método consiste principalmente en seleccionar una serie de AUs basadas en los puntos previamente obtenidos y en establecer cuáles son más relevantes para cada una de las seis posibles emociones. Cuando una nueva imagen entre en el sistema, se detecta cuáles de las AUs de interés están presentes en ella y utilizando esa relevancia se procede a clasificar la emoción.

En la Tabla 7 se muestra la denominación de las AUs utilizadas en la implementación de este método de clasificación que corresponden con AUs del método FACS que se comentó en el apartado 2. Algunas de ellas que serían relevantes para la clasificación de emociones no se han implementado debido a la dificultad de asociarlas con los puntos detectados o la dificultad de percibir su relevancia. Un ejemplo de ello sería la AU9 definida en FACS que corresponde con arrugar la nariz.

Se han definido dos nuevas AUs, una de ellas consiste en bajar el interior de las cejas y la segunda consiste en elevar las cejas, y se han denominado respectivamente AU3 y AU8, ya que no estaban

definidas en FACS. La primera de ellas se ha decidido crearla debido a que, como se verá en los resultados obtenidos en el apartado 5, la expresión *anger* es comúnmente confundida con *sadness* y *fear* debido a que tienen características (en este caso distribución de puntos) similares y esta AU ayuda a diferenciar de ellas la emoción *anger* (recordar que se está implementando un método muy sencillo de clasificación en el que es necesario una mayor diferenciación entre emociones). Para la segunda, se ha realizado una prueba que obtiene el porcentaje de la presencia de esta AU en las imágenes con emoción *surprise* de la base de datos de estudio y su presencia es del 90%. Junto con la AU27 definida en FACS, es la más característica de esta emoción, y debido a que identificar una emoción sólo por una AU (AU27) no es adecuado (ya que si se produce un error en su detección la clasificación ya sería errónea) se ha decidido incluirla.

Tabla 7. Implementación de AUs para programa Matlab

	AU (FACS)	**Descripción de la AU**	**Puntos implicados (de 19)**
Parte superior	1	Interior de las cejas elevado	3,4
	2	Exterior de las cejas elevado	1,6
	3*	Interior de las cejas bajado	3,4
	4	Cejas bajadas	2,5
	8*	Cejas subidas: AU1+AU2	2,5
Parte inferior	11	Nasolabial: Labio superior elevado> 10 pixeles	16
	12	Comisuras de la boca estiradas y elevadas	15,17
	15	Comisuras de los labios hacia abajo	15,17
	16	Labio inferior hacia abajo (0-5 pixeles)	18
	20	Labios estrechados y estirados en horizontal (comisuras hacia los lados <- ->)	15,17
	23	Labios apretados y encogidos (comisuras hacia dentro -> <-)	15,17
	24	Labios comprimidos (superior e inferior)	16,18
	26	Labio inferior hacia abajo (15-30 pixeles)	18
	27	Labio inferior hacia abajo (>30 pixeles)	18

**Añadidas por su relevancia para algunas expresiones como surprise (cejas elevadas).*

Además, existen combinaciones de AUs definidas en FACS que no son apropiadas, por ejemplo la AU4 (cejas bajadas) con AU5 (párpado superior elevado) como se muestra en la Figura 42, por ello se ha descartado el uso de este tipo de combinaciones utilizándose solo las indicadas en la Tabla 8, en la que se presentan las AUs relevantes para cada emoción [17].

AU4 AU5

Figura 42. Ejemplo de una combinación de AUs inapropiada

Para clasificar la emoción, el primer paso se realiza de la misma forma que se definió en el apartado 4.3.1, cuando entra en el sistema una nueva imagen, se busca la imagen neutral correspondiente al sujeto analizado y se obtienen los 19 puntos tanto de la imagen neutral como de la imagen a analizar. A continuación, comparando las posiciones de los puntos entre la imagen neutral y la analizada, y

mediante la comparación de condiciones, se van identificando las diferentes Unidades de Acción. Concretamente, se crea una matriz de diferencias entre los 19 puntos extraídos en la imagen neutral y en la imagen analizada, es decir, la diferencia entre las matrices *P* de ambas imágenes, y para decidir si una AU está activa o no, se evalúan las variaciones en dicha matriz correspondientes a los puntos implicados en la AU, mostrados en la Tabla 7.(Véase la distribución de los 19 puntos en la Figura 40). Por ejemplo, la AU27 definida en el método FACS consiste en abrir considerablemente la boca, por lo que se evalúa el punto del labio inferior y se establece que si su posición en la imagen analizada está desplazada hacia abajo con respecto a ese punto en la imagen neutral, se activa la AU27.

Una vez que se decide que AUs están activas y cuáles no, se procede a clasificar la expresión. Para ello se utiliza la información contenida en la Tabla 8 y mediante seis contadores (uno por cada emoción) se evalúa para que emoción se han detectado más AUs. Si una AU que se considera relevante para una emoción concreta está activada, se procede a incrementar en 1 el valor de dicho contador. Finalmente, se evalúa cuál de los 6 contadores es mayor y esa será la expresión detectada.

Si se produce un empate entre el valor de los contadores de emociones diferentes, se procede a evaluar AUs individuales, por ejemplo si hay empate entre *surprise* y otra expresión, se evalúa la boca, concretamente se comprueba si la AU27 está presente o no en la imagen, de forma que si está presente, la expresión sería *surprise*. Las posibilidades de que exista otro empate al evaluar AUs de forma individual es escasa ya que se eligen AUs que sean muy restrictivas para cada emoción, pero es posible que ocurra si ha habido algún previo error en la detección de puntos o AUs, por ello para cada caso de empate entre los contadores de dos emociones y la primera AU individual elegida, se evalúa otra AU.

Tabla 8. Selección de AUs para cada expresión

Emotion	**AUs (FACS)**
Anger	3**,4,15,16,20,23,24
Disgust	4,11,16,23
Fear	1,8*,20,26
Happy	12,20,26
Sadness	1,15
Surprise	8*,27

**Nuevas AUs Definidas en la Tabla 7: cejas elevadas*
*** Nuevas AUs Definidas en la Tabla 7 Interior cejas bajado*

El uso de estos contadores es una forma muy sencilla de aplicar *Actions Units*, y por lo tanto no la más adecuada (lo mejor es aplicar árboles de decisión). La ventaja de este método es que su ejecución es rápida y para su implementación ha sido necesario hacer un estudio en profundidad sobre *Action Units*. Concretamente, averiguar qué AUs son más relevantes a la hora de detectar una emoción, la necesidad de añadir AUs nuevas a las descritas en FACS o descartar el uso de algunas combinaciones , y además permite comprobar el resultado de su utilización para futuros métodos de clasificación de la emoción, o futuras mejoras.

El principal inconveniente es consecuencia de los puntos detectados, ya que al ser un número reducido, hay AUs que no pueden ser detectadas y por lo tanto se limita la efectividad de los métodos de clasificación de la emoción que requieran su uso. Otro defecto es que hay casos en los que se produce un empate entre los contadores de diferentes emociones y hay que recurrir a una nueva evaluación de AUs. Esto aumenta la dependencia de su correcta detección que deriva en la necesidad de una extracción de características muy precisa (en este caso puntos), ya que un fallo en esta fase podría cambiar por completo la emoción clasificada.

En la Figura 43 se muestra el diagrama de bloques del método de clasificación de la emoción usando Unidades de Acción, con un ejemplo para la expresión *disgust*.

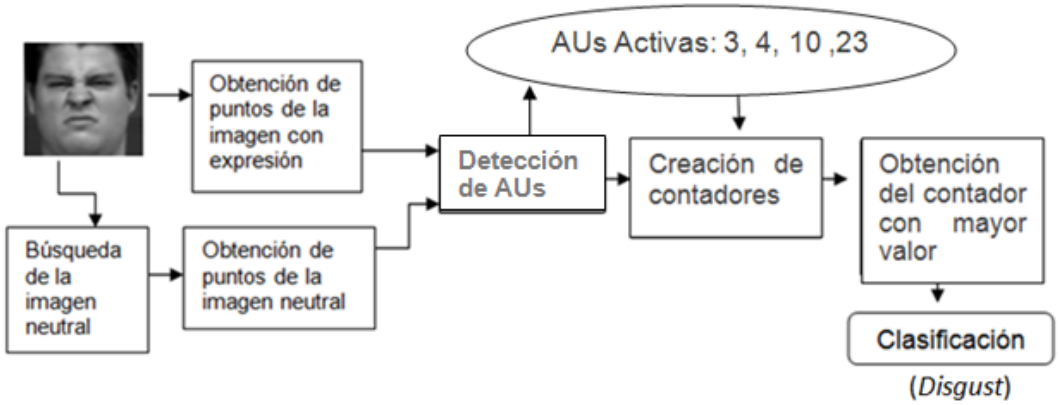


Figura 43. Diagrama de bloques del método usando Unidades de Acción.

5. Resultados

Para comprobar el funcionamiento de la herramienta y de los distintos métodos implementados, se realizan pruebas con las bases de datos *CohnKanade+* [30], JAFFE [31] e imágenes capturadas mediante la cámara del ordenador. Estas pruebas son realizadas de manera automática mediante el uso de la interfaz gráfica creada, que permite seleccionar las imágenes cuya expresión se quiere analizar, hace los cálculos necesarios, y guarda una tabla Excel con los resultados de acierto y fallo.

Las expresiones para las que se van a realizar las pruebas, y para las que se han creado los métodos de clasificación de la emoción explicados en apartados anteriores, son las seis más analizadas en la mayoría de los trabajos existentes: *anger*, *disgust*, *fear*, *happiness, surprise* y *sadness*.

Para determinar la mejor solución en cuanto a precisión y tiempo de procesado, se evalúan los resultados al elegir distintos tamaños de resolución del rostro, extraer 12 o 19 puntos característicos y las dos formas de segmentación propuestas en el apartado 4.2.

Para el método de comparación de matrices se entrena al sistema creando matrices representativas para cada base de datos. Se han hecho pruebas de dos formas. En una de ellas, se entrena el sistema utilizando la base de datos completa y se prueba sobre las mismas imágenes, y en la otra, se divide la base de datos en *N* fragmentos, de manera que se seleccionan $\frac{n-1}{n}$ de las imágenes para entrenamiento y se prueba sobre el $\frac{1}{n}$ restante. Se repite el proceso N veces, modificando los fragmentos seleccionados para entrenamiento y prueba.

Para el método basado en Unidades de Acción no es necesario un entrenamiento previo, aunque hay que tener en cuenta el trabajo anteriormente realizado en el análisis de Unidades de Acción relevantes para cada emoción, y del cual dependen directamente los resultados obtenidos.

Además, se realizan pruebas tanto usando normalización en la extracción de puntos, como sin normalización para evaluar su efecto.

La forma de presentar los resultados a lo largo de este apartado consiste en mostrar tablas de confusión. Estas tablas permiten conocer el porcentaje de acierto para la expresión analizada, y en caso de fallo, conocer con qué expresión o expresiones se confunde y poder aplicar medidas correctoras.

5.1. Método basado en comparación de desplazamiento

La base de datos *CohnKanade+* ha sido la más usada por otros investigadores en el área de reconocimiento de emociones como se comentó en el apartado 2.5 y por ello sobre esta base de datos se van a realizar las pruebas que evalúan distintas características del sistema, como la mejor resolución de la cara y el número de puntos extraídos en la fase de extracción de características. Dicha base de datos consta de 105 sujetos, para todos ellos hay una imagen con expresión neutral, pero la mayoría no tienen imágenes correspondientes a las seis expresiones básicas, sino que sólo contienen algunas de ellas.

5.1.1. Resolución del rostro y distinto número de puntos característicos

Para determinar el mejor tamaño del rostro es importante tener en cuenta la resolución con la que las caras de la base de datos de estudio son detectadas. Dependiendo de la cercanía del sujeto a la cámara el tamaño de la región que enmarca al rostro detectado será mayor o menor. Para igualar el tamaño del rostro de todas las imágenes de estudio se realiza una interpolación bilineal. Para la base de datos CK+ el tamaño medio de las caras en las imágenes es de 300x300 pixeles, por lo tanto se han hecho pruebas con diferentes resoluciones alrededor de este valor para elegir el más apropiado. En la Tabla 9 se pueden ver

los resultados de transformar el tamaño de cara a 200x200 pixeles y en la Tabla 10 a 300x300 pixeles. En el caso de seleccionar un tamaño de 200x200 pixeles (o mayores a 300x300 pixeles) la precisión en la detección de las expresiones es afectada negativamente, sobre todo en las expresiones de *sadness* y *disgust*, en las que se obtienen porcentajes menores al 20% y al 40%, respectivamente. Por otro lado, al elegir como resolución un tamaño de 300x300 pixeles, que es el tamaño medio de las caras en las imágenes de esta base de datos, se mejora la precisión en dichas expresiones en un 40-50%. Cuanto más difiere la resolución elegida de la resolución media, la distancia entre los puntos extraídos es menos precisa y por lo tanto su importancia en la media final que se hace cuando se comparan la imagen analizada con cada una de las matrices representativas, es menor.

Tabla 9. Pruebas en CohnKanade+ con tamaño de la cara 200x200 y 19 puntos.

Expresión Analizada -->	*Anger*	*Disgust*	*Fear*	*Happiness*	*Sadness*	*Surprise*
Expresión Detectada						
Anger	**0,64**	0,14	0,13	0,07	0,11	0,05
Disgust	0,02	**0,40**	0,13	0,00	0,07	0,01
Fear	0,07	0,0	**0,42**	0,03	0,44	0,03
Happiness	0,27	0,47	0,25	**0,88**	0,19	0,01
Sadness	0,00	0,00	0,04	0,02	**0,15**	0,12
Surprise	0,00	0,00	0,04	0,00	0,04	**0,78**
Precisión	**0,54**					

Tabla 10. Pruebas en CohnKanade+ con tamaño de la cara 300x300 y 19 puntos.

Expresión Analizada -->	*Anger*	*Disgust*	*Fear*	*Happiness*	*Sadness*	*Surprise*
Expresión Detectada						
Anger	**0,51**	0,05	0,08	0,04	0,11	0,01
Disgust	0,07	**0,82**	0,04	0,04	0,00	0,01
Fear	0,20	0,02	**0,42**	0,07	0,15	0,03
Happiness	0,16	0,10	0,13	**0,84**	0,04	0,00
Sadness	0,04	0,00	0,30	0,00	**0,67**	0,04
Surprise	0,02	0,00	0,04	0,00	0,04	**0,91**
Precisión	**0,70**					

Para comparar la eficacia del sistema en relación al número de puntos característicos extraídos, se muestran en la Tabla 11 los resultados obtenidos al extraer 12 puntos (número de puntos del método base), mientras que en la Tabla 10 se mostraban los resultados con el método mejorado a 19 puntos (se añade el punto intermedio de la ceja, los puntos superior e inferior de los ojos, y el punto de la nariz).

Es importante mencionar que en la clasificación de expresiones, se obtienen mejores resultados en las emociones *happiness*, *surprise* y *disgust*, debido a que tienen características bastante diferentes entre sí, y no se suelen confundir con el resto. Todo lo contrario ocurre con *sadness*, *fear* y *anger*, ya que suelen tener características comunes o poco diferenciadas.

Al aumentar el número de puntos de 12 a 19, la precisión en las expresiones que más tienden a confundirse (*sadness*, *anger* y *fear*) aumentan entre un 3% y un 18%. A partir de ahora las pruebas se realizarán con 19 puntos.

Tabla 11. Pruebas en CohnKanade+ con tamaño de la cara 300x300 y 12 puntos.

Expresión Analizada -->	*Anger*	*Disgust*	*Fear*	*Happiness*	*Sadness*	*Surprise*
Expresión Detectada						
Anger	**0,49**	0,05	0,13	0,04	0,19	0,01
Disgust	0,11	**0,76**	0,04	0,03	0,11	0,01
Fear	0,18	0,02	**0,46**	0,04	0,15	0,05
Happiness	0,11	0,12	0,25	**0,88**	0,04	0,00
Sadness	0,02	0,00	0,013	0,0	**0,48**	0,05
Surprise	0,09	0,05	0,00	0,00	0,04	**0,88**
Precisión	**0,66**					

5.1.2. Expresiones analizadas

Como se ha mencionado en el apartado anterior, las expresiones más difíciles de diferenciar son *anger*, *sadness* y *fear*. Además, en este sistema la dificultad se incrementa debido a que en la base de datos *CohnKanade+* hay muchas imágenes en las que es difícil diferenciar visualmente si se trata de la expresión *anger*, *sadness* o *fear*. Un ejemplo de ello se muestra en la Figura 44. Se observa cómo estas imágenes, que en la base de datos están clasificadas como *anger*, visualmente podrían ser identificadas como *sadness* o *disgust*. Esto afecta también a la clasificación de la emoción realizada por los algoritmos implementados en el sistema, y por lo tanto los porcentajes de acierto en estas expresiones se reducen.

Teniendo en cuenta esta característica de la base de datos *CohnKanade+*, se realizan pruebas descartando una de estas expresiones (*anger*, *sadness* y *fear*) y se muestran los resultados para evaluar los porcentajes de acierto con cinco expresiones. En la Tabla 12 se ha descartado la expresión *sadness*, en la Tabla 13 la expresión *anger* y en la Tabla 14 la expresión *fear*.

Figura 44. Imágenes de la base de datos CohnKanade+ visualmente difíciles de clasificar [36].

Tabla 12. Pruebas en CohnKanade+ con todas las expresiones excepto Sadness

Expresión Analizada -->	*Anger*	*Disgust*	*Fear*	*Happiness*	*Sadness*	*Surprise*
Expresión Detectada						
Anger	**0,53**	0,05	0,08	0,04		0,01
Disgust	0,07	**0,83**	0,04	0,04		0,01
Fear	0,22	0,02	**0,63**	0,07		0,05
Happiness	0,15	0,01	0,17	**0,84**		0,00
Sadness						
Surprise	0,02	0,00	0,08	0,00		**0,93**
Precisión	**0,75**					

Tabla 13. Pruebas en CohnKanade+ con todas las expresiones excepto Anger

Expresión Analizada -->	*Anger*	*Disgust*	*Fear*	*Happiness*	*Sadness*	*Surprise*
Expresión Detectada						
Anger						
Fear		0,05	**0,5**	0,10	0,22	0,04
Disgust		**0,83**	0,04	0,04	0,00	0,01
Happiness		0,10	0,13	**0,85**	0,04	0,00
Sadness		0,00	0,29	0,00	**0,70**	0,04
Surprise		0,02	0,04	0,00	0,04	**0,91**
Precisión	**0,76**					

Tabla 14. Pruebas en CohnKanade+ con todas las expresiones excepto Fear

Expresión Analizada -->	*Anger*	*Disgust*	*Fear*	*Happiness*	*Sadness*	*Surprise*
Expresión Detectada						
Anger	**0,62**	0,07		0,09	0,19	0,03
Disgust	0,07	**0,83**		0,04	00,0	0,01
Fear						
Happiness	0,18	0,10		**0,87**	0,04	0,00
Sadness	0,11	00,0		0,00	**0,74**	0,04
Surprise	0,02	0,00		0,00	0,04	**0,93**
Precisión	**0,80**					

Si se comparan los resultados obtenidos al analizar seis expresiones (Tabla 10) con los obtenidos al analizar cinco expresiones (Tabla 12, Tabla 13 y Tabla 14) se observa que el porcentaje de acierto para las expresiones *happiness*, *surprise* y *disgust* prácticamente no varía. Sin embargo, se aprecian mejoras significativas en el caso de *sadness* (hasta un 10%), *anger* (hasta un 10%) y *fear* (hasta un 12%), y además se obtienen porcentajes para todas las expresiones analizadas por encima del 50%, e incluso del 60% en el caso de descartar la expresión *fear* (Tabla 14).

La significativa mejora en el caso de no considerar la expresión *fear*, se debe a que existen numerosas diferencias en cómo los sujetos expresan dicha emoción. Esto hace que para el análisis de muchos sujetos, la matriz obtenida difiera de la matriz representativa creada para la emoción *fear*, y la expresión puede ser confundida con *happiness*, *sadness* o *anger*. En la Figura 45 se muestran algunos ejemplos de esa diferencia en la forma de expresar la emoción *fear* dependiendo del sujeto. En el caso (a) se confunde con *sadness* debido a la boca (comisuras hacia abajo), en (b) se confunde con *anger* por las cejas, en el caso (c) con *happiness* por la forma en la apertura de la boca.

Figura 45. Imágenes de la base de datos CohnKanade+ clasificadas como fear

5.1.3. Independencia entre imágenes de entrenamiento y prueba

Como se ha mencionado en la introducción del apartado 5, se realizan pruebas de dos formas. Una de ellas, que ha sido la utilizada en las tablas mostradas hasta el momento, consiste en entrenar el sistema con la base de datos completa y realizar las pruebas sobre las mismas imágenes. La segunda forma, consiste en dividir la base de datos en *N* fragmentos de manera que las imágenes con las que se realicen pruebas sean distintas de las imágenes empleadas en el entrenamiento del sistema.

En la mayoría de los trabajos realizados por los investigadores en el área de reconocimiento de emociones, la división de la base de datos de estudio se realiza de tal forma que el sistema sea entrenado con un 90% de las imágenes y con el 10% restante se pruebe. Por ello, para la base de datos *CohnKanade+,* el número de fragmentos elegido es de 10, de manera que se entrena el sistema con 9/10 de las imágenes totales (90%) y se prueba con el 1/10 restante (10%). Esto se repite 10 veces, cambiando los fragmentos de entrenamiento/prueba, y se obtiene el resultado como promedio sobre las 10 pruebas. Cada fragmento de imágenes está almacenado en un directorio diferente, que se crean de forma automática por la herramienta distribuyendo las imágenes de la base de datos de forma aleatoria. Cada directorio contiene 61 imágenes (sin incluir las imágenes neutrales) y no hay uniformidad en el número de imágenes correspondientes a cada emoción, es decir, en un directorio puede haber 25 imágenes para una emoción y tan solo cinco para otra.

Si se prueba el sistema sobre las mismas imágenes con las que fue entrenado, el porcentaje de aciertos será mayor que si se ejecuta sobre imágenes nuevas e independientes a las que se usaron en el entrenamiento. Esto se comprueba al comparar los resultados mostrados en la Tabla 15, que son obtenidos al entrenar al sistema de forma independiente, con los porcentajes de acierto obtenidos en la Tabla 10, caso en el que el sistema es entrenado de forma dependiente. Se observa que la precisión total disminuye y que sobretodo se ve drásticamente reducida en el caso de la expresión *sadness*. Es importante tener en cuenta que en la base de datos *CohnKanade+,* el número de imágenes de las que se dispone para cada expresión es diferente, por ejemplo para el caso de *surprise* existen 81 imágenes, mientras que para el caso de la expresión *sadness*, hay tan solo 27. Debido al número reducido de imágenes, el entrenamiento será menos preciso y por ello los porcentajes de acierto si se independizan las imágenes de prueba y entrenamiento, son menores.

Tabla 15. Pruebas en CohnKanade+ independizando las imágenes de entrenamiento de las de prueba

Expresión Analizada→	*Anger*	*Disgust*	*Fear*	*Happiness*	*Sadness*	*Surprise*
Expresión Detectada						
Anger	**0,42**	0,06	0,10	0,03	0,15	0,01
Disgust	0,05	**0,63**	0,05	0,06	0,13	0,06
Fear	0,34	0,01	**0,42**	0,09	0,37	0,02
Happiness	0,17	0,28	0,04	**0,83**	0,03	0,00
Sadness	0,03	0,00	0,19	0,00	**0,32**	0,18
Surprise	0,00	0,01	0,00	0,00	0,00	**0,73**
Precisión	**0,56**					

Como se comentó en el apartado 4.2, el método de extracción de características se mejoró para evitar que algunos de los puntos fuesen erróneamente detectados a raíz de la previa separación de regiones de la cara. Por ello se realiza una prueba comparativa entre los dos métodos. En las tablas mostradas hasta ahora, la detección de puntos era realizada por el método denominado en el apartado 4.2 como *Canny* y en la Tabla 16 se muestran los resultados de aplicar el método denominado como *Thres* (con independencia entre las imágenes de entrenamiento y prueba).

Al implementar el método *Thres*, la detección de puntos en los ojos y las cejas es más correcta que en el caso de *Canny*, pero su detección en la boca sigue siendo deficiente. En general, los resultados obtenidos en la clasificación de la emoción son similares para ambos casos y son resultados que se pueden mejorar mediante el uso de métodos de clasificación más avanzados como pueden ser SVM o redes neuronales.

Tabla 16. Pruebas en CohnKanade+ utilizando mejoras en la detección de puntos

Expresión Analizada→	*Anger*	*Disgust*	*Fear*	*Happiness*	*Sadness*	*Surprise*
Expresión Detectada						
Anger	**0,53**	0,17	0,15	0,00	0,10	0,04
Disgust	0,06	**0,66**	0,03	0,03	0,07	0,00
Fear	0,37	0,03	**0,22**	0,03	0,64	0,00
Happines	0,04	0,14	0,12	**0,93**	0,00	0,02
Sadness	0,00	0,00	0,27	0,00	**0,14**	0,20
Surprise	0,00	0,00	0,03	0,02	0,06	**0,74**
Precisión	**0,53**					

5.1.4. Normalización

En el método de comparación de desplazamiento, se establece una distribución representativa de los puntos para cada una de las expresiones. Como se explicó en el apartado 4.3.1, suponiendo una emoción X, para calcular la distribución representativa (matriz representativa) correspondiente a esa emoción, en primer lugar se obtienen las distribuciones de puntos de todas las imágenes etiquetadas con dicha emoción X en la base de datos. A continuación, para cada uno de los 19 puntos, se realiza una media de su valor en todas las imágenes etiquetadas con dicha emoción X. Por lo tanto, los desplazamientos Δx_i y Δy_i y las coordenadas x_i y y_i de los puntos en la matriz representativa P no se forman solamente por movimiento de los puntos, sino que están afectados por desplazamientos de la posición de la propia cara. Para resolver este problema, se procede a la normalización, mediante la transformación de los puntos a un espacio normalizado, como se explicó en el apartado 4.2.

Al aplicar normalización a las imágenes se pretende que cada uno de los puntos análogos en cada una de las caras se junten (por ejemplo las caras anchas se estrechan). Los puntos de cualquier cara se transforman a una posición normalizada (recordar que se toma como base de la normalización un triángulo formado por los puntos externos en los ojos y el punto en la nariz) y de esta forma las desviaciones deberían disminuir

En la Tabla 17 se pueden ver los resultados tras aplicar normalización cuando se extraen 19 puntos característicos y se aprecia que la precisión total, comparada con la Tabla 15 (resultados obtenidos sin aplicar normalización), se ve reducida en un 26%.

Tabla 17. Pruebas en CK+ con normalización de la posición de 19 puntos.

Expresión Analizada→	*Anger*	*Disgust*	*Fear*	*Happiness*	*Sadness*	*Surprise*
Expresión Detectada						
Anger	**0,58**	0,30	0,30	0,07	0,33	0,10
Disgust	0,07	**0,43**	0,00	0,09	0,04	0,16
Fear	0,11	0,05	**0,17**	0,03	0,22	0,00
Happiness	0,16	0,10	0,25	**0,74**	0,15	0,03
Sadness	0,09	0,12	0,29	0,07	**0,26**	0,31
Surprise	0,00	0,00	0,00	0,00	0,00	**0,41**
Precisión	**0,43**					

5.1.5. Base de datos JAFFE

Otra base de datos normalmente utilizada por los investigadores es JAFFE. Consta de 10 sujetos japoneses femeninos, y cada uno de ellos representa las seis emociones básicas entre uno y cuatro grados de intensidad.

Para esta base de datos ya no se muestran los resultados para determinar el mejor tamaño de la cara, ni para evaluar la aplicación de normalización, puesto que el número de sujetos es demasiado reducido y son sólo femeninos, a pesar de ser comúnmente utilizada por los investigadores, sus resultados para evaluar características generales no es relevante.

Sin embargo, como esta es la base de datos utilizada en el proyecto de fin de máster del que partía este proyecto [34], sí se comparan los resultados obtenidos al extraer 12 o 19 puntos de interés, y se muestran en la Tabla 18. Se puede observar que en todas las expresiones la precisión mejora, excepto para el caso de la expresión *fear*, que al igual que ocurre en la base de datos CK+ y como se puede observar en la Figura 46, existen grandes diferencias en la forma en que cada sujeto expresa dicha emoción, e incluso en algunas de las imágenes, es visualmente difícil identificar que se trata de la expresión *fear*.

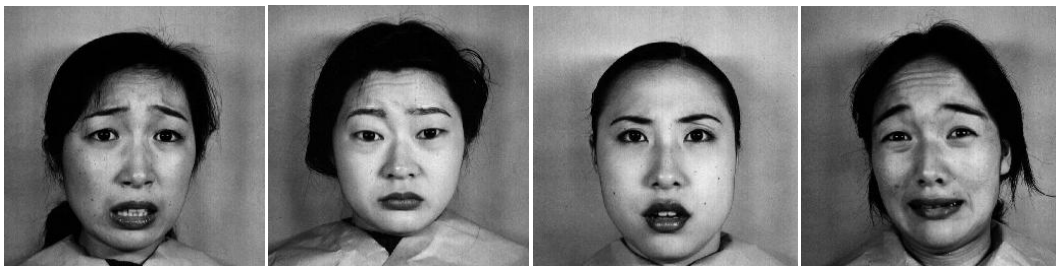


Figura 46. Imágenes de la base de datos JAFFE clasificadas como fear.

Al igual que en las pruebas realizadas sobre la base de datos CK+, se procede a dividir la base de datos JAFFE en varios fragmentos, en este caso cuatro debido a tener un menor número de imágenes. Al comparar los resultados donde no hay independencia entre las imágenes de entrenamiento y las de prueba, con el caso en que si están independizadas (ambos mostrados en la Tabla 18) se puede observar que la precisión total disminuye en un 12%. Al igual que se explicó para la base de datos CK+ la precisión, tanto total como de cada emoción, se ve reducida por ser distintas las imágenes de prueba que las de test, lo que aumenta la dificultad de la detección.

Tabla 18. Comparativa de resultados para JAFFE

Expresión Analizada→	*Anger*	*Disgust*	*Fear*	*Happiness*	*Sadness*	*Surprise*	**Precisión**
12 puntos	0,73	0,79	0,34	0,75	0,53	0,60	**0,62**
19 puntos	0,73	0,86	0,28	0,91	0,63	0,73	**0,69**
19 puntos e independencia de imágenes entrenamiento/prueba	0,43	0,71	0,25	1,00	0,43	0,57	**0,57**

Si se comparan los resultados obtenidos en la base de datos CK+ con los obtenidos en JAFFE mostrados en la Tabla 19, la diferencia es mínima en cuanto a precisión total, pero en la base de datos JAFFE, la precisión para la expresión *surprise* se reduce. Esto se debe a que algunos de los sujetos asiáticos no abren la boca para expresar sorpresa y la clasificación llevada a cabo por el sistema, se ve afectada. Lo mismo ocurre con *fear*, debido a que es la expresión más variante a la hora de ser expresada, como se ha explicado con anterioridad.

Tabla 19. Comparativa resultados para JAFFE y CK+

Expresión Analizada→	*Anger*	*Disgust*	*Fear*	*Happiness*	*Sadness*	*Surprise*	**Precisión**
CohnKanade+	0,51	0,83	0,42	0,83	0,67	0,91	**0,70**
JAFFE	0,73	0,86	0,28	0,91	0,63	0,73	**0,69**

5.1.6. Base de datos con imágenes de la webcam

En el apartado 5.1.3 se ha comentado que la base de datos CK+ no contiene el mismo número de imágenes para cada una de las expresiones a analizar y esto afecta negativamente a la clasificación de aquéllas que se encuentran en minoría.

Por otro lado, en el apartado 3 se explicó que la herramienta desarrollada permite analizar imágenes tomadas con la cámara del ordenador. Con el fin de probar esta característica e igualar el número de imágenes de las que se dispone para cada una de las expresiones, se crea una base de datos que contiene tanto imágenes pertenecientes a la base de datos CK+, como fotos tomadas de la webcam. Para obtener las imágenes desde la webcam, se ha pedido a familiares y amigos que posen ante la cámara del ordenador y que indicasen la expresión con la que posaban, para así poder etiquetar correctamente las imágenes obtenidas. Además, se han unido estas bases de datos debido a que en ambas la mayoría de los sujetos son europeos, por lo que su forma de expresarse es similar.

La base de datos creada con imágenes de CK+ junto con la webcam contiene 45 imágenes para cada una de las expresiones. Para el aprendizaje del sistema, se ha dividido la base de datos en cinco fragmentos debido a que no se tiene un número elevado de imágenes, de manera que se entrena con el 80% de las imágenes disponibles y se prueba con el 20% restante. En la Tabla 20 se muestran los resultados obtenidos al emplear el método *Thres* para la extracción de características y la comparación de desplazamientos como método de clasificación.

Tabla 20. Resultados obtenidos con imágenes de la webcam y de la base de datos CK+

Expresión Analizada→	*Anger*	*Disgust*	*Fear*	*Happiness*	*Sadness*	*Surprise*
Expresión Detectada						
Anger	**0,67**	0,16	0,23	0,10	0,14	0,02
Disgust	0,17	**0,74**	0,10	0,46	0,08	0,00
Fear	0,10	0,00	**0,35**	0,05	0,38	0,07
Happiness	0,00	0,06	0,00	**0,28**	0,02	0,02
Sadness	0,05	0,02	0,17	0,09	**0,11**	0,00
Surprise	0,02	0,02	0,15	0,02	0,27	**0,89**
Precisión	**0,51**					

En la base de datos CK+ las expresiones con menor número de imágenes eran *sadness, fear* y *anger*, por lo que el resultado esperado al añadir imágenes de estas expresiones es mejorar su clasificación en comparación con la Tabla 15. Para el caso de *anger* sí mejora, al contrario que para *sadness* y *fear*, esto se debe a que a pesar de tener un número mayor de imágenes que mejoría el entrenamiento para dichas expresiones, en las imágenes tomadas de la webcam, cada uno de los sujetos expresan de forma muy distinta la expresión *fear* como ocurre en CK+ y que se explicó en el apartado 5.1.2. Además, los sujetos femeninos tienen en su mayoría flequillo, por lo que las cejas (y a veces ojos) son total o parcialmente tapadas y esto reduce la información para la clasificación. Por otro lado, estas nuevas imágenes han sido tomadas bajo diferentes condiciones de luz y algunas sombras en la región de la boca producen una detección de puntos errónea. La solución de estos problemas de oclusión y deficiencia de iluminación son un reto en el área de reconocimiento de emociones.

5.2. Método basado en Unidades de Acción.

Para evaluar el método basado en AUs, al igual que para el método de comparación de matrices, se utilizan las bases de datos *CohnKanade+* y JAFFE. Se parte de las decisiones tomadas anteriormente, es decir, se elige un tamaño de cara de 300x300 pixeles y se detectan 19 puntos a partir de los cuales se obtienen las correspondientes AUs y se procede a la clasificación.

En la Tabla 21 se muestran los resultados obtenidos al realizar pruebas sobre la base de datos *CohnKanade+* utilizando el método basado en AUs con independencia entre las imágenes de entrenamiento y de prueba. Para compararlo con los resultados obtenidos al utilizar el método basado en comparación de desplazamientos, cuyos resultados se mostraron en la Tabla 10, se muestran de nuevo de forma abreviada en la Tabla 22. Se observa que la precisión al clasificar las emociones de *anger, happiness* y *sadness* es considerablemente reducida, entre 10-20%. Como ya se explicó en apartados anteriores, *anger*, *sadness* y *fear* son las expresiones que mayor confusión proporcionan en su clasificación. Para el caso de *happiness* es muy confundida con *fear*, esto se debe a que como se explicó con anterioridad, para la base de datos CK+, debido a la forma que tiene los sujetos de expresar la emoción *fear*, en numerosas ocasiones tiene características comunes con *happiness*, y para ayudar al sistema a clasificar *fear* (una de las "expresiones complicadas"), se ha priorizado esta expresión en los casos de empate entre emociones.

Los resultados están directamente influenciados por las condiciones que se establecen en el algoritmo implementado y la modificación de una de ellas o el añadir una nueva AU (que pueda dar confusión entre varias expresiones) puede cambiar por completo los resultados en alguna o varias expresiones. Además, en el caso del método utilizando AUs se implementan reglas fijas, que no están adaptadas a las diferentes bases de datos (sin entrenamiento), por lo que aumenta la dificultad de diferenciar dichas emociones. Esta carencia de adaptación del sistema, también influye al resto de expresiones, reduciéndose los resultados alrededor de un 10%.

Tabla 21. Pruebas en CohnKanade+ con el método basado en AUs.

Expresión Analizada->	*Anger*	*Disgust*	*Fear*	*Happiness*	*Sadness*	*Surprise*
Expresión Detectada						
Anger	**0,21**	0,12	0,20	0,01	0,19	0,11
Disgust	0,18	**0,71**	0,00	0,14	0,13	0,01
Fear	0,15	0,03	**0,34**	0,20	0,11	0,00
Happiness	0,12	0,06	0,03	**0,45**	0,00	0,00
Sadness	0,27	0,01	0,17	0,14	**0,53**	0,00
Surprise	0,06	0,06	0,08	0,06	0,05	**0,88**
Precisión	**0,52**					

Tabla 22. Pruebas en CohnKanade+ con el método comparación de desplazamientos

Expresión Analizada→	*Anger*	*Disgust*	*Fear*	*Happiness*	*Sadness*	*Surprise*	**Precisión**
Aciertos	0,42	0,63	0,42	0,83	0,32	0,91	**0,73**

También se realizan pruebas sobre la base de datos JAFFE y se muestran en la Tabla 23 para ser comparada con el caso del uso de comparación de matrices que se mostraba en la Tabla 10, pero que se añade resumido en la Tabla 24.

Tabla 23. Pruebas en JAFFE con el método basado en AUs.

Expresión Analizada->	*Anger*	*Disgust*	*Fear*	*Happiness*	*Sadness*	*Surprise*
Expresión Detectada						
Anger	**0,33**	0,29	0,38	0,00	0,37	0,20
Disgust	0,20	**0,29**	0,16	0,28	0,27	0,00
Fear	0,00	0,00	**0,03**	0,00	0,00	0,10
Happiness	0,40	0,21	0,09	**0,50**	0,13	0,10
Sadness	0,07	0,18	0,16	0,13	**0,20**	0,00
Surprise	0,00	0,04	0,19	0,09	0,03	**0,60**
Precisión	**0,47**					

Tabla 24. Pruebas en JAFFE con el método basado en comparación de matrices.

Expresión Analizada→	*Anger*	*Disgust*	*Fear*	*Happiness*	*Sadness*	*Surprise*	**Precisión**
Aciertos	0,73	0,86	0,28	0,91	0,63	0,73	**0,69**

Se aprecia que los porcentajes de acierto son reducidos al utilizar el método basado en unidades de acción. Esto se debe a las diferencias que se comentaron en apartados anteriores entre la forma de expresarse de los sujetos asiáticos (JAFFE) respecto a los sujetos occidentales (CK+). Las diferentes AUs y condiciones se realizaron basadas en la base de datos CK+ no son apropiadas para JAFFE. Si se observan imágenes de JAFFE correspondientes a las expresiones de *sadness* y *anger* (en la Tabla 23 *sadness* es muy confundida con *anger*) se aprecia que los sujetos asiáticos para expresar la emoción *sadness* no mueven la comisura de los labios hacia abajo (AU15 que se define como importante para la base de datos CK+), si no que encogen los labios como ocurre para la expresión *anger*. Se puede ver un ejemplo en la Figura 47. Para otros sujetos, la expresión *anger* no se expresa encogiendo los labios como ocurre en los sujetos occidentales de la base de datos CK+, sino que estiran la comisura de los labios y esto hace que sea confundido con la expresión *happiness*. Se puede ver un ejemplo en la Figura 48 Para el caso de la expresión *fear*, también se vio en el apartado 5.1.5 que cada sujeto la reproducía con diferentes características, por lo que para la clasificación mediante el uso de AUs, también afecta.

(a) (b)

Figura 47. Imágenes de la base de datos JAFFE clasificadas como (a) anger, (b) sadness.

Figura 48. Imágenes de la base de datos JAFFE clasificadas como anger

5.3. Otros estudios

Los comienzos de este proyecto partían del trabajo realizado en [34], que identificaba tan sólo tres expresiones (*anger, sadness* y *surprise*) sobre la base de datos JAFFE. Con el método final de extracción de puntos y comparación de desplazamientos, se han mejorado los resultados como se muestra en la Tabla 25.

Tabla 25. Comparación de resultados entre el algoritmo inicial [34] *y el algoritmo final*

Base de datos	**Algoritmo de prueba**	*Anger*	*Disgust*	*Fear*	*Happiness*	*Sadness*	*Surprise*
JAFFE	Algoritmo inicial	0,53				0,37	0,57
	Algoritmo final 12 puntos	0,73	0,79	0,34	0,75	0,53	0,60
	Algoritmo final 19 puntos	0,73	0,86	0,28	0,91	0,63	0,73
CK+	Algoritmo final 19 puntos	0,51	0,83	0,42	0,84	0,67	0,91

En primer lugar, con el nuevo algoritmo se aumenta el número de expresiones detectadas de tres a seis. Por otro lado, para las tres expresiones que se detectaban en el algoritmo inicial, se ha mejorado la clasificación alrededor de un 15-20% para el caso de detectar 12 puntos, y alrededor de un 20-30% para el caso de extraer 19 puntos. Además, si se cambia la base de datos a una más completa y con un mayor número de sujetos, como es el caso de *CohnKanade+*, los resultados mejoran hasta un 40%.

En [18] se puede apreciar que al emplear técnicas más complejas de detección de puntos, como máscaras parametrizadas similares a *Candide*, y de clasificación de emociones como multi-class SVM o Adaboost, los resultados son mejores que los obtenidos con una detección de puntos básica y la posterior clasificación mediante comparación de desplazamientos o unidades de acción que se ha desarrollado en este proyecto (Tabla 26).

Tabla 26. Comparación de diferentes métodos de clasificación en CK+

Expresión Analizada→	*Anger*	*Disgust*	*Fear*	*Happiness*	*Sadness*	*Surprise*	**Precisión**
Multi-class SVM [18]	0,71	0,92	0,64	0,86	0,86	0,95	**0,77**
Adaboost [18]	0,73	0,92	0,56	0,91	0,82	0,95	**0,83**
Comparación de desplazamientos	0,42	0,63	0,42	0,83	0,32	0,73	**0,70**
AUs	0,21	0,71	0,34	0,45	0,53	0,88	**0,46**

En [17] se emplea extracción de características basada en la forma o en la apariencia, y se clasifica la emoción mediante SVM. Los resultados obtenidos, mostrados en la Tabla 27, son similares a los obtenidos en el método propuesto (Tabla 26), siendo positiva la mejora en las expresiones *fear* y *sadness*, por su dificultad de clasificación, y negativo la reducción de aciertos para *happiness*. De igual forma ocurre si se aplica el método LBP con diferentes características [22], y cuyos resultados se muestran también en la Tabla 27.

Tabla 27. Comparación de diferentes métodos de extracción de características en CK+

Expresión Analizada→	*Anger*	*Disgust*	*Fear*	*Happiness*	*Sadness*	*Surprise*	**Precisión**
SVM (*shape*) [17].	0,35	0,68	0,22	0,98	0,04	1,00	**0,50**
SVM (*appearance*) [17].	0,70	0,95	0,22	1,00	0,60	0,99	**0,67**
LBPms [22].	0,55	0,63	0,49	0,77	0,63	0,82	**0,65**
LBPgm [22].	0,30	0,40	0,17	0,81	0,65	0,88	**0,53**

En [22], los resultados al aplicar el método LGBP (*Local Gabor Binary Patterns*), Tabla 28, mejoran en comparación con los obtenidos al utilizar el método LBP con diferentes características, algunos de los cuales se mostraron en la Tabla 27. Por otro lado, los resultados presentados en [20] son considerablemente mejores. En este caso las características son extraídas utilizando Gabor Wavelets y se clasifica la emoción, primero con el método de clasificación SVM, obteniendo una precisión media de 96.17 ± 4.0 y segundo con el método de redes neuronales, obteniéndose una precisión de 97.14 ± 3.7.

Tabla 28. Resultado en CK+ empleando LGBP [22]*.*

Expresión Analizada→	*Anger*	*Disgust*	*Fear*	*Happiness*	*Sadness*	*Surprise*	**Precisión**
Aciertos	0,63	0,63	0,51	0,79	0,68	9,83	**0,68**

En vista de las tablas mostradas anteriormente, la extracción de características es muy relevante, siendo mejor la extracción basada en la apariencia que en la forma. Pero la solución más óptima es el uso de Gabor Wavelets, ya que se aumenta la precisión de la clasificación hasta un 30%. También es mejor utilizar un número de puntos superior a los 19 utilizados en este proyecto o el uso de máscaras parametrizadas del tipo *Candide.*

En cuanto a los métodos de clasificación de la emoción, en este proyecto se han desarrollado métodos sencillos para poder comparar sus resultados con otros más complejos utilizados actualmente como son Adaboost y SVM. Se comprueba que para obtener una precisión aceptable (junto con otras características como la robustez) es necesario el uso de dichas técnicas más elaboradas, y por ello en la actualidad se pretende mejorarlas y utilizar otras como el uso de redes neuronales.

6. Conclusiones

En este proyecto se ha tratado de analizar y resumir el estado del arte de métodos y técnicas para el reconocimiento automático de emociones. Se han definido distintos tipos de expresiones y la influencia que tiene en el reconocimiento reproducir una emoción en un entorno controlado, y para el análisis del sistema se han establecido siete expresiones básicas (incluyendo la neutral) [1].

Debido a que en la actualidad el reconocimiento de emociones ha sido, y está siendo, muy estudiado, se ha desarrollado e implementado una herramienta intuitiva y accesible que permite analizar y comparar métodos y técnicas de reconocimiento automático de emociones sobre distintos materiales (imágenes estáticas, vídeo…). Además, esta herramienta permite evaluar métodos de extracción de características y métodos de clasificación de la emoción, con diferentes combinaciones entre ellos.

Por otro lado, se ha desarrollado un método de extracción de características basado en la detección de puntos característicos de la cara, realizando una previa segmentación para mejorar la precisión. A su vez, se han implementado dos métodos de clasificación de la emoción, uno basado en comparación de matrices, y otro basado en unidades de acción, en ambos evaluando los desplazamientos entre los puntos detectados.

Para evaluar el funcionamiento de la herramienta y la precisión de cada uno de los métodos, se han llevado a cabo una serie de pruebas, de las cuales se han obtenido tablas de resultado. El sistema se ha probado sobre diferentes bases de datos, de manera que permita evaluar cómo le afectan características como la calidad de la imagen, las condiciones de iluminación y el número de imágenes analizadas. Además, se ha evaluado la influencia que tiene probar el sistema con las mismas imágenes de entrenamiento, entrenar y probar con distintas imágenes (mediante la separación en fragmentos de la base de datos) y la obtención de resultados al probar las imágenes una base de datos, sobre el entrenamiento realizado con otra. Tras la realización de estas pruebas se han extraído varias conclusiones que se comentan a continuación.

La herramienta desarrollada es intuitiva y permite numerosas configuraciones. Se pueden seleccionar diversas bases de datos u obtener imágenes desde la cámara del ordenador, combinar métodos de extracción de características con métodos de clasificación de la emoción de diferentes formas, y obtener resultados tanto de forma visual como en forma de tabla y porcentajes de acierto y fallo. De esta forma, se pueden comparar de forma rápida y sencilla numerosas combinaciones de métodos con el objetivo de llegar a las mejores soluciones en el reconocimiento automático de emociones. Además esta herramienta incluye la fase de entrenamiento del sistema y de tal forma que pueda ser ejecutada eficazmente.

Al igual que los resultados obtenidos por otros investigadores, se ha confirmado que las cuatro regiones más significativas en el rostro cuando se muestran emociones son, por orden de importancia, la boca, las cejas, los ojos y la nariz. Por lo que los métodos de reconocimiento de emociones se basan principalmente en las variaciones que sufren estas regiones para la clasificación de la emoción, obviando otras como las orejas o la frente.

El número de imágenes utilizadas en el entrenamiento es muy relevante. A mayor número, mejores resultados se obtienen, pero hay que tener en cuenta que el rendimiento del sistema empeora, como se comprobó en el trabajo de prácticas previo a este proyecto.

En el método de clasificación de la emoción basado en comparación de matrices, el modelo resultante se ajusta demasiado a los datos de entrenamiento, y su generalización es inadecuada. Como se mostró en el apartado 5, los resultados de evaluar las imágenes de una base de datos, sobre el sistema entrenado con una diferente, proporciona peores resultados.

Para el método de clasificación de la emoción basado en AUS, es importante tener en cuenta la base de datos sobre la que se trabaja, sobre todo debido a la raza de los sujetos analizados. Como se mostró en el apartado 5, existen diferencias en la forma de expresarse (acciones y partes del rostro que actúan)

entre los sujetos asiáticos y occidentales. Esto repercute directamente en las AUs elegidas como más representativas para cada emoción o simplemente en la decisión de qué AUs implementar y cuáles no, para hacer el sistema lo más eficiente posible. Los resultados obtenidos al utilizar este método son peores que si se emplean otros en los que se ha realizado un entrenamiento previo. Una forma de mejorar estos resultados sería aplicar las AUs ya definidas a un árbol de decisión en el que sí se haga un entrenamiento previo.

En general, las expresiones *fear* y *anger* son las más difíciles de diferenciar. Como se vio en el apartado 5, para expresar la emoción *fear*, cada sujeto utiliza gestos de la boca y cejas muy diferentes, y ocurre de forma similar para la expresión *anger*. Sin embargo, las expresiones de *surprise* y *happiness* son las más fáciles, ya que las características principales de estas emociones son muy representativas y bastante diferenciadas del resto de emociones (por ejemplo abrir la boca para la expresión *surprise*). Por otro lado, las emociones de *fear* y *anger* son comúnmente confundidas con *sadness*, debido a la similitud de las regiones que actúan al ser expresadas.

Es necesaria una base de datos que contenga un número suficientemente representativo de sujetos y con similar número de imágenes para cada expresión que se pretende detectar. Además, los sujetos deben ser de diferentes etnias y edades, y encontrarse bajo diferentes condiciones de iluminación, posición y calidad de la imagen, de forma que la base de datos se convierta en un estándar, sin aumentar significativamente el gasto computacional. Es también importante tener en cuenta que las imágenes tomadas en un laboratorio son usualmente exageradas y artificiales, pero crear una base de datos de imágenes o vídeos en las que los sujetos expresen las emociones de forma espontánea es muy difícil.

Para corregir el problema de la inclinación del rostro o de la diferencia de su tamaño entre las distintas imágenes de una base de datos, se pueden llevar a cabo diferentes tipos de normalizaciones, que dependen directamente del método utilizado en extracción de características y clasificación de la emoción, para ello en este proyecto se realizó una transformación afín de los puntos detectados.

7. Trabajo futuro

Siguiendo la línea de trabajo de este proyecto, se podrían implementar nuevos métodos más complejos tanto de extracción de características, como aplicar Gabor Wavelets, y de clasificación de la emoción, como un árbol de decisión basado en Unidades de Acción o el método SVM. Además, se emplearía la herramienta desarrollada en este proyecto para comparar los resultados de los métodos implementados en este proyecto, con otros métodos desarrollados por investigadores en éste ámbito.

Como se ha comentado anteriormente, es necesaria una base de datos estándar que sea muy completa y que contenga numerosas y diferentes características, por lo que sería muy enriquecedor el estudio exhaustivo y la creación de una de ellas.

Otra línea de investigación que se propone es el estudio del marcado de puntos, desde cómo mejorar su precisión hasta evaluar la influencia de los errores cometidos en su detección. Es importante tener en cuenta que la incorrecta detección está muy influenciada por la presencia en el rostro de arrugas, sombras o inclinaciones de la cara.

En cuanto a los métodos de clasificación de la emoción, se podrían añadir a la matriz de representativa de clasificación de la expresión más columnas, por ejemplo, una que contuviera información de las distancias entre puntos de la cara que aumentase la información de la que se dispone para la clasificación de la emoción. También se podrían detectar un mayor número de puntos con el fin de poder definir un mayor número de Unidades de Acción y así mejorar la clasificación de emociones.

Cabe destacar, que combinar estos métodos con información contextual o proveniente de otras formas de expresión, como la voz, permitiría clasificar de forma más efectiva las emociones, siendo relevante en la discriminación de las expresiones más confusas, como son *anger*, *sadness* y *fear*.

Por otro lado, los métodos desarrollados en este proyecto están destinados al análisis de imágenes estáticas, por lo que un gran avance sería desarrollar métodos basados en vídeo, con la correspondiente creación de una base de datos adecuada.

Relacionado también con la aplicación a vídeo, aunque la herramienta y los métodos implementados trabajan de manera muy rápida, se podría intentar mejorar la eficacia del sistema para conseguir tiempos adecuados para su aplicación en tiempo real.

Por último, una relativamente nueva línea de investigación que es bastante importante es el estudio de las expresiones espontáneas, y la correspondiente mejora o implementación de métodos para su correcta clasificación.

8. Referencias

[1] P. Ekman y W. V. Friesen, “Constants across cultures in the face and emotion.,” *J. Pers. Soc. Psychol.*, vol. 17, no. 2, págs. 124–129, 1971.

[2] V. Bettadapura, “Face expression recognition and analysis: the state of the art,” *College of Computing, Georgia Institute of Technology*, 2012.

[3] T. Wu, S. Fu, y G. Yang, “Survey of the Facial Expression Recognition Research,” *Adv. Brain Inspired Cogn. Syst.*, págs. 392–402, 2012.

[4] B. D. Lucas y T. Kanade, “An iterative Image Registration Technique with an Application to Stereo Vision,” in *DARPA Image Understanding Workshop*, 1981, págs. 121–130.

[5] H. Schneiderman and T. Kanade, “Object Detection Using the Statistics of Parts,” *Int. J. Comput. Vis.*, vol. 56, no. 3, págs. 151–177, Feb. 2004.

[6] P. Viola y M. Jones, “Rapid object detection using a boosted cascade of simple features,” *Proc. 2001 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognition. CVPR 2001*, vol. 1, págs. I–511–I–518, 2001.

[7] G. Lemaıtre y M. Radojevic, “Directed Reading: Boosting algorithms,” Univ. Heriot-Watt Univ. de Girona, Univ. de Bourgogne, 2009.

[8] J. Ahlberg, “Candide-3-an updated parameterised face,” Dept. de Ingeniería Eléctrica, Univ. Linköping, Suecia, 2001.

[9] T. Cootes, G. Edwards, y C. Taylor, “Active appearance models,” *IEEE Trans. pattern analysis and machine intelligence,* vol. 23, no. 6, págs. 681–685, 2001.

[10] K. Yu, Z. Wang, M. Hagenbuchner, y D. Dagan Feng, “Spectral embedding based facial expression recognition with multiple features,” *Neurocomputing*, vol. 129, págs. 136–145, Apr. 2014.

[11] L. Blazquez, “Reconocimiento Facial Basado en Puntos Característicos de la Cara en entornos no controlados,” Proyect o fin de Carrera, Universidad Autónoma de Madrid, Madrid, En. 2013.

[12] L. Wiskott y J. Fellous, “Face recognition by elastic bunch graph matching,” *Pattern Analysis and machine intelligence,* págs. 1–23, July, 1997.

[13] S. P. Khandait, R. C. Thool, y P. D. Khandait, “ANFIS and BPNN based Expression Recognition using HFGA for Feature Extraction,” vol. 2, no. 1, págs. 11–22, 2013.

[14] P. E. Group, “FACS Archives - Paul Ekman Group, LLC.” [Online]. Available: http://www.paulekman.com/product-category/facs/.

[15] M. Singh, A. Majumder, y L. Behera, “Facial expressions recognition system using Bayesian inference,” *2014 Int. Jt. Conf. Neural Networks*, págs. 1502–1509, Jul. 2014.

[16] P. Lucey, J. Cohn, y T. Kanade, “The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression,” *Comput. Vision and Pattern Recognition Workshops (CVPRW),* July, 2010.

[17] N. Khan, “A comparative analysis of facial expression recognition techniques,” *Adv. Comput. Conf. (IACC), 2013 IEEE 3rd International,* págs. 1262–1268, 2013.

[18] M. Taner Eskil y K. S. Benli, “Facial expression recognition based on anatomy,” *Comput. Vis. Image Underst.*, vol. 119, págs. 1–14, Feb. 2014.

[19] M. Pantic y M. Bartlett, “Machine analysis of facial expressions,” ,Dept. Computación, *Imperial College London*, Dept. Computacion de redes, Universidad de California2007.

[20] S. Zhang, X. Zhao, y B. Lei, “Robust facial expression recognition via compressive sensing.,” *Sensors (Basel).*, vol. 12, no. 3, págs. 3747–61, Jan. 2012.

[21] T. Ojala, “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns,” *Pattern Analysis and Machine Inteliggence, IEEE Transactions on* vol. 24, no. 7, págs. 971–987, 2002.

[22] S. Moore y R. Bowden, “Local binary patterns for multi-view facial expression recognition,” *Comput. Vis. Image Underst.*, vol. 115, no. 4, págs. 541–558, Apr. 2011.

[23] D. G. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints,” *Int. J. Comput. Vis.*, vol. 60, no. 2, págs. 91–110, Nov. 2004.

[24] L. Zhang, D. Tjondronegoro, y V. Chandran, “Facial expression recognition experiments with data from television broadcasts and the World Wide Web,” *Image Vis. Comput.*, vol. 32, no. 2, págs. 107–119, Feb. 2014.

[25] “Arboles decision id3.” [Online]. Available: http://es.slideshare.net/FernandoCaparrini/arboles-decision-id3.

[26] A. Gil y D. Zapatero, “Reconocimiento facial de emociones.” Informe final de prácticas, CITSEM, Madrid, 2014.

[27] “Red neuronal artificial,” *wikipedia- Enciclopedia libre*. [Online]. Available: http://es.wikipedia.org/wiki/Red_neuronal_artificial.

[28] G. Mascorro y G. Torres, “Sistema para identificación de hablantes robusto a cambios en la voz,” *ingenius.ups.edu.ec*, págs. 45–53, 2012.

[29] L. Peterson, “K-nearest neighbor,” *Scholarpedia*, 2009.

[30] C. Truesdell, “Acknowledgments,” *Pure Appl. Math.*, pág. xxv, 1980.

[31] “JAFFE Database.” [Online]. Available: http://www.kasrl.org/jaffe.html.

[32] “MMI Database.” [Online]. Available: http://ibug.doc.ic.ac.uk/research/mmi-database/.

[33] “eNTERFACE’05 Database,” 2005. [Online]. Available: http://www.enterface.net/enterface05/main.php?frame=emotion.

[34] R. N. Rojas Bello, “Identificación de características relevantes para reconocimiento de emociones en el rostro,” Proyecto fin de Mater, Dep. Ingeniería Informática, Univ. Autónoma de Madrid, 2009.

[35] C. Tanchotsrinon, S. Phimoltares, S. Maneeroj, y A. Virtual, “Facial expression recognition using graph-based features,” Dept. Matemáticas, Univ. Chulalongkon, Bangkok, 2011.

[36] S. González y Á. Martínez, “Interfaces Naturales para Realidad Aumentada,” Informe final de prácticas, CITSEM, págs. 1–16, 2014.

[37] B. Seddik, H. Maâmatou, S. Gazzah, T. Chateau, N. Essoukri, y B. Amara, “Unsupervised Facial Expressions Recognition and Avatar Reconstruction from Kinect,” in *10th International Multi-Conference on Systems, Signals & Devices (SSD)*, 2013, págs. 1–6.

[38] F. Wikipedia, “k-Nearest Neighbors Algorithm.” págs. 1–7. [Online] Available: http://scholarpedia.org/article/K-Nearest_Neighbor

ANEXO. Manual del programador

1. Introducción

El siguiente documento tiene como objetivo entregar las pautas para gestionar y ampliar la herramienta desarrollada en este proyecto.

Mediante el uso de esta herramienta se pretende comparar diferentes métodos para el desarrollo de sistemas de reconocimiento automático de emociones. Gracias a este manual se permite al usuario conocer la interfaz gráfica utilizada y cómo modificarla o mostrar resultados visuales en ella. También explica como incluir, modificar o eliminar métodos de extracción de características y de clasificación de la emoción, y se presenta la estructura para poder exportar/importar los datos del entrenamiento y las tabas Excel de resultados.

2. Acceso a la herramienta

Para poder utilizar la herramienta desarrollada para el análisis de métodos de reconocimiento de emociones es necesario tener instalado el programa Matlab versión 2012b junto con la *Computer Vision Toolbox* de dicho programa.

La función principal que se debe ejecutar (*Run*) en el programa Matlab para poder comenzar a usar la herramienta es denominada *Interface_Faces.m*. Esta función tiene asociada su correspondiente imagen *Interface_Faces.fig* que permite el acceso a la interfaz gráfica.

3. Gestión de la interfaz gráfica

Para poder editar la apariencia y la funcionalidad de la interfaz gráfica se debe abrir *Interface_Faces.fig* con el gestor de interfaces gráficas GUIDE [1] tal como muestra la Figura 49, obteniendo como resultado la ventana de edición mostrada en la Figura 50. En dicha ventana, mediante los botones que aparecen a la izquierda, se pueden añadir nuevos botones, texto, imágenes, etc. Las propiedades de cada uno de ellos pueden ser editadas pulsando dos veces con el ratón sobre ellos o abriendo la opción *Property Inspector* mediante el botón derecho del ratón. La principal propiedad a tener en cuenta en cada una de las opciones es el nombre (*Tag*), ya que es su identificativo dentro del código del programa.

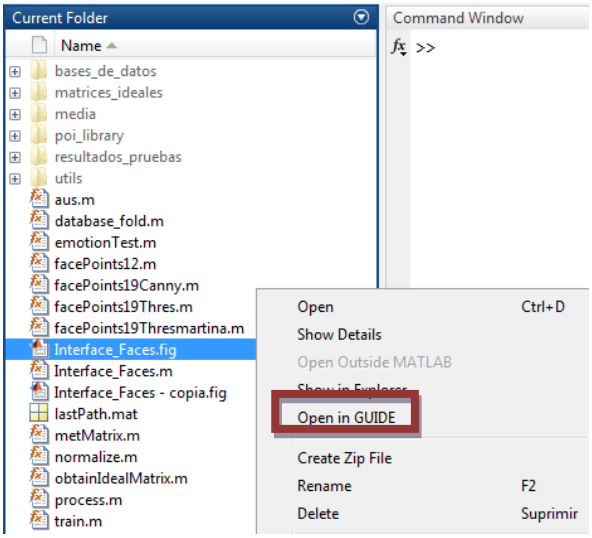


Figura 49. Abrir editor de interfaces gráficas en Matlab (GUIDE)

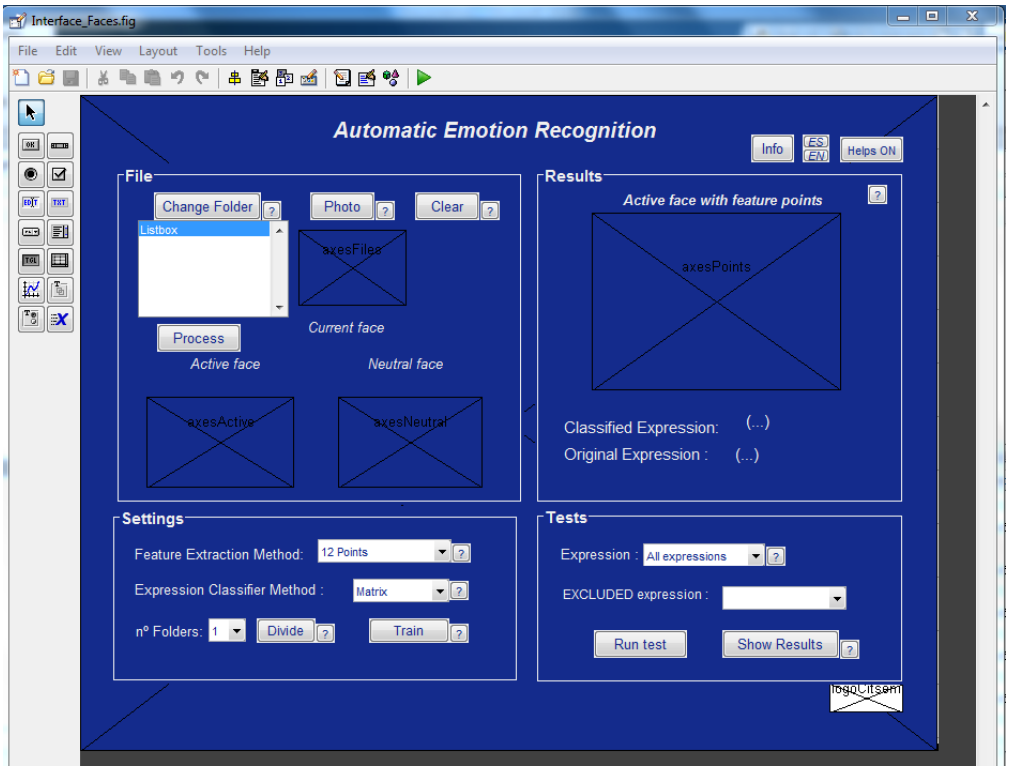


Figura 50. Ventada de edición GUIDE

Cuando se añade una nueva opción a la herramienta, en la función principal (*Interface_faces.m*) aparece su correspondiente apartado en el código como una subfunción (generalmente *Callback*) en la cual se realizan las acciones básicas como extraer el texto que ha sido seleccionado de una lista, conversiones de tipos de variables/formatos o re-direccionar a la función independiente que se desee ejecutar. Por ejemplo, para el caso de seleccionar el botón *Photo* (opción de tipo *pushbutton*), el código asociado podría ser el siguiente (después de '%' se escriben comentarios que no se implementan):

```
Function buttonPhoto_Callback(hObject, ~, handles)

handles.mode='photo'; %mode es una variable previa que diferencia
                      %archivos importados de una base de datos o de
                      %una webcam

[handles.imOrig] = LoadImages(handles.mode,,handles.axesActive);
                      % funcion externa que se encarga de adquirir la %imagen a analizar y la %muestra
                      en la interfaz en la %opcion nombrada como 'axesActive'
guidata(hObject,handles); %Actualiza la información en handles y
                      %hObject
```

Y si se selecciona un texto de una lista desplegable (opción de tipo *popup*):

```
function popupFMethod_Callback(hObject, ~, handles)
Fmethod=get(handles.popupFMethod,'String'); %Se extrae el texto
handles.Fmethod=cell2mat(Fmethod(get(handles.popupFMethod,'Value')));
                      %se convierte al tipo de variable óptimo para su uso %en el programa
guidata(hObject,handles);
```

Es también importante destacar la función que se ejecuta al abrir el programa en la cual es necesario incluir todos aquellos directorios que contengan información útil para su ejecución e inicializar las variables globales que se utilizan mediante *handles* ([1]).

```
function Interface_Faces_OpeningFcn(hObject, ~, handles, varargin)
```

```
addpath('bases_de_datos'); %directorio que contiene las bases de datos
  %que se pueden utilizar
addpath('media/ES'); %directorio contiene las funciones de ayuda
%escritas en español
addpath('media/EN'); %directorio contiene las funciones de ayuda
%escritas en ingles
imshow(imread('CITSEM.png'),'Parent',handles.logoCitsem %se carga en
                        %el programa la imagen del Logo del CITSEM y se muestra en
                        %la interfaz en la opción nombrada como 'logoCitsem'
handles.output = hObject;
```

4. Gestión de las funciones

Para implementar el carácter modular y ampliable de la herramienta, se ha creado la función *process.m* que re-direcciona a funciones independientes que contienen los diversos métodos de extracción de características y de clasificación de la emoción. Si se desea introducir un nuevo método se deben realizar las siguientes acciones:

Crear la función .m correspondiente

Utilizando GUIDE de Matlab, añadir un nombre representativo para dicho método en la opción desplegable correspondiente (*popupFMethod* si corresponde a un método de extracción de características o *popUpEmethod* si se trata de un método de clasificación) de la interfaz gráfica.

Dicho nombre otorgado a cada método se utiliza para ser seleccionado en la función *process.m*, que funciona de la siguiente manera: establece una condición (*if-elseid-end*) que compara (utilizando el comando *strcmp*), el texto contenido en la variable *FMethod* (definida en *Interface_Faces.fig*) con cada uno de los textos incluidos en el desplegable mencionado, y si coincide se ejecuta la función correspondiente. Por ejemplo:

- Función : *facePoints19Canny.m*;
- Opción de la Interfaz: *popupFMethod*
- Nombre en la interfaz: '19 Points Canny'
- En el código incluido en *process.m*:

```
if (strcmp(Fmethod,'19 Points Canny'))
   [Pn]=facePoints19Canny(faceNeutral);
   [P]=facePoints19Canny(faceOrig);
   nPoints=19;

elseif (strcmp(Fmethod,'otro'))%nuevo método
   []=otro();
end
```

En los sistemas de reconocimiento de emociones se requiere un previo entrenamiento del sistema, por lo que se crea una función modular del mismo modo que se crea *process.m*, llamada *train.m*, que además realiza la lectura automática de todas las imágenes de la base de datos seleccionada. De manera similar trabaja la función que extrae los resultados y los almacena en una tabla Excel, lee todas las imágenes de una base de datos, y mediante la función *process.m* procesa las imágenes utilizando el entrenamiento de dichos métodos.

Es importante comentar que la detección del rostro y su división en regiones, como se comentó en el apartado 4.2, se realiza en la función *process.m*, previo a la selección de los métodos a implementar.

5. Otras características

En este apartado se comentan algunos comandos y librerías necesarios para la importación/exportación de datos en el programa.

Para poder exportar los datos resultado del entrenamiento del sistema, se utiliza *save*, eligiendo como directorio destino el creado con un nombre significativo del método del cual se realiza el entrenamiento. Esto se hace al final de la función *train.m.*

Para importar los datos del entrenamiento de un método concreto y poder utilizarlo para procesar una imagen, se utiliza el comando *load* dentro de la propia función del método de clasificación. Es importante saber que el nombre con el que se importa en el programa es aquél que tenía asignado en el programa, no con el nombre que se guarda. Por ejemplo, se tiene una matriz llamada *P* que se guarda en el directorio *matrices* con el nombre *Prueba* de la siguiente forma: *save (finalP, matrices/Prueba.mat)*. En dicho directorio el nombre que aparece es *Prueba*, pero al volver a importar dicha matriz al programa mediante l*oad (matrices/Prueba),* no existirá en el programa ninguna matriz llamada *Prueba*, sino que se habrá cargado la matriz con el nombre *P*.

Por otro lado, para almacenar/leer los porcentajes resultado de las pruebas (función *emotionTest.m*), es necesario cargar la librería *poi_library* en Matlab, y después se usan las funciones correspondientes a escribir datos en una tabla Excel (*xlswrite*) o a importarlos (*uiimport*).

Para reducir el número de funciones que se muestran en la ventana principal de Matlab, se han creado dos directorios denominados *utils* y m*edia*. *Utils* contiene pequeñas funciones que son básicas y no es necesario modificar como las destinadas a detectar la cara en una imagen, a crear barras de estado o a leer el nombre con el que una imagen está guardada. El directorio *Media* contiene cada una de las pequeñas funciones de ayuda escritas en español e inglés, así como imágenes que se muestran a lo largo del programa (ejemplo el logo del centro CITSEM).

6. Gestión de folds

En los sistemas de reconocimiento de emociones es necesario realizar una fase de aprendizaje del sistema, para la cual se dispone de una base de datos. El entrenamiento es realizado con un 90% de las imágenes para posteriormente probarlo con el 10 % restante.

Para implementar esta característica, se divide la base de datos en 10 *folds* entre las cuales se reparten las imágenes de una base de datos. A la hora de crear el entrenamiento el funcionamiento actual es el siguiente: se ejecutan los métodos elegidos sobre las imágenes de todas las *folds* y se obtienen los datos correspondientes, se crea un directorio por cada *fold* de manera que cada uno de ellos contiene la media de los datos obtenidos al procesar cada una de las imágenes de las restantes 9 f*olds (entrenamiento).* Al realizar las pruebas (*test*), el sistema obtiene los resultados de cada una de las *folds* por separado, para ello procesa las imágenes de la *current_fold* usando los datos almacenados en el directorio destinado a dicha *fold* (como se ha comentado, contiene el entrenamiento con el 90% de las imágenes restantes). Para obtener la tabla de resultado final se hace una media de las 10 pruebas obtenidas (una por cada *fold*).

Referencias

[1] D. Orlando Barragán Guerrero. “Manual de interfaz Gráfica de usuario en Matlab”, 2008.