Reconocimiento facial de emociones

Memoria final de “Practicas Externas en CITSEM”

30 de enero de 2015

Almudena Gil, Diego Zapatero

Tutora: Martina Eckert

Grupo de trabajo: Realidad Aumentada

Resumen

La comunicación no verbal entre los seres humanos cada día cobra más importancia, y las expresiones faciales son el mejor ejemplo de ello. En los últimos años, gracias al avance en el procesado digital de la imagen se han desarrollado numerosos sistemas capaces de reconocer estas emociones de una manera automática. En esta investigación se realiza primeramente un estudio del arte acerca de todos estos sistemas, analizando qué métodos utilizan y cuáles son sus ventajas. A raíz de ello, se implementa como primer paso un método para extraer las características faciales mediante el análisis de acciones de músculos. Como segundo paso se crea un algoritmo capaz de clasificar la expresión de una cara en concreto a partir de las particularidades de ella. Con ello, y a través de una interfaz gráfica para mayor comodidad, se pretende lograr el objetivo de reconocer la expresión. Por último, se realizan diferentes pruebas con el fin de obtener el mayor número de conclusiones posibles para poder determinar qué mejoras serían necesarias y cuáles podrían ser las posibles líneas de trabajo a partir de ellas.

Abstract

Non-verbal communication between humans every day is becoming more important, and facial expressions are the best example of it. In recent years, thanks to advances in digital image processing, a lot of systems able to recognize these emotions automatically have been developed. In this research, firstly, an art studio about all these systems is performed, analyzing which methods they use and its advantages. Consequently, as a first step, a method is implemented in order to extract facial features by analyzing muscle actions. As a second step, an algorithm is created to classify the expression of a specific face using the particularities of it. With this, and through a graphical interface for convenience, the main goal is to recognize the expression. Finally, different tests are performed in order to obtain the largest number of conclusions to determine which improvements would be needed and which lines of work could be possible from them.

1. Introducción

Cada día la interacción entre las personas y las máquinas que les rodean es más frecuente. A día de hoy se realizan numerosas investigaciones para hacer posible el reconocimiento de una expresión de manera automática gracias al desarrollo en estos últimos años del procesado de la imagen.

Las expresiones faciales son el medio más importante para expresar emociones y estados de ánimo. A través de ellas, se puede obtener una mejor compresión de lo que nos intentan comunicar los demás, reforzando o engañando la comunicación.

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

En esta investigación se persigue el reconocimiento de estas expresiones de manera automática. Se desea detectar seis emociones básicas: *fear* (miedo), *anger* (ira), *happiness* (felicidad), *surprise* (sorpresa), *sadness* (tristeza) y *disgust* (aversión).

Para lograr ese objetivo, se crea una herramienta llevando a cabo los pasos detallados a continuación:

- Obtención de la imagen, ya sea por medio de una base de datos o por la webcam para su posterior procesado.

- Extracción de características de la imagen, es decir, se selecciona el método de extracción de características adecuado para poder localizar y analizar diferentes regiones de la cara y llevar a cabo el reconocimiento de la emoción.

- Módulo de Aprendizaje, a partir del cual se crean una especie de rutas o caminos para poder llegar al paso siguiente. Para ello se utilizará un árbol de decisión.

- Clasificación de las emociones, es decir, el reconocimiento de las expresiones faciales anteriormente comentadas.

A continuación se detallan los apartados de los que consta el documento. En primer lugar se realiza un estudio del estado del arte en el ámbito del reconocimiento facial de emociones. En segundo lugar se lleva a cabo el desarrollo del trabajo, en el cual se explican el método de extracción de características y el modelo de aprendizaje utilizado para la detección y clasificación de la emoción. En tercer lugar se muestran los resultados obtenidos tras la aplicación del método y modelo seleccionado. En cuarto lugar se detallan las conclusiones obtenidas tras el desarrollo de esta investigación. Finalmente se plantean posibles trabajos futuros en este ámbito.

2. Estado del Arte

Antes de llevar a cabo los pasos comentados en la introducción, se realizan estudios y análisis sobre el estado del arte en el ámbito del reconocimiento facial de emociones, para poder así conocer los diferentes métodos y algoritmos que existen hoy en día.

Al mismo tiempo, se examina el trabajo realizado anteriormente por compañeros del CITSEM [1], a partir del cual se detectan ciertos problemas (por ejemplo, relacionados con el método de extracción de características entre otros) y se plantean nuevas mejoras para intentar solucionarlos.

A continuación se destacan los métodos más relevantes encontrados en el ámbito de la extracción de características:

- La detección de bordes y segmentación de regiones de la cara, (utilizado en [2], [3], [4], [5] y [6]) a partir de la cual se basa la investigación realizada con anterioridad en el CITSEM.

- Método de extracción de características de deformación: que consiste en la extracción de una cierta información de deformación en la cara, como puede ser la deformación geométrica (utilizado en [7]) o cambios de textura.

- Método de extracción de características de movimiento: usado principalmente para extraer puntos característicos o información de movimiento de áreas características a partir de imágenes, usado en [8] y [9].

- Método de extracción de características de estadística: describe las características de las imágenes con el uso de la estadística (histogramas o momento invariante), utilizado en [10].

- Método FACS (*Facial Action Coding System*) [11]: que permite identificar los músculos faciales que causan cambios en la cara. Al movimiento de estos músculos se le denomina *Action Units* (AUs) y existen 46. Es el método en el cual se basa esta investigación. Actualmente ha evolucionado a FACES (*Facial Action Coding Expression System*), que permite disminuir el tiempo de detección de las emociones.

- Método LBP (*Local Binary Patterns)*: es uno de los métodos que más se usan en el ámbito de la extracción de características, se basa en la asignación de un valor binario a cada uno de los pixeles de la imagen. Es utilizado en [12] y [13].

Respecto a los métodos de clasificación de la expresión, se detallan a continuación los más relevantes:

- Método basado en el Modelo Hidden Markov (HMM): [14] puede describir con eficiencia el modelo estadístico de una señal aleatoria. Tiene como ventajas que permite cambios de expresión, rotación de la cabeza o que no haya necesidad de entrenar de nuevo añadir nuevas imágenes.

- Método basado en Redes Neuronales Artificiales (ANN): [15], [16] y [17] inspirado en la observación de los sistemas de aprendizaje biológico que desarrollan las redes neuronales de un ser humano.

- Método Support Vector Machine (SVM), cuya utilidad se puede ver en [18] y [19]. Consiste en un modelo de "*Active Learning*" que analiza los datos según se van obteniendo para poder reconocer y clasificar patrones que se repiten.

- Métodos basados en la red Bayesian: [14] modelo gráfico probabilístico basado en la fórmula Bayesiana. Mejora la precisión de la clasificación pero requiere numerosos parámetros y un gran número de muestras para que los resultados sean reales.

- Métodos basados en Adaboost Algoritmo: [14] combina un método de clasificación débil y otro fuerte. Es considerablemente más rápido que el SVM pero su clasificación no es óptima con un número pequeño de muestras.

Para concluir, en los estudios [20], [21 y [22], se analiza de forma extensa y detallada el estado del arte que permite tener una mejor ideal general del desarrollo actual del reconocimiento de expresiones faciales y sus aplicaciones. Además, se comparan diferentes métodos para ver las ventajas y desventajas de cada uno y saber cuál es el más adecuado.

3. Procedimiento para el reconocimiento automático de expresiones

El objetivo principal de esta investigación es el reconocimiento facial de emociones. Como se ha comentado anteriormente, se pretenden detectar seis emociones básicas (*happiness*, *anger*, *fear*, *surprise*, *sadness* y *disgust*).

Para ello se utiliza el trabajo realizado anteriormente por alumnos en prácticas en el CITSEM [1], que consiste en una herramienta que permite comparar métodos para el reconocimiento de emociones y en el que se implementó un método básico. A partir de este trabajo se han desarrollado e implementado nuevos métodos y algoritmos con los que se mejora la identificación de las expresiones.

En la Fig.1 se muestra un esquema de los pasos que se siguen para el reconocimiento de una imagen, comentados anteriormente en la Introducción.

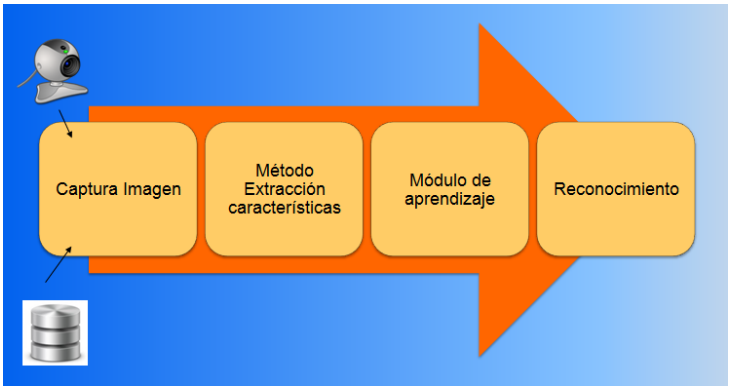


Figura 1 Pasos a seguir para el reconocimiento de una expresión.

Una vez comentado el procedimiento de forma global, se ha de destacar que este trabajo se divide en dos partes. La parte de extracción de características de la imagen realizada por Almudena Gil y el módulo de aprendizaje desarrollado por Diego Zapatero. Con el fin de llegar a la detección de la emoción, se unen ambos trabajos como se explicará posteriormente.

3.1. Método de extracción de características

Tras el análisis de las etapas que intervienen en el reconocimiento de expresiones, a continuación se procede a desarrollar la parte de extracción de características.

En primer lugar se detecta la cara y se localizan las regiones boca, ojos y cejas. El siguiente paso es la implementación del método de extracción de características, en el que se realiza una detección de puntos de interés de la cara tanto para la imagen neutral como para la imagen que se pretende analizar.

En las investigaciones anteriores el número de puntos relevantes detectados en la cara era doce, para ello se dividía la cara en regiones de interés (boca, ojos y cejas) y se usaba la detección de bordes *Canny* y el ajuste de contraste, todo esto se explica en el trabajo realizado anteriormente por alumnos en prácticas en el CITSEM [1].

Con el estudio de este trabajo previo, se detectó que existían algunos problemas en la identificación de las emociones (como por ejemplo: número de puntos de interés en la cara insuficientes, confusión en la detección de ciertas emociones) que se han tratado de solucionar con el aumento de puntos de interés en la cara, pasando de doce a diecinueve puntos. Los nuevos puntos introducidos son: los puntos intermedios de las cejas, el punto de la nariz y los puntos del párpado inferior y superior. Estos puntos se detectan de la misma manera que en el caso de los doce anteriormente detectados. Como se muestra en la Figura 2.

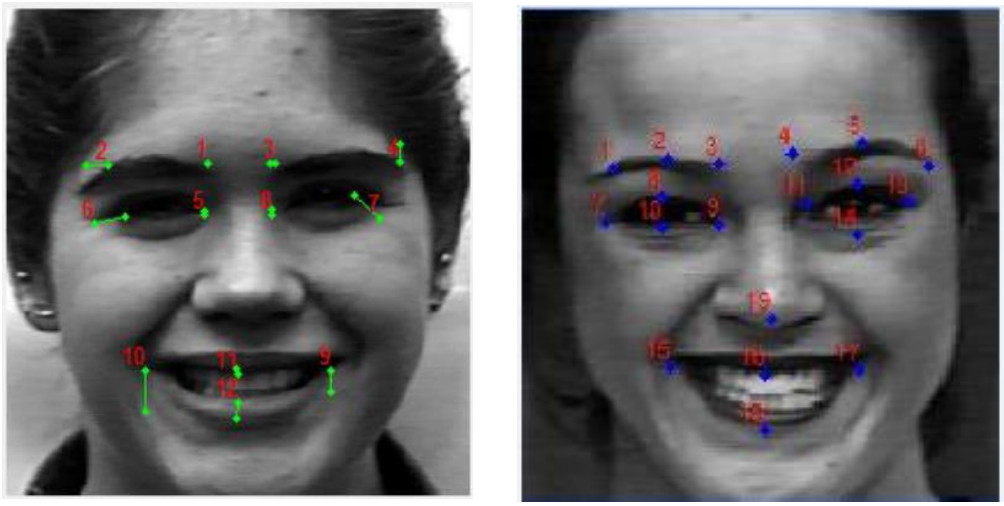
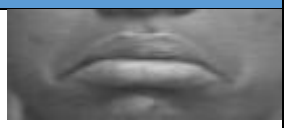
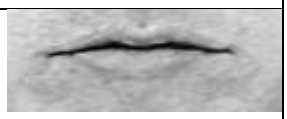
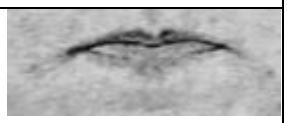
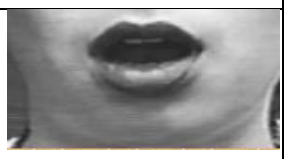
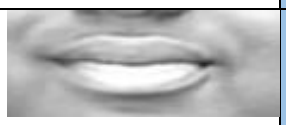


Figura 2 Detección de doce puntos (izquierda) y diecinueve puntos (derecha).

Posteriormente, tras el estudio del estado del arte de métodos de extracción de características, se ha seleccionado un nuevo método a implementar. Este método es el *Facial Action Coding System* (FACS) [11] que asocia cambios que se producen en la cara con acciones de músculos. A estas acciones se las denomina *Action Units* (AUs) y se pueden referir a la acción de un único musculo o al de un conjunto de ellos, según el FACS existen 46 Unidades de Acción. En la Tabla 1 se muestran las *Action Units* que más se utilizan según el FACS, cabe destacar que algunas de éstas se codifican según cinco niveles de intensidad como por ejemplo la AU 25, 26 y 27.

Tabla 1 *Action Units* según las FACS.

AUs (FACS)	Denominación	Imagen	AUs (FACS)	Denominación	Imagen
1	Interior de las cejas elevado		**15**	Comisuras de los labios hacia abajo	
2	Exterior de las cejas elevado		**16**	Labio inferior hacia abajo	
4	Cejas bajadas		**17**	Barbilla elevada	
5	Párpado superior elevado		**20**	Labios estrechados y estirados en horizontal (comisuras hacia los lados <- ->)	

6	Mejillas elevadas		**22**	Labios crateriformes (comisuras hacia dentro, labios abiertos)	
7	Párpados tensos		**23**	Labios tirantes, tensos	
9	Nariz arrugada		**24**	Labios presionados	
10	Labio superior elevado		**25**	Labios separados	
11	Nasolabial pronunciado		**26**	Boca entreabierta	
12	Comisuras de la boca estiradas y elevadas		**27**	Boca abierta	

A continuación, se relacionan estas *Action Units* con cada una de las expresiones que se pretenden detectar, es decir las AUs que están presentes en cada una de las emociones según el FACS, como se muestra en la Tabla 2, y se detecta que existen incoherencias entre ciertas Action Units y las expresiones con las que se asocian. Por ejemplo, en la expresión *disgust*, según el FACS interviene el AU 1 (interior de las cejas elevado), pero según propias observaciones esto no sucede así, como se puede apreciar en la Figura 3.

Tabla 2 Correspondencia AUs para cada expresión según las FACS

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

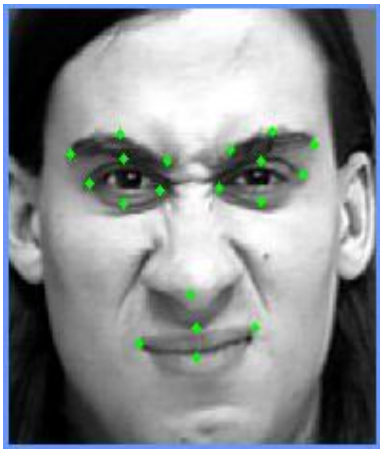


Figura 3 Imagen con la emoción de *disgust*.

Por ello, al establecer las *Action Units* que se van a implementar, se decide descartar algunas por no estar presentes en ninguna de las seis emociones que se desean reconocer y crear otras nuevas que son significativas. Al asignar los diecinueve puntos a las *Action Units* se notó que algunas de las *Action Units* definidas en FACS no eran óptimas para los puntos localizados en la cara. Finalmente se establecen veintiuna *Action Units* entre las más significativas de las FACS más dos propias (nuevas) como se puede ver en la Tabla 3. En esta tabla se observa que hasta la *Action Unit 12* son iguales que las FACS pero a partir de esta cambian, numerándose seguidamente las *Action Units* utilizadas en la herramienta.

Por ejemplo: el AU 2 (exterior de las cejas elevado) de la herramienta creada es coincidente con el método FACS y en él están presentes los puntos 1 y 6 de los 19 detectados, en cambio el AU 3 de dicha herramienta no existe en las FACS y se le ha denominado interior de las cejas bajado, con los puntos 3 y 4 presentes.

Tabla 3 Implementación de AUs para la herramienta creada

AUs (Programa)	AUs (FACS)	Denominación AUs	Puntos (de 19)
1	1	Interior de las cejas elevado	3,4
2	2	Exterior de las cejas elevado	1,6
3	-	Interior de las cejas bajado (INVENTADO)	3,4
4	4	Cejas bajadas	2,5
5	5	Párpado superior elevado	8,12
6	6	Mejillas elevadas =ojos entrecerrados	8,10,12,14
7	7	Párpados tensos	10,14
8	-	Cejas subidas (INVENTADO): AU1+AU2	2,5
9	9	Nariz arrugada	19
10	10	Labio superior elevado	16
11	11	Nasolabial: Labio superior elevado> 10 pixeles	16
12	12	Comisuras de la boca estiradas y elevadas	15,17
13	15	Comisuras de los labios hacia abajo	15,17
14	16	Labio inferior hacia abajo (0-5 pixeles)	18

15	20	Labios estrechados y estirados en horizontal (comisuras hacia los lados <- ->)	15,17
16	22	Labios crateriformes (comisuras hacia dentro, labios abiertos)	15,16,17,18
17	23	Labios apretados y encogidos (comisuras hacia dentro -> <-)	15,17
18	24	Labios comprimidos (superior e inferior)	16,18
19	25	Labio inferior hacia abajo (5-10 pixeles)	18
20	26	Labio inferior hacia abajo (10-30 pixeles)	18
21	27	Labio inferior hacia abajo (>30 pixeles)	18

En la Tabla 4 se puede ver como se han relacionado las veintiuna *Action Units* definidas con las emociones que se pretenden detectar. De esta manera se definen las *Action Units* activas en cada emoción para llevar a cabo su posterior clasificación de la expresión.

Tabla 4 Correspondencia de AUs (programa) para cada emoción

Emoción	AUs (Programa)
Fear	1,8,13,15,20
Anger	3,4,13,14,15,17,18
Sadness	1,13
Disgust	2,3,4,11,14,17,18
Happy	3,4,12,15,20
Surprise	8,21

En la Tabla 5 se puede ver de una manera más rápida y visual la comparación de las *Action Units* que intervienen en cada emoción según las FACS, correspondiendo la **X** a lo establecido según las FACS y * a lo establecido para la herramienta creada.

Tabla 5 Comparación de AUs según lo establecido por las FACS y por la herramienta desarrollada

AUs	Anger	Disgust	Fear	Happy	Sadness	Surprise
1		X	X *		X *	X
2		*	X		X	X
3						
4	X *	X *	X	*	X	
5	X		X			X
6				X	X	
7	X					
8						
9		X				
10	X	X				
11		*			X	
12				X *		
13						
14						
15	X *	X	*		X *	
16	*	X *				

17	X	X			X	
18						
19						
20	*		X *	*		
21						
22	X					
23	X *	*				
24	X *	*				
25	X	X	X	X		X
26		X	X *	*		X
27	X		X			X *

Antes de desarrollar la relación entre puntos fáciles y *Action Units*, se lleva a cabo la normalización de las imágenes entrantes, que proporcionan caras de diferente tamaño, inclinación y con diferente aspecto ya que cada persona posee diferentes distancias de ojos, boca etc. Para realizar un proceso exacto de reconocimiento se necesita restringir las zonas de movimiento, lo cual solo se consigue con una normalización. Esto se realiza ya que se desea eliminar la inclinación de las caras de ciertas imágenes presentes en la base de datos.

Se ha optado por crear una transformación afín que abarca el escalado, la rotación y desplazamientos. Para ello se extrae un triángulo (denominado triangulo de normalización) formado por los puntos exteriores de los ojos y el punto de la nariz asignándole unos valores fijos, y a su vez, se crea el triángulo equivalente de la imagen de entrada (imagen a analizar). Una vez obtenidas estas 6 coordenadas, se calculan los parámetros necesarios que permiten transformar toda la imagen de entrada según los valores del triángulo de normalización:

$$y = Ax + b \tag{1}$$

En la Figura 4 se puede ver una imagen en la cual se muestran los puntos iniciales de la imagen de entrada (en verde) y los puntos que se obtienen tras la normalización (en rojo), así como el triangulo de normalización (en morado).

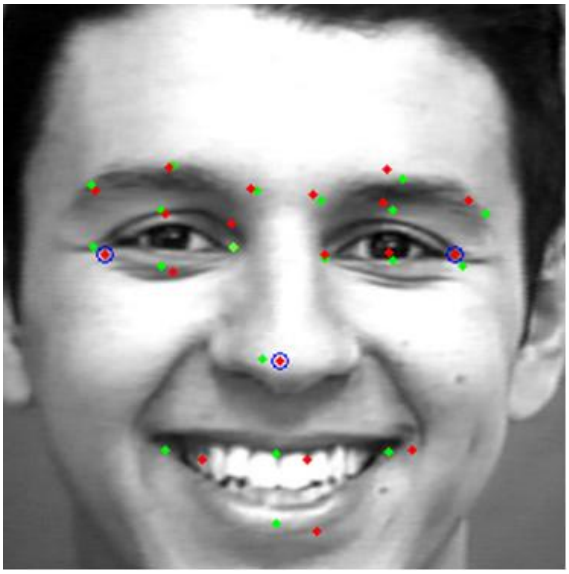


Figura 4 Triángulo de normalización (morado), puntos iniciales (verde) y finales (rojo) tras aplicar la transformación afín.

3.2. Módulo de aprendizaje

Para poder aplicar las *Action Units* en el modelo de aprendizaje para la clasificación de la emoción, se crea una matriz de entrenamiento (de unos y ceros) con todas las imágenes presentes en la base de datos, sus respectivas *Action Units* (un "1" significa que una AU está presente en la imagen y un "0" que no lo está) y la emoción correspondiente asociada (cada una de ellas identificada por un número del 1 al 6, siendo por ejemplo el '1' *happiness*).

El algoritmo de aprendizaje utilizado en este trabajo para la clasificación de emociones es un árbol de decisión.

3.2.1 Introducción

Un árbol de decisión es un modelo de aprendizaje utilizado en el ámbito de la inteligencia artificial. Dada una base de datos se construye un diagrama de decisiones secuenciales que llevan a distintas decisiones. Es una técnica para ayudar a realizar elecciones adecuadas entre muchas posibilidades, ya que ayudan a tomar la decisión "más acertada", desde un punto de vista probabilístico.

Parecido a un árbol natural, el árbol de decisión se compone de una raíz y un tronco, de unas ramas y de unas hojas. En la Figura 4 se puede observar la estructura gráficamente.

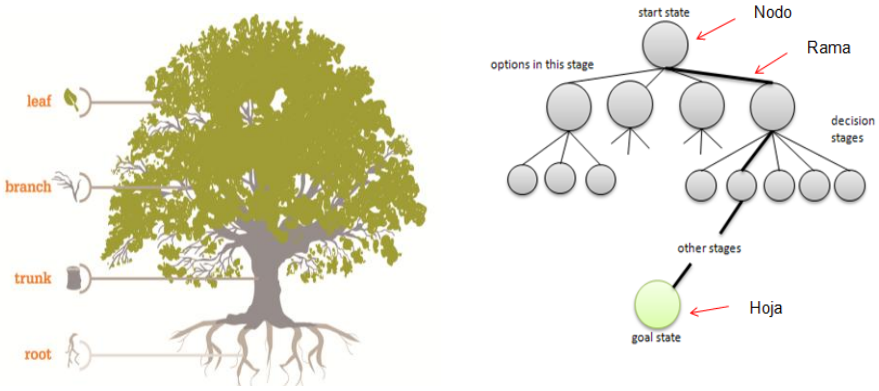


Figura 5 Estructura árbol de decisión: árbol natural (izq.), árbol de decisión (dcha)

La raíz del árbol es el nodo principal (en un árbol natural la raíz se sitúa en la parte baja del mismo, mientras que en el árbol de decisión el nodo principal se sitúa en la parte superior), que lleva la *Action Unit* más decisiva. De él parten las ramas principales, que llevan a otros nodos secundarios y así sucesivamente hasta llegar a las hojas, que serían los posibles resultados.

Como ya se ha comentado anteriormente, el módulo de aprendizaje (en este caso para crear el árbol de decisión) necesita una matriz con todas las imágenes de la base de datos y sus respectivas *Action Units*, así como un vector de resultados que asocie la emoción real con cada imagen de la base de datos. Después, a partir de esta matriz, se crea el árbol de decisión. En la Tabla 7 se puede observar un ejemplo sencillo de lo que sería esta matriz:

Tabla 6 Ejemplo sencillo matriz de datos

		Action Units				**Emoción real**
		AU1	AU2	AU3	…	
IMÁGENES	1	1	0	1		Happyness
	2	0	1	1		Sadness
	3	1	1	1		Disgust
	4	0	0	1		Anger
	5	1	1	0		Fear
	…					

3.2.2 Algoritmo

Para crear este árbol de decisión se ha utilizado el algoritmo ID3, por ser uno de los más conocidos y utilizados debido a su sencillez frente a otros. Se ha de analizar cada una de las emociones por separado (en nuestro caso 6 emociones, lo que implica 6 árboles diferentes), ya que este algoritmo sólo devuelve un valor binario, es decir, solución positiva o solución negativa.

Para la implementación de este algoritmo se ha usado como base el ejercicio de un curso sobre Machine Learning del Imperial College London [23].

El pseudocódigo base es el siguiente:

```
function DECISION-TREE-LEARNING(examples,attributes,binary_targets) returns a decision tree for a given target
label
        if all examples have the same value of binary_targets
        then return a leaf node with this value
        else if attributes is empty
            then return a leaf node with value = MAJORITY-VALUE(binary_targets)
        else
            best_attribute ← CHOOSE-BEST-DECISION-ATTRIBUTE(examples,attributes, binary_targets)
            tree ← a new decision tree with root as best_attribute
            for each possible value v_i of best_attribute do        (note that there are 2 values: 0 and 1)
                add a branch to tree corresponding to best_attibute = v_i
                {examples_i , binary_targets_i}← {elements of examples with best_attribute = v_i and the
                                                  corresponding binary_targets_i }
                if examples_i is empty
                then return a leaf node with value = MAJORITY-VALUE(binary_targets)
                else subtree ← DECISION-TREE-LEARNING(examples_i ,attributes-{best_attribute}, binary_targets_i)

        return tree
```

Figura 6 Peudocódigo para implementar el algoritmo del árbol de decisión

El algoritmo ID3 crea un árbol de decisión con una estructura descendente. La idea general de este algoritmo es identificar cuáles son los atributos (*Action Units*) más importantes y que tienen mayor influencia, o sea, aquel que posea el mayor poder discriminatorio para dicho conjunto e ir creando subconjuntos a partir de dicho atributo. Cuanta más importancia tenga un atributo, más cerca de la raíz del árbol se encontrarán. Para calcular la influencia de los atributos, se analiza cada atributo estadísticamente para saber cuánta información aporta a la hora de clasificar una emoción.

Un concepto muy importante para ese análisis estadístico es la Entropía (E), que determina la cantidad de incertidumbre de un determinado conjunto de ejemplos:

$$E\ (S) = -\sum_{i=1}^{N} p_i log_2\ p_i\ , \tag{2}$$

siendo 'S' un conjunto de ejemplos, 'N' es el número total de ejemplos,' p_i ' la probabilidad de los posibles valores de 'i', o lo que es lo mismo, si 'i' puede valer '0' ó '1', la entropía se calcula de la siguiente manera:

$$E\ (p,n) = -\frac{p}{p+n} log_2 \left(\frac{p}{p+n}\right) - \frac{n}{p+n} log_2 \left(\frac{n}{p+n}\right)\ , \tag{3}$$

donde 'p' es el número de ejemplos positivos y 'n' el número de ejemplos negativos.

A mayor entropía, habrá mayor incertidumbre, es decir, más difícil le será al árbol tomar una decisión. En general, un atributo que puede ayudar a discriminar más ejemplos, tiende a reducir más la entropía, y por tal motivo, debe ser seleccionado como un nodo de selección para la siguiente subdivisión.

Sin embargo, en general se suele tener en cuenta otro parámetro más preciso para seleccionar el mejor atributo: la *ganancia de información*:

$$Ganancia\ (atributo) = E\ (p, n) - Resto\ (atributo), \tag{4}$$

que mide lo bien o mal que un atributo separa los ejemplos. A mayor ganancia de información más importante será el atributo en cuestión.

El algoritmo utiliza esta ganancia calculada para ir eligiendo los mejores atributos en cada paso, crear subconjuntos e ir construyendo el árbol descendentemente.

El Resto de un atributo calcula la entropía de cada rama, es decir, de cada subconjunto y realiza una suma proporcional para calcular la entropía del total:

$$Resto(p,n) = \frac{p_0+n_0}{p+n} log_2 \left(\frac{p_0+n_0}{p+n}\right) * E(p_0, n_0) + \frac{p_1+n_1}{p+n} log_2 \left(\frac{p_1+n_1}{p+n}\right) * E(p_1, n_1), \tag{5}$$

donde 'p_0' el número de ejemplos positivos dentro del subconjunto para el cual el atributo tiene valor '0', 'n_0' el número de ejemplos negativos dentro del subconjunto para el cual el atributo tiene valor '0', 'p_1' el número de ejemplos positivos dentro del subconjunto para el cual el atributo tiene valor '1' y 'n_1' el número de ejemplos negativos dentro del subconjunto para el cual el atributo tiene valor '1'.

La explicación más detallada con ejemplos incluidos se puede encontrar en [24].

Así, de forma general se podría decir que para crear el árbol de decisión este algoritmo realiza los siguientes pasos:

- A partir de todo el conjunto de ejemplos, se selecciona el mejor atributo entre todos, que corresponde al nodo principal.

- Una vez elegido, se subdivide el conjunto en dos subconjuntos: uno correspondiente a todos los ejemplos cuyo atributo sea '1' y otro correspondiente a todos los ejemplos cuyo atributo sea '0'

- Para cada uno de los subconjuntos se vuelven a realizar los mismos pasos hasta que:

    - Todos los ejemplos sean iguales, es decir, todos sean positivos o negativos (en este caso la hoja o terminación del árbol tomaría dicho valor)

    - Se hayan analizado todos los atributos (en cuyo caso el valor de la hoja o terminación del árbol estaría determinado en función de si existe un mayor número de ejemplos positivos o negativos para dicha expresión). Un detalle importante a tener en cuenta es

que los atributos que vayan seleccionándose como mejores atributos de cada subconjunto se van descartando para las siguientes subdivisiones inferiores ya que no tiene sentido que se vuelva a analizar.

Tras la implementación de este algoritmo se crearía el árbol de decisión con una estructura como la de la Figura 7. En ella, se puede observar el árbol de decisión de la expresión *happyness*. Los '0' y '1' corresponden a los valores de las hojas del árbol, mientras que el resto de números corresponden a los distintos atributos, siendo por ejemplo el 12 el mejor atributo y por tanto el nodo principal del árbol.

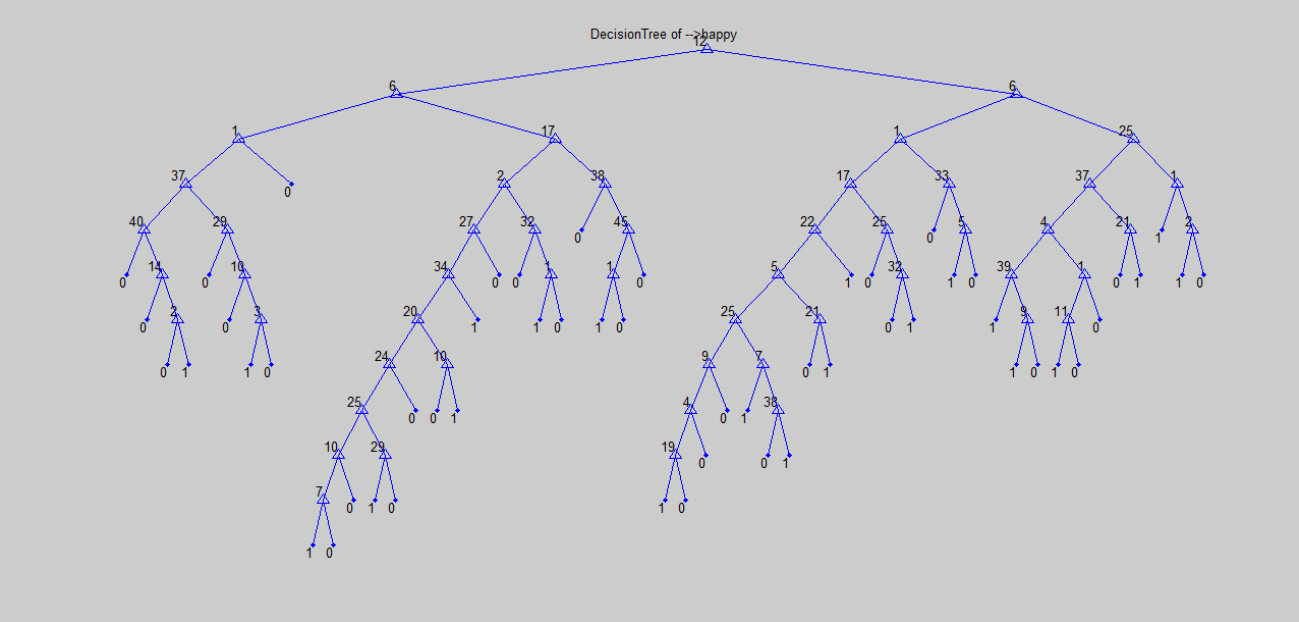


Figura 7 Ejemplo árbol de decisión para la expresión *happyness*

3.2.3 Validación cruzada

Para realizar las pruebas y comprobar con qué precisión la herramienta es capaz de clasificar las emociones se ha utilizado la técnica de validación cruzada, que consiste en utilizar una parte del conjunto total para entrenar y crear el árbol y el resto para poner a prueba ese árbol creado.

Se divide el conjunto de datos en K subdivisiones, de tal manera que se utiliza una subdivisión para pruebas y K-1 subdivisiones para el entrenamiento. Este proceso es repetido K veces, una vez por cada subdivisión. En la Figura 8 se muestra un esquema de este proceso si por ejemplo K=10:

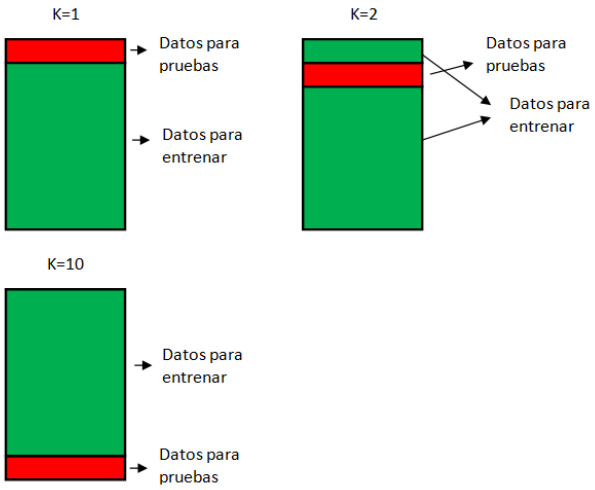


Figura 8 Proceso de validación cruzada con K subdivisiones

Como se ha comentado anteriormente, es necesario crear un árbol por emoción. De esta manera, cada imagen se probaría con cada árbol con la esperanza de que solamente un árbol de un resultado positivo. Como esto no sucede en muchos casos, se debe combinar los 6 árboles para lograr una solución única. Este proceso se comentara más adelante.

Para cada una de las K veces que se tiene que repetir este proceso, una vez que se crean los 6 árboles a partir del conjunto de entrenamiento, se analizan una a una las imágenes del conjunto de pruebas.

Así, el primer paso consiste en obtener todos los valores de los atributos pertenecientes a cada imagen. Una vez extraídos, se analizan para cada uno de los árboles todos los atributos de la imagen.

Primero se localiza a partir del árbol creado anteriormente cuál es el atributo del nodo principal y se busca cuál es el valor de dicho atributo en la imagen de entrada. Si el atributo o *Action Unit* está presente en la imagen habrá un '1' y por tanto se encaminará por la rama correspondiente, pasando a otro subconjunto donde se comprobará de nuevo cuál es el mejor atributo y así sucesivamente hasta que se llegue a una solución es decir, a una hoja del árbol. Sin embargo, si por el contrario el atributo no está presente en la imagen el proceso se encaminará por la rama correspondiente al '0' y seguiría los mismos pasos descritos para el otro caso.

Cuando se ha analizado la imagen de entrada con los 6 árboles, se obtienen 6 soluciones binarias que hay que intentar convertir en solución única. Si solamente hay un árbol que haya dado una solución positiva (lo ideal), la emoción correspondiente a dicho árbol será la solución final. Si esto no ocurre, se pensaron varias posibles estrategias:

- Tener en cuenta como mejor solución el árbol que antes llegue a la solución positiva, es decir, el árbol que haya analizado menos atributos.

- Tener en cuenta como mejor solución el árbol que más atributos positivos haya analizado, ya que indica que un mayor número de unidades de acción estaría presente en la imagen y por tanto mayores posibilidades de que sea la emoción real.

Se hicieron pruebas con las dos estrategias obteniendo resultados parecidos, por lo que se pensó en combinar las dos estrategias para obtener mejores resultados. De tal manera que primero se analizaría cuál es el árbol que ha analizado una menor cantidad de atributos, y si dos árboles coinciden entonces se pasaría a investigar cuál de esos dos árboles tiene mayor cantidad de *Action Units* presentes.

Por otro lado, tras la implementación de este algoritmo y la realización de las pruebas con la base de datos proporcionada por ejercicio en el cual se basó todo este proceso, se procedió a implementar un árbol de decisión con las funciones que aporta el programa *Matlab* para ver las diferencias y cuál de los dos era más adecuado. Dichas pruebas se encuentran en el apartado de Resultados.

3.3. Interfaz utilizada

Para poder realizar las distintas pruebas y obtener los resultados se ha utilizado como base la interfaz gráfica implementada por los alumnos en prácticas en el CITSEM durante el periodo de otoño de 2014 [1].

A partir de ella se han desarrollado e implementado nuevas mejoras, con el uso de la herramienta *GUIDE* de Matlab, que se comentan a continuación:

- Se modifica el aspecto visual de la interfaz, tanto de colores como de organización.
- Se añaden botones para poder seleccionar el idioma (Inglés o Español), así como un botones de información para poder ayudar al usuario.
- Un control con el cual se permite seleccionar el método de extracción de características:
    - 12 puntos
    - 19 puntos
    - 12 puntos + normalización
    - 19 puntos + normalización
- Un control para el método de clasificación de la emoción, con dos opciones posibles:
    - Matriz de desplazamiento
    - *Action Units* + árbol de decisión
- Se incorpora un módulo de pruebas donde se permite al usuario, seleccionar entre el uso de cinco o todas (seis) las emociones para obtener una matriz de resultados como las que se muestran en el apartado de resultados.

En la Figura 9 se puede observar la interfaz actual y la realizada por antiguos compañeros, apreciando claramente los cambios mencionados antes.

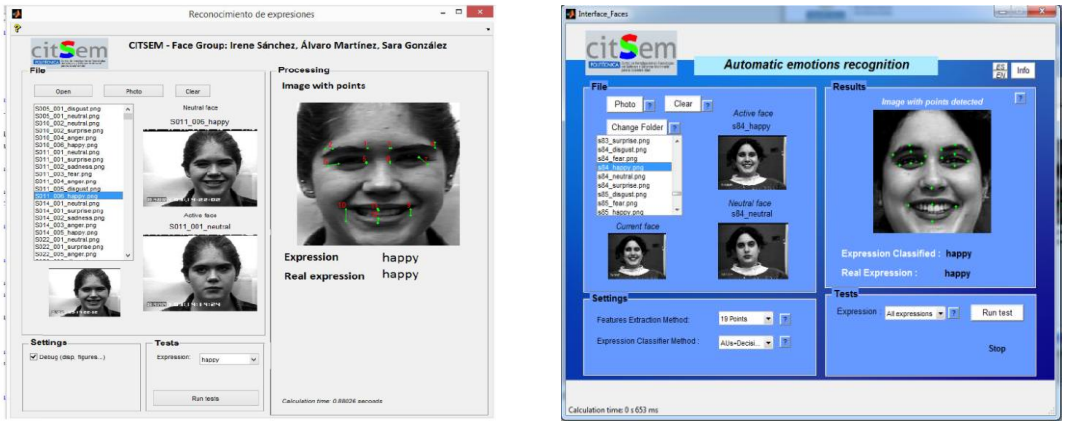


Figura 9 Interfaz gráfica antigua (izquierda) y actual (derecha)

4. Resultados

Para realizar las pruebas se utiliza la base de datos *CohnKanade+* [25], que contiene imágenes de 210 sujetos, tanto hombres como mujeres y de distintas nacionalidades, que expresan las seis emociones básicas que se quieren detectar, más la imagen neutral y una expresión llamada *Contempt* (desprecio).

Además se realiza la búsqueda de una nueva base de datos con el objetivo de ampliar el número de sujetos y emociones a analizar debido a que la base de datos *CohnKanade+* se queda corta y no es adecuada al haber diferente número de imágenes para cada emoción como se observa en la Tabla 7, influyendo en los resultados obtenidos. Por ejemplo para la expresión *surprise* hay 81 imágenes y para *sadness* solo 27. Una de las bases de datos encontrada es la llamada MMI [26], que es usada por algunos investigadores, pero finalmente se descarta ya que después de su análisis y organización no se considera adecuada al tener presente muchas emociones que no se han podido clasificar y reconocer.

Tabla 7 Número de imágenes de la base de datos CohnKanade+

Emoción	**Número de imágenes en la *CohnKanade+***
Fear	24
Anger	45
Sadness	27
Disgust	58
Happy	68
Surprise	81
Contempt	6
Neutral	105
Total de imágenes	414

Para llevar a cabo la obtención de resultados se utiliza la interfaz gráfica, con la cual se puede seleccionar la expresión que se pretende analizar así como la realización de matrices de confusión, que permiten analizar los porcentajes de acierto y de fallo en la detección de la emoción.

En primer lugar se realizan una serie de pruebas con las *Action Units* establecidas por la herramienta creada para clasificar la expresión, antes de la implementación del árbol de decisión como módulo de aprendizaje. Con estas pruebas se pretende averiguar qué *Action Units* son válidas o no para cada emoción, es decir si nos sirven o no para detectar la emoción de las imágenes de la base de datos.

Una vez se ha seleccionado cuáles son las *Action Units* son más relevantes para cada una de las emociones, como se muestra en la Tabla 4 anterior, se definen unos contadores (uno para cada emoción) de tal manera que si una AU que se considera relevante para una emoción concreta está activa (es válida), se incrementa en 1 el valor de dicho contador. Finalmente, se evalúa cuál de los 6 contadores es mayor y esa será la expresión detectada (*happiness*, *surprise*, *disgust*, *sadness*, *anger* o *fear*).

En la Tabla 8 se pueden ver los resultados obtenidos tras la realización de la prueba anteriormente comentada, sin el uso del árbol de decisión. Esta tabla muestra los porcentajes de acierto (en verde) de cada expresión y los porcentajes de confusión (en blanco) entre cada una de las emociones. Por ejemplo: se analizan todas las imágenes de la base de datos con la emoción *happy* y se obtiene un 50% de aciertos, un 13,24 % que se confunden con la emoción *surprise*, un 19,12 % con *sadness*, un 4,41 % con *disgust*, 10,29 % con *anger* y un 2,94 % de las imágenes *happy* analizadas es confundido con la expresión *fear*.

Tabla 6 Matriz de confusión, sin el uso del árbol de decisión.

expresion Analizada-> expresion Detectada	Happy	Surprise	Sadness	Disgust	Anger	Fear
Happy	50,00	0,00	7,41	18,97	13,33	8,33
Surprise	13,24	87,65	14,81	20,69	11,11	16,67
Sadness	19,12	1,23	62,96	5,17	26,67	33,33
Disgust	4,41	1,23	3,70	43,10	11,11	0,00
Anger	10,29	1,23	3,70	10,34	31,11	16,67
Fear	2,94	8,64	7,41	1,72	6,67	25,00

Como se puede ver los porcentajes de acierto más altos se obtienen para las expresiones *happy*, *surprise* y *sadness* y unos resultados peores para *disgust*, *anger y fear* al confundirse unas emociones con otras y ser las más difíciles de detectar.

Cabe destacar que hasta llegar a la obtención de los resultados que se pueden ver en la Tabla 6, se realizaron múltiples pruebas y combinaciones de *Action Units* para cada expresión, mostrándose aquí sólo ésta por considerarse la más relevante al obtener mejores porcentajes de acierto.

En segundo lugar, se realizaron pruebas con el árbol de decisión utilizando la base de datos proporcionada por [24]. Como se ha comentado en el apartado 3.2.3, las pruebas se realizaron en base a una validación cruzada y por tanto se usaron una serie de ejemplos para entrenar y otra serie

de ejemplos para hacer las pruebas. Este proceso se realiza 10 veces, y los resultados obtenidos en la Tabla 9 son la media de cada una de esas veces que se han realizado las pruebas.

Tabla 7 Matriz de confusión usando árbol de decisión y base de datos proporcionada

		Real Emotion					
		Anger	Disgust	Fear	Happyness	Sadness	Surprise
Predicted Emotion	Anger	45,00	5,00	7,00	1,00	9,00	0,00
	Disgust	18,00	75,00	5,00	5,00	17,00	4,00
	Fear	6,00	2,00	63,00	3,00	4,00	20,00
	Happyness	5,00	7,00	3,00	75,00	5,00	2,00
	Sadness	21,00	9,00	8,00	7,00	58,00	5,00
	Surprise	5,00	3,00	14,00	9,00	7,00	67,00

Observando los resultados, se puede comprobar que salvo en la expresión *Anger* los porcentajes superan el 50% en todas las demás, pero están aún muy lejos del ideal 100%, por lo que aún tiene margen de mejora.

Como paso siguiente, se realizan de nuevo las mismas pruebas pero ahora utilizando la funciones que proporciona Matlab, Tabla 8, para comparar las dos matrices de confusión y ver las diferencias obtenidas.

Tabla 8 Matriz de confusión usando árbol de decisión, base de datos proporcionada y funciones Matlab

		Real Emotion					
		Anger	Disgust	Fear	Happyness	Sadness	Surprise
Predicted Emotion	Anger	64,00	4,00	5,00	1,00	8,00	1,00
	Disgust	8,00	80,00	1,00	2,00	6,00	2,00
	Fear	4,00	3,00	76,00	0,00	2,00	9,00
	Happyness	2,00	5,00	4,00	94,00	3,00	2,00
	Sadness	20,00	8,00	5,00	1,00	76,00	4,00
	Surprise	2,00	1,00	8,00	1,00	6,00	82,00

Comparando la Tabla 7 y la Tabla 8, se pueden apreciar diferencias en los porcentajes, siendo mayores cuando se aplican las funciones de Matlab, pero se puede observar también que la relación es parecida; el porcentaje de acierto en la expresión *Anger* también es el menor en de todos los porcentajes, y que el porcentaje de aciertos en la expresión *Happyness* es el mayor en ambos casos. Por contrapartida, existe un aumento de aciertos en la expresión *Surprise* al usar las funciones de Matlab, siendo incluso mayor que el porcentaje de la expresión *Disgust*, un hecho que con el algoritmo implementado no sucede, ya que el porcentaje de *Disgust* es mayor que el de *Surprise*.

Debido a estas diferencias, se decide utilizar la implementación con las funciones de Matlab para combinar la matriz creada a partir de las *Action Units* utilizadas y creadas como método de extracción de características con el árbol de decisión como método de aprendizaje. Los resultados de combinar estos dos métodos se pueden observar en la Tabla 9.

Tabla 9 Matriz de confusión usando el árbol de decisión y la base de datos CohnKanade+

expresion Analizada-> expresion Detectada	Happy	Surprise	Sadness	Disgust	Anger	Fear
Happy	78,00	10,00	17,65	14,85	36,00	8,08
Surprise	6,00	73,00	29,41	10,89	11,00	8,08
Sadness	1,00	3,00	37,25	10,89	0,00	8,08
Disgust	12,00	8,00	7,84	63,37	15,00	25,25
Anger	3,00	3,00	7,84	0,00	33,00	17,17
Fear	0,00	3,00	0,00	0,00	5,00	33,33

Los resultados que se obtienen son mejores para las emociones de *happiness*, *surprise* y *disgust*, y peores para *sadness*, *anger* y *fear*. Esto se debe a que estas últimas expresiones se confunden entre sí en muchos casos.

En general los resultados de esta tabla no son tan buenos como los de la Tabla 9 y la Tabla 10 debido a que estas tablas se han obtenido a partir de la base de datos y el método de extracción de características (también basado en *Action Units*) proporcionada por [24].

5. Conclusiones

Tras la investigación que se ha realizado en estos meses y tras el estudio del estado del arte, se pueden sacar las siguientes conclusiones:

Por un lado se procede a comentar los logros obtenidos tras la realización de esta investigación:

- Con la introducción de los diecinueve puntos de interés en la cara se mejora la detección de algunas de las emociones que se pretenden identificar.

- Con la normalización de las imágenes se elimina la inclinación de algunas de las imágenes que están presentes en la base de datos que se utiliza en esta investigación [18], ya que esto supone un problema a la hora de llevar a cabo el reconocimiento facial de una emoción.

- Gracias a la introducción del árbol de decisión como método de aprendizaje se ha conseguido optimizar el método usado hasta entonces, ya que ayuda a realizar las mejores decisiones con la información existente. Además, permite que todas las opciones sean analizadas, pudiendo incluso analizar las consecuencias de llevar a cabo una alternativa gracias a la fácil interpretación de la decisión tomada.

- Con el uso de las *Actions Units* y el árbol de decisión se reconoce la expresión de una manera más sencilla que con la matriz de diferencias que se implementaba anteriormente en el trabajo realizado en el CITSEM [1].

Por otro lado se mencionan algunos de los problemas encontrados:

- Surgen problemas en la detección de los puntos cuando en la imagen aparecen sombras y personas con gorras o gafas. Además, el método utilizado no es muy robusto a diferentes iluminaciones y colores de piel de los sujetos.

- La base de datos utilizada *CohnKanade+* [25], como ya se ha comentado anteriormente no tiene un número adecuado de imágenes.

- Se produce confusión entre las emociones *sadness, anger* y *fear* en el reconocimiento de la emoción al ser las 3 expresiones que más similitudes tienen y por tanto más complicadas de distinguir.

Por último cabe comentar que se han cumplido los objetivos establecidos durante el desarrollo de las prácticas, aunque al tratarse de un trabajo de investigación siempre existen posibles mejoras que se pueden realizar.

6. Trabajo futuro

Así pues, de cara al trabajo futuro se procede a mencionar algunos avances y mejoras:

- Mejora de la eficiencia del método utilizado, para obtener un reconocimiento a tiempo real.

- Ampliar la base de datos, para obtener la misma cantidad de imágenes para cada una de las expresiones a detectar.

- Mejorar la relación entre puntos característicos de la cara y las AUs.

- Desarrollo e implementación de otros métodos de extracción de características y modelos de aprendizaje, como puede ser el uso de redes neuronales para ver si producen mejoras y se consiguen mejores resultados que utilizando el árbol de decisión.

- Detección de expresiones espontáneas y micro-expresiones, ya que las imágenes presentes en la base de datos utilizada [25] posee imágenes con emociones muy exageradas.

- Expandir la herramienta a secuencias de vídeo analizando el movimiento en las expresiones.

- Combinación de los métodos y modelos desarrollados con la voz, permitiendo así un reconocimiento más efectivo de las emociones.

7. Referencias

[1] Sara González, Álvaro Martínez, "Interfaces Naturales para Realidad Aumentada", Junio 2014

[2] R. N. Rojas Bello, "Identificación de características relevantes para reconocimiento de emociones en el rostro," 2009.

[3] L. Blazquez, "PFC-Reconocimiento Facial Basado en Puntos Característicos de la Cara en entornos no controlados," 2013.

[4] P. Tome and L. Bl, "UNDERSTANDING THE DISCRIMINATION POWER OF FACIAL REGIONS IN FORENSIC CASEWORK," pp. 13–16, 2013.

[5] A. Saeed, A. Al-Hamadi, R. Niese, and M. Elzobi, "Effective geometric features for human emotion recognition," *2012 IEEE 11th Int. Conf. Signal Process.*, pp. 623–627, Oct. 2012.

[6] D. M. Aung, "Automatic Facial Expression Recognition System using Orientation Histogram and Neural Network," vol. 63, no. 18, pp. 35–39, 2013.

[7] Essa, I.A.: "Coding, analysis, interpretation, and recognition of facial expressions". IEEE Transactions on Pattern Analysis and Machine Intelligence, 757–763 (1997).

[8] James Lien, J.-J.: "Automatic recognition of facial expressions using hidden markov models and estimation of expression Intensity". Doctoral dissertation, Robotics Institute, Carnegie Mellon University (April 1998).

[9] Mase, K.: "Recognition of facial expressions for optical flow". IEICE Transactions on Special Issue on Computer Vision and Its Applications E74(10), 3474–3483 (1991).

[10] Choudhury, T.A.: "Pentland Motion field histograms for robust modeling of facial expressions". In: Proceeding of the 15th ICIP, USA, vol. 2, pp. 929–932 (2000).

[11] J. Hamm, C. G. Kohler, R. C. Gur, and R. Verma, "Automated Facial Action Coding System for dynamic analysis of facial expressions in neuropsychiatric disorders.," *J. Neurosci. Methods*, vol.200, no. 2, pp. 237–56, Sep. 2011.

[12] S. Yang and B. Bhanu, "Understanding Discrete Facial Expressions in Video Using an Emotion Avatar Image.," *IEEE Trans. Syst. Man. Cybern. B. Cybern.*, vol. 42, no. 4, pp. 980 992, May 2012.

[13] T. Senechal, V. Rapp, H. Salam, R. Seguier, K. Bailly, and L. Prevost, "Features via Multikernel Learning," vol. 42, no. 4, pp. 993–1005, 2012.

[14] Ting Wu, Siyao Fu, and Guosheng Yang, "Survey of the Facial Expression Recognition Research", pp. 392–402, 2012

[15] C. Tanchotsrinon, S. Phimoltares, S. Maneeroj, and A. Virtual, "FACIAL EXPRESSION RECOGNITION USING GRAPH-BASED FEATURES," 2011.

[16] R. R. Londhe and V. P. Pawar, “Analysis of Facial Expression and Recognition Based On Statistical Approach,” no. 2, pp. 391–394, 2012.

[17] D. M. Aung, “Automatic Facial Expression Recognition System using Orientation Histogram and Neural Network,” vol. 63, no. 18, pp. 35–39, 2013.

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

[19] K. Yurtkan and H. Demirel, “Feature selection for improved 3D facial expression recognition,” Pattern Recognit. Lett., vol. 38, pp. 26–33, Mar. 2014.

[20] V. Bettadapura, “Face Expression Recognition and Analysis : The State of the Art,” pp. 1–27, 2012.

[21] S. Consultant and I. F. E. Databases, “A SURVEY ON FACIAL EXPRESSION DATABASES,” vol. 2, no. 10, pp. 5158–5174, 2010.

[22] N. U. Khan, “A Comparative Analysis of Facial Expression Recognition Techniques,” pp. 1262–1268, 2012.

[23] Maja Pantic, “Computer Based Coursework Manual”, Imperial College London, 2013

[24] Tom M. Mitchell, “*Machine Learning*”, McGraw-Hill Science, pp. 52-78, 1997

[25] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, I. Matthews, and F. Ave, “The Extended CohnKanade Dataset (CK +): A complete dataset for action unit and emotion-specified expression,” no. July, 2010.

[26] M. Pantic, M.F. Valstar, R. Rademaker, and L. Maat. 2005, Web-based database for facial expression analysis. In IEEE International Conference on Multimedia and Expo, pages 317–321.