

## Interfaces Naturales para Realidad Aumentada

Informe final de la asignatura "Prácticas Externas"

18 de junio de 2014

Sara González, Álvaro Martínez

Tutor(a): Martina Eckert

---

### Resumen

El reconocimiento automático de expresiones faciales es un ámbito en el que se centran muchos equipos de investigación de todo el mundo. En los últimos años gracias a la aparición del procesado digital de imágenes, ha habido muchos avances en el ámbito de la detección de la cara, extracción de características y clasificación de la expresión. En este informe, en primer lugar se realiza un análisis del estado del arte en el reconocimiento de emociones para poder saber qué métodos existen y cuáles son sus ventajas e inconvenientes. Posteriormente se implementa un método de obtención de la expresión mediante la extracción de regiones y segmentación para poder obtener 12 puntos de interés en la cara que nos permitan clasificar la expresión. Tras la obtención de estos puntos de interés se crean unas matrices que contienen información útil para poder realizar la clasificación de las emociones. Una vez implementado éste se crea una interfaz gráfica para que el manejo del programa y la realización de pruebas puedan ser accesibles por cualquier tipo de usuario. Por último se realizan pruebas de este algoritmo con diferentes parámetros y diferentes expresiones para comprobar su funcionamiento y poder determinar qué mejoras serían necesarias. Finalmente se extraen las conclusiones del trabajo y se detallan las futuras líneas de investigación con las que se puede continuar.

---

### Abstract

Automatic recognition of facial expressions is a field in which many research teams have focused. In recent years due to the appearance of digital image processing, automatic recognition of facial expressions has been an active research topic. This is the reason why face detection, feature extraction and expression classification are constantly evolving. First of all, this work surveys some of the most important techniques in automatic facial expressions recognition to give an overview of which is the most important methods currently being researched. Afterwards, an expression recognition algorithm is implemented. It focuses on extracting 12 facial feature points with help of a segmentation, that allow the expression classification. Then, a matrix, with useful information regarding the classification of the expression, is created. Besides, a user graphical interface is created to ease of use and testing of the program by any kind of user. Finally, some tests are carried out with different parameters and expressions to evaluate its operating mode, and to be able to determine which improvements would be needed. At the end the conclusions of the present work are extracted and the future works are detailed.

---

## 1. Introducción

Las expresiones faciales son respuestas a estímulos que tienen dos objetivos principales durante la comunicación social: reforzar la comunicación o engañar en la comunicación. La posibilidad de reconocer estas expresiones de forma automática lleva siendo investigada desde hace más de 30 años, aunque ha sido en esta última década donde la investigación ha experimentado un impulso, gracias al desarrollo del procesado de la imagen y los gráficos por ordenador.

Hay muchas áreas donde puede resultar muy interesante el desarrollo de aplicaciones de reconocimiento automático de emociones. Mediante estas aplicaciones se pretende mejorar la comunicación de las personas con las máquinas (*Affective Computing*). Estas tecnologías tienen numerosos ámbitos de aplicación son: robótica, videojuegos, seguridad de los automóviles, medicina, psicología, psiquiatría y rehabilitación.

Se ha desarrollado un sistema de reconocimiento de emociones basado en la localización de 12 puntos característicos de la cara, concretamente en la boca, ojos y cejas, de forma que se puedan detectar automáticamente las emociones expresadas por las personas, basándose en este caso en la expresión facial.

Los objetivos principales de este proyecto son:

1. Investigar el estado del arte en el área del reconocimiento de expresiones faciales para conocer los diferentes métodos e implementaciones existentes actualmente, así como algunas de sus posibles aplicaciones.
2. Analizar los problemas del método propuesto en el Proyecto Fin de Carrera de R. N. Rojas Bello "Identificación de características relevantes para reconocimiento de emociones en el rostro" [1] y plantear un nuevo método que mejore y optimice los resultados. Para ello se plantean determinadas modificaciones, tales como la división de la cara por regiones, el pre-procesado de la imagen para obtener segmentaciones más precisas y la extracción de puntos de interés.
3. Conforme se desarrollan los algoritmos, extraer conclusiones acerca de la utilidad de las distintas características de la cara en el reconocimiento de expresiones.
4. Evaluar la eficacia del sistema de reconocimiento de expresiones que se ha desarrollado comparando los resultados con los obtenidos por otros autores y realizando pruebas en varias bases de datos.

Se detallan a continuación los apartados en los que se divide el documento. En el siguiente apartado se realiza un análisis del estado del arte en el reconocimiento de expresiones resaltando los métodos más relevantes. Posteriormente, en el tercer apartado, se explica el método implementado para la detección y clasificación de expresiones y se describe la interfaz gráfica desarrollada. En el cuarto apartado se explican los resultados obtenidos tras probar el método en diferentes bases de datos y con diferentes parámetros. Finalmente en el quinto apartado se detallan las conclusiones extraídas y en el sexto se proponen futuras líneas de trabajo posibles en este ámbito.

## 2. Estado del Arte

Se resumen a continuación las conclusiones más importantes extraídas del análisis del estado del arte en reconocimiento de expresiones faciales. El análisis de las expresiones faciales es muy importante para mejorar la comunicación entre las personas, ya que aumenta la cantidad de datos que se intercambian durante la comunicación. Actualmente, debido a los avances en procesamiento de imágenes y gráficos de ordenador, el análisis de expresiones de forma automática esta adquiriendo gran importancia.

La mayoría de los sistemas desarrollados presenta determinadas limitaciones relacionadas con la inclinación de la cara, la variación en las condiciones de iluminación o la resolución de las imágenes; así como la dificultad al tratar las expresiones espontáneas y la posibilidad de clasificación de la expresión en tiempo real. Las investigaciones más recientes se centran en la implementación de métodos que solucionen estas limitaciones.

Existen muchos métodos de extracción de la emoción de una persona, partiendo de una imagen, y cada uno utiliza una combinación de métodos diferente para realizar los dos pasos principales: extracción y selección de características de interés y clasificación de la expresión.

Respecto a los métodos de **extracción de puntos de interés de la cara** se destacan a continuación los más utilizados. El primer método, propuesto por Ekman y Friesen, fue el FACS (*Facial Action Coding System*) [2]. Consiste en definir una serie de puntos a observar en el rostro y describir todas las actividades faciales en base los movimientos de 46 unidades de acción de diferentes partes de la cara. Este método ha evolucionado actualmente a FACES (*Facial Action Coding Expression System*), que reduce el tiempo de detección de las emociones y no necesita una codificación individual de cada músculo como FACS, ya que entrega información de frecuencia, intensidad, velocidad y duración de la expresión facial. Otro método utilizado en [1], [3], [4], [5] y [6] es el de la obtención de puntos de interés mediante el uso de teoría de grafos, detección de bordes y segmentación de regiones de la cara, en el que se ha basado la realización de las prácticas. En [7] se propone utilizar una máscara parametrizada para la codificación de caras, basándose en modelos predefinidos. Se han publicado varias versiones de esta máscara en las que varía el número de vértices, entre ellos, uno de los modelos más nuevos e integrado en las librerías publicadas para la *Kinect* es el *CANDIDE-3* [7]. El método de *Gabor Wavelets* permite extraer las características de interés de la imagen con el filtrado de *Gabor* y es también muy utilizado, como se puede ver en [8], [9], [10], [11] y [12]. Finalmente en [13] y [14] se propone el método de AAM (*Active Appearance Model*) que se basa en comparar un modelo de puntos con una nueva imagen usando una variante del algoritmo de maximización de la esperanza. Con esto se encuentran los parámetros del modelo que minimizan la diferencia entre dicha imagen y la síntesis del modelo proyectado en la imagen. Uno de los últimos métodos más utilizados en el campo de extracción de características es el de LBP (*Local Binary Patterns*) que da un valor binario a cada píxel de la imagen en función de los píxeles de su entorno, utilizado en [10], [15], [16], [17] y [18].

Respecto a los métodos de **clasificación de la expresión** se detallan a continuación los más significativos. El método propuesto en [19] y [20] de *Eigenfaces* consiste en interpretar cada imagen facial en escala de grises como un conjunto bidimensional de patrones claros y oscuros. Con esto se representan los patrones de los niveles de gris vectorialmente formando una base de vectores capaz de representar caras que tienen características comunes. El método más utilizado con diferencia es SVM (*Support Vector Machine*), cuya utilidad se puede ver en [21] y [22]. Consiste en un modelo de “*Active Learning*” que analiza los datos según se van obteniendo para poder reconocer y clasificar patrones que se repiten. Con SVM se obtiene una representación de las caras que expresan una emoción como puntos en el espacio, formando bases. Cada ejemplo se caracteriza por unos vectores en relación a las bases creadas. Otro método muy utilizado es ANN (*Active Neural Network*) [23], [24] y [6] en el que se usan redes neuronales para interpretar la expresión detectada tras obtener las características de los puntos de interés.

Dentro de los estudios más destacados sobre esta temática se podrían destacar [25], [26] y [27]. En todos ellos se detalla un extenso estado del arte para poder tener una mejor visión del desarrollo actual de las emociones faciales y además se detallan sus aplicaciones y los retos de la investigación actuales en éste ámbito. Algunos

de ellos analizan diferentes métodos realizando comparaciones entre ellos, para poder concluir cuáles son los más adecuados.

En este documento se hace también referencia al algoritmo *Viola & Jones* [28] para detección de caras. Éste algoritmo se basa en una estructura en cascada de clasificadores sencillos que permite descartar rápidamente zonas de la imagen que no contienen caras para concentrarse en las zonas de interés; esto lo convierte en un detector rápido y que ha sido la elección de muchos autores. Este tipo de clasificadores pueden ser entrenados para detectar otro tipo de objetos, por ejemplo, los ojos.

### 3. Desarrollo de sistemas de reconocimiento de expresiones faciales

Se ha observado que hay varios métodos diferentes de análisis de expresiones faciales y que la mayoría no permite el procesado en tiempo real, debido a la gran cantidad de datos a procesar. Se ha optado por desarrollar un método basado en segmentación y detección de bordes para la obtención de puntos de interés de la cara, ya que permiten un procesado en tiempo real y un tiempo de implementación acorde con el desarrollo de las prácticas, además de conseguir buenos resultados de clasificación de las expresiones tras la obtención de características de interés.

En primer lugar se optó por implementar el método propuesto en el Proyecto Fin de Carrera de R. N. Rojas Bello, "Identificación de características relevantes para reconocimiento de emociones en el rostro" [1], para poder tener un punto de partida. Este método se basa en la segmentación de las diferentes regiones de la cara para obtener información de inclinación de las cejas o apertura de la boca y usarlo en la clasificación de las emociones. Este algoritmo es muy limitado y proporciona resultados mejorables, por lo que se desarrolla un nuevo método, explicado a continuación, para solucionar los problemas y aumentar el número de expresiones detectadas.

#### 3.1. Implementación de método propuesto en Rojas\_08

El algoritmo propuesto en [1] consiste en la detección de sorpresa, ira o tristeza. En él se localiza la cara y los ojos con ayuda del algoritmo *Viola & Jones* implementado en la *Computer Vision Toolbox* de Matlab, pero con una modificación, se segmenta la boca y los ojos seleccionando un umbral de forma automática en función de las características estadísticas del histograma de la imagen. Posteriormente, se realizan operaciones morfológicas para extraer la forma de las regiones boca y cejas y se estudia la inclinación de las cejas y la apertura de la boca para distinguir si la expresión es sorpresa, ira o tristeza.

En la Tabla 1 se muestran los resultados obtenidos tras la implementación del sistema descrito en [1] y la realización de pruebas sobre la base de datos *Jaffe* [28]. Se obtienen porcentajes muy bajos, ya que en ningún caso superan el 60%. Además, el sistema propuesto solo detecta 3 expresiones, frente a las 6 o 7 del resto de los sistemas existentes. Por todo esto, el método de procesado se cambia de manera significativa para poder mejorar la detección y clasificación de las expresiones.

**Tabla 1.** Resultados obtenidos del método implementado en [1]

Expresión analizada	Tristeza	Sorpresa	Ira
Porcentaje de acierto	53,3%	36,6%	56,6%

### 3.2 Desarrollo de un nuevo método de obtención de puntos de interés

La obtención de información del tamaño o inclinación de las regiones de la cara no permite detectar muchas expresiones, por lo tanto en el nuevo método desarrollado se propone la extracción de puntos de interés de la cara para poder ampliar el rango de expresiones detectadas y la veracidad de éstas. En dicho método se localizan 12 puntos de interés y se crea una matriz de características del sujeto analizado, para posteriormente compararla con las matrices representativas correspondientes a 6 expresiones (*happy*, *surprise*, *sadness*, *disgust*, *anger* y *fear*, que son las más significativas) y proceder a la de la expresión con un algoritmo que se explica en profundidad en el apartado 3.3. Se muestra en la Figura 1 el diagrama de bloques de la solución implementada, que se explica a continuación.

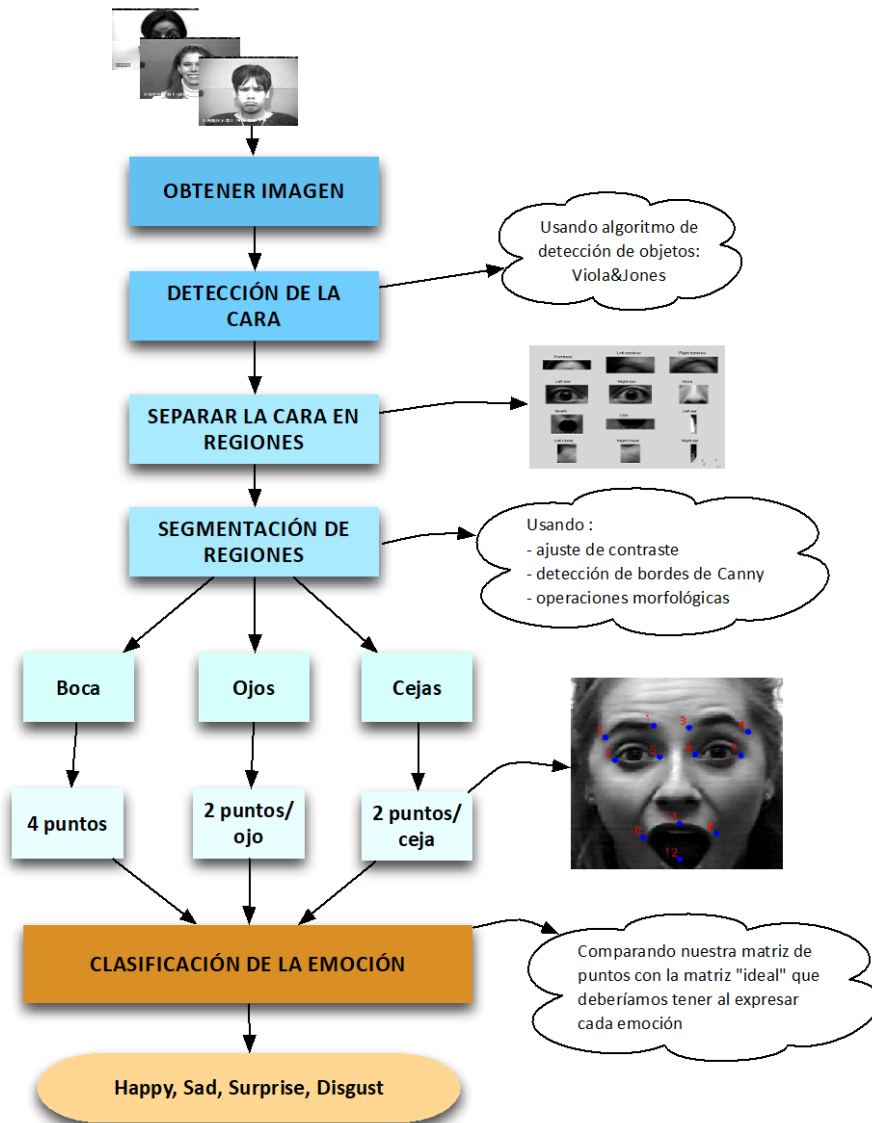


Figura 1. Diagrama de bloques de la solución implementada

Cualquier sistema de reconocimiento de expresiones debe implementar un reconocimiento facial. En primer lugar, tras la obtención de la imagen, se localiza la cara dentro de ella. Para ello se utiliza una función incluida en *Computer Vision Toolbox* de Matlab que implementa el algoritmo de *Viola & Jones*, y permite la localización de las caras en una imagen. Tras detectar la cara, se recorta y normaliza con un tamaño de  $N \times N$  píxeles para independizar el programa del tamaño de las caras. En este trabajo se escogió un tamaño de  $300 \times 300$  píxeles por ser el que, manteniendo un tiempo razonable de procesamiento de cada imagen, proporcionaba una buena

precisión en la detección de los puntos. Tamaños mayores aumentaban el tiempo de procesado de sensiblemente y tamaños menores dificultaban la correcta detección de los puntos.

A continuación, se divide la cara en regiones de interés según las proporciones y regiones expuestas en [29], de las cuáles se han seleccionado cejas, ojos y boca. En este artículo se establecen unas posiciones  $x$  e  $y$  en función de la altura y anchura de la cara. Se han realizado determinados ajustes en las proporciones propuestas en [29], para obtener mejores resultados ya que en algunos casos no se detectaban las cejas o la boca correctamente. Por ello se ha disminuido el ancho de la región extraída 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 poder segmentarla en su totalidad cuando se encuentra abierta.

Como se puede observar en el diagrama de bloques, el siguiente paso es la segmentación de las regiones que han sido extraídas para obtener 12 puntos significativos, en vez de los 14 indicados en [23], ya que no se observa gran mejoría al añadir estos dos puntos (puntos intermedios de las cejas) y se considera que 12 puntos son suficientes para la clasificación de las expresiones faciales.

Para la extracción de cada uno de ellos, se probaron varios métodos de segmentación con el objetivo de encontrar el que proporcionase mejores resultados y mayor precisión en la detección de los puntos. Finalmente el implementado ha sido el aumento del contraste, el uso de detección de bordes de *Canny* y las operaciones morfológicas de dilatación y erosión para eliminar ruido y objetos innecesarios en la imagen, mejorando la delimitación de las regiones analizadas. Para el caso de extracción de los puntos significativos de las cejas, se soluciona el problema de que debido a la cercanía entre las cejas y el pelo, se detectasen ambos como una sola región. Los métodos descartados son: utilización de máscaras elípticas, segmentación por agrupamiento de píxeles, segmentación por texturas y segmentación mediante el rango local de una imagen. Los extremos de la boca (4), cejas (2/ceja) y ojos (2/ojo), son identificados con la función *regionprops()* de Matlab. Para los puntos superiores e inferiores de la boca se calcula el punto intermedio entre ambos extremos de la boca en el eje  $x$  ya calculados y se recorre el eje  $y$  hasta encontrar una variación en dos píxeles consecutivos (ya que la imagen resultante de *Canny* es binaria).

### 3.3 Desarrollo de un método de clasificación de emociones

Tras obtener los 12 puntos de interés es necesario crear un algoritmo que permita la clasificación de las emociones con la información obtenida de estos puntos. Se muestra en la Figura 2 el diagrama de bloques de clasificación de la emoción implementado.

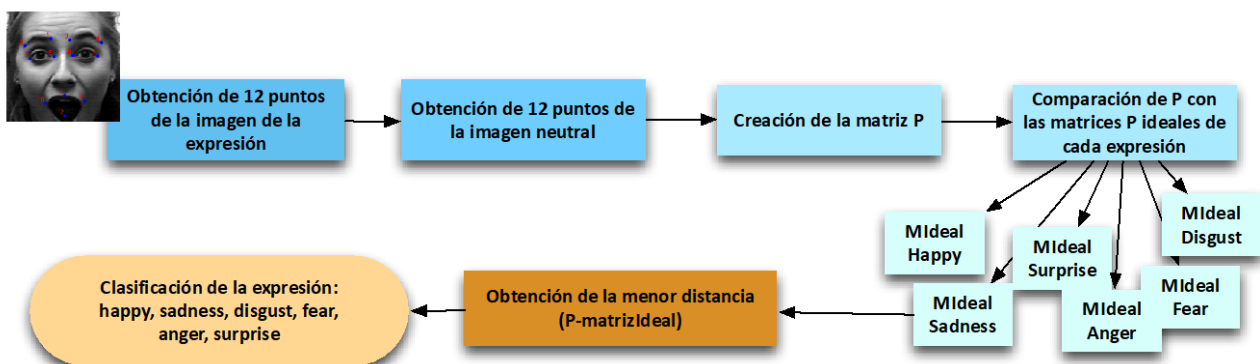


Figura 2. Diagrama de bloques de la clasificación de las expresiones

En primer lugar, se busca la imagen neutral correspondiente al sujeto analizado. Se obtienen los 12 puntos tanto de la imagen neutral como de la imagen a analizar. En segundo lugar se crea una matriz que da información sobre las posiciones de los puntos y sus movimientos respecto de la posición neutral cuando se

expresa la emoción. En concreto se trata de restar la posición de los 12 puntos detectados en la imagen que se está analizando, con la posición de esos mismos 12 puntos en la imagen correspondiente a la expresión neutral de ese mismo sujeto. Esta matriz (a la que se ha denominado “*P*” en el programa) tiene un tamaño de 12 x 4, cuyo formato se muestra en (1). Contiene en las dos primeras columnas la dirección y cantidad del desplazamiento, horizontal y vertical ( $\Delta h$  y  $\Delta v$ ), de cada punto respecto a ese mismo punto en la imagen de la expresión neutral. En las dos siguientes contiene la posición *x* e *y* de cada uno de los 12 puntos de la imagen analizada.

$$(1) \quad P = \begin{bmatrix} \Delta h_1 & \Delta v_1 & x_1 & y_1 \\ \Delta h_2 & \Delta v_2 & x_2 & y_2 \\ \dots & \dots & \dots & \dots \\ \Delta h_{12} & \Delta v_{12} & x_{12} & y_{12} \end{bmatrix} \quad (1)$$

Tras crear esta matriz se procede al entrenamiento del sistema para poder clasificar las diferentes emociones.

El diseño de un sistema de clasificación, en general, suele comprender dos fases. En primer lugar, se utiliza un conjunto de datos (en este caso, imágenes) para proporcionar la información necesaria al sistema de clasificación. Éste conjunto de datos se denomina comúnmente “conjunto de entrenamiento”. A continuación se proporcionan nuevos datos al sistema para que los clasifique y poder así evaluar su efectividad.

En nuestro sistema la fase de entrenamiento consiste en crear una “**matriz representativa**” de cada expresión, por lo que se tendrán en total 6 matrices representativas. La forma de crear cada una de ellas es, suponiendo una emoción *X*:

1. Obtener las matrices *P* de todas las imágenes de la emoción *X* que hay en una determinada base de datos de entrenamiento.
2. Hacer la media de todas las matrices *P* obtenidas de esa emoción, para hallar la matriz representativa.

Con esto se consigue tener una matriz representativa de cada expresión que debería parecerse lo más posible a las matrices *P* de las imágenes analizadas.

Para clasificar una emoción tras obtener su matriz *P* se resta a cada una de las matrices representativas de cada expresión y se obtiene la media de todos los elementos de la matriz diferencia. Aquella expresión cuya diferencia con la matriz representativa sea menor, es decir, cuya media sea menor, será la expresión detectada: *happy, surprise, disgust, sadness, anger o fear*.

### 3.4 Creación de una interfaz gráfica

Para garantizar la usabilidad de cualquier programa a menudo es necesario proporcionar una interfaz gráfica que permita al usuario acceder correctamente a sus funciones y visualizar los resultados. En este caso era necesario disponer de una forma cómoda de abrir las imágenes a procesar y representar gráficamente la posición de los puntos en la cara.

Se utiliza la herramienta GUIDE de Matlab para crear una ventana con 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 una forma visual y unificada.

También se implementan los controles necesarios para poder realizar las pruebas de porcentajes de aciertos y fallos, analizando de forma automática todas las imágenes de una base de datos, especificando la expresión que se quiere evaluar.

Se incorpora también una función de ayuda destinada a explicar el funcionamiento del programa a nuevos usuarios, y permitir en el futuro expandir sus funcionalidades de forma sencilla.

La interfaz proporciona también un acceso a la captura de imagen mediante cámara web, para poder realizar pruebas con otras imágenes diferentes de las de la base de datos.

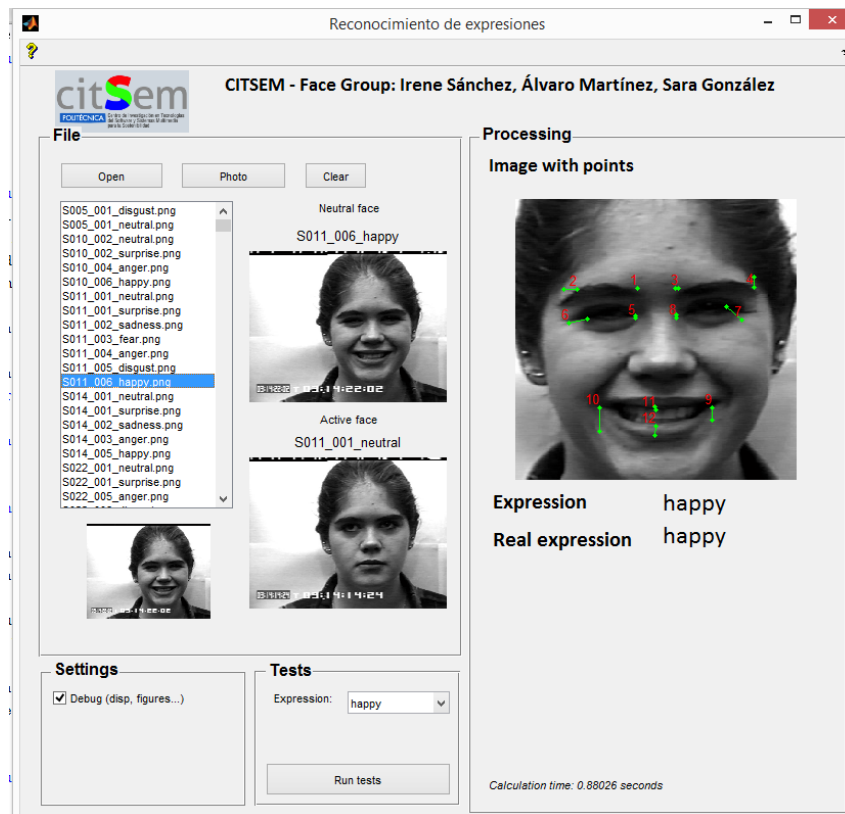


Figura 3. Diseño de la interfaz gráfica

## 4. Resultados

Para poder comprobar el funcionamiento del sistema se realizan pruebas en las bases de datos *CohnKanade+* [30], *Jaffe* [31] y en imágenes capturadas mediante la cámara del ordenador. Estas pruebas se realizan de forma automática utilizando la interfaz gráfica, que permite seleccionar la expresión que se quiere analizar y calcula y muestra los porcentajes de acierto o fallo.

Se crean matrices representativas diferentes para cada una de las dos bases de datos y se prueba sobre las mismas imágenes a partir de las cuáles se crearon dichas matrices. Por otro lado, se usan las matrices representativas creadas a partir de la base de datos *CohnKanade+* para realizar pruebas con imágenes capturadas por la webcam. No se usa la base de datos *Jaffe* para el análisis de las imágenes obtenidas con la cámara del ordenador, debido a que esta base de datos es limitada en cuanto al reducido número de sujetos y principalmente porque se trata de sujetos asiáticos, y las imágenes capturadas por la webcam pertenecen a sujetos europeos y tanto los rasgos faciales como la forma de expresar emociones varían significativamente.

Las expresiones que detecta el algoritmo son las 6 más analizadas en la mayoría de los trabajos existentes: *happy*, *disgust*, *anger*, *surprise*, *sadness*, y *fear*.



#### 4.1 Realización de pruebas con la base de datos *Cohn Kanade +* con 6 expresiones

Se realizan pruebas con diferentes tamaños de cara para comprobar cuál es el tamaño óptimo de normalización respecto a la relación calidad/tiempo de procesado. Los resultados obtenidos se muestran en la Tabla 2 y la Tabla 3. Estas tablas son unas tablas de confusión, es decir, al analizar una expresión muestran el porcentaje de acierto que se obtiene de cada una de las 6 expresiones analizadas, para, en caso de error, saber con qué expresión o expresiones se confunde y poder tomar medidas correctoras. Así se muestran en las filas de la primera columna cada una de las expresiones analizadas y en las siguientes columnas la expresión que se ha clasificado al analizar la emoción de una fila determinada.

**Tabla 2.** Pruebas en *CohnKanade+* con 6 expresiones y tamaño de la cara de 200x200

Expresión detectada → Expresión analizada ↓	Happy	Disgust	Anger	Surprise	Sadness	Fear
Happy	86,6	4,5	5,9	1,5	0,0	1,5
Disgust	8,7	60,0	17,6	7,0	0,0	7,0
Anger	6,8	4,5	68,2	2,3	6,8	11,4
Surprise	1,2	1,2	0,0	93,8	3,7	0,0
Sadness	0,0	3,7	18,5	14,8	30,0	33,3
Fear	16,7	4,2	29,2	4,17	4,2	41,7

**Tabla 3.** Pruebas en *CohnKanade+* con 6 expresiones y tamaño de la cara de 300x300

Expresión detectada → Expresión analizada ↓	Happy	Disgust	Anger	Surprise	Sadness	Fear
Happy	85,1	4,4	7,4	0,0	0,0	2,9
Disgust	10,5	75,4	5,2	5,2	1,8	1,8
Anger	20,5	11,3	47,7	9,0	9,0	2,2
Surprise	0,0	2,5	1,2	91,3	3,7	1,2
Sadness	0,0	11,1	7,4	11,1	40,7	29,6
Fear	4,2	8,3	16,7	0,0	20,8	50,0

Las pruebas se realizan con un tamaño de cara de 200 x 200 (tabla 2) y con 300 x 300 (tabla 3). Se observa que en el tamaño de 200 x 200 se tienen problemas significativos con la detección de *sadness* y *fear* ya que se obtienen porcentajes inferiores al 42% en ambos casos, pues se reduce el tamaño medio de la cara original en la base de datos. Sin embargo, al aumentar el tamaño de la cara a 300 x 300, que es el tamaño común correspondiente a la cara para las imágenes de esta base de datos, estos porcentajes tienen una mejora de un 10%. Esto es debido a que a menor resolución las distancias entre los puntos son menos precisas y tienen menos importancia en la media final que se hace al comparar la imagen analizada con cada una de las matrices representativas.

Por otro lado, si se aumenta el tamaño de 200 x 200 a 300 x 300, el tiempo de procesado aumenta en unos 40 ms. Lo mismo ocurre si se continúa aumentando la resolución, por tanto, de ahora en adelante se continúan las pruebas con una resolución de 300 x 300.

Los mejores resultados en la clasificación de las expresiones se obtienen con *happy*, *disgust* y *surprise*. Esto es debido a que estas expresiones son muy diferentes entre sí y no se suelen confundir con otras, al contrario que *sadness*, *fear* y *anger*.

## 4.2 Realización de pruebas con la base de datos *Cohn Kannade +* con 5 expresiones

A continuación, se descarta una de las expresiones que tiene menor porcentaje de acierto y se muestran los resultados en las siguientes tablas, para ver los porcentajes de acierto con 5 expresiones. Los peores porcentajes al analizar las 6 expresiones son en *anger*, *sadness* y *fear*, ya que se confunden entre sí. Por ello, se analizan 5 expresiones, eliminando en la Tabla 4 la expresión *sadness*, en la Tabla 5 la expresión *anger* y en la Tabla 6 la expresión *fear*.

Se muestra a continuación una de las razones por las cuales el algoritmo diseñado no distingue bien entre *anger*, *sadness* y *fear*. Hay muchas imágenes en la base de datos en las que visualmente es difícil diferenciar si se trata de la expresión *fear*, *sadness* o *anger*. Esto se muestra en la **Figura 4**, con unas imágenes que la base de datos las clasifica como *anger* y visualmente podrían ser *sadness* o *disgust*. Esto afecta a la determinación de la emoción por el algoritmo implementado, y a la reducción de los porcentajes de acierto.



**Figura 4.** Imágenes de la base de datos *CohnKanade+* clasificadas como *anger*, que visualmente pueden ser identificadas como *sadness*

Se puede observar en la Tabla 4, comparándola con la Tabla 3, que los resultados obtenidos en *happy*, *surprise* y *disgust* prácticamente no varían. Sin embargo, se aprecian mejoras significativas en *anger* (de 47 a 50%) y *fear* (de 50 a 62%).

**Tabla 4.** Pruebas en *CohnKanade+* con todas las expresiones menos *sadness*

Expresión detectada → Expresión analizada ↓	Happy	Surprise	Disgust	Anger	Fear
Happy	85,0	0,0	4,5	7,5	2,9
Surprise	0,0	92,5	2,5	2,5	2,5
Disgust	10,5	5,2	75,5	5,2	3,5
Anger	30,5	9,09	11,3	50,0	9,1
Fear	4,1	0,0	8,3	25,0	62,5

Comparando igualmente los resultados obtenidos en la Tabla 5 con los de la Tabla 3, los porcentajes en *happy*, *surprise* y *disgust* prácticamente no varían, y se aprecian mejoras también significativas en *sadness* (de 40,7 a 44,4%) y *fear* (de 50 a 66,6%). Uno de los problemas es que aun mejorando los porcentajes no se consigue que la expresión *sadness* pase del 50% de aciertos.

**Tabla 5.** Pruebas en *CohnKanade+* con todas las expresiones menos *anger*

Expresión detectada → Expresión analizada ↓	Happy	Surprise	Disgust	Sadness	Fear
Happy	85,1	0,0	7,5	0,0	7,4
Surprise	0	92,5	2,4	3,7	1,2
Disgust	10,5	7,0	75,4	3,5	3,5
Sadness	0,0	14,8	11,1	44,4	29,6
Fear	4,1	0,0	8,3	20,8	66,6

Con 6 expresiones el tercer peor porcentaje se detectaba en *fear* por lo que se realizan también pruebas eliminando esta expresión y se muestran los resultados en la Tabla 6. En este caso *happy*, *surprise* y *disgust* también permanecen casi iguales, así como *anger*. Por otro lado, *sadness* aumenta los porcentajes de 40,7 a 55,5.

**Tabla 6.** Pruebas en *CohnKanade+* con todas las expresiones menos *Fear*

Expresión detectada → Expresión analizada ↓	Happy	Surprise	Disgust	Anger	Sadness
Happy	85,1	0,0	4,5	8,9	1,4
Surprise	0,0	92,6	2,4	1,2	3,7
Disgust	12,2	5,2	75,5	5,2	1,7
Anger	20,4	9,1	11,3	47,7	11,4
Sadness	3,7	14,8	11,1	14,8	55,5

Se puede concluir que los mejores resultados obtenidos del análisis de 5 expresiones se observan al elegir las expresiones: *happy*, *disgust*, *anger*, *surprise* y *fear*, ya que todos los porcentajes de acierto en este caso se encuentran por encima del 50%.

### 4.3 Realización de pruebas con la base de datos *Jaffe*

A continuación se realizan pruebas en una base de datos distinta. Se trata de una base de datos con reducido número de sujetos de origen asiático cuya principal diferencia con la base de datos *CohnKanade+* es que en este caso los sujetos tienen una forma distinta de expresar emociones, por ejemplo para expresar sorpresa no suelen abrir la boca. Además, contiene únicamente imágenes de mujeres, al contrario que la *CohnKanade+*.

En la Tabla 7 se aprecia que los resultados para esta emoción (sorpresa) son peores que cuando se analizó en la base de datos *CohnKanade+* (con un 91,3 %). Para el resto de emociones los resultados son similares y los porcentajes más bajos de acierto se obtienen en las expresiones de *sadness* y *fear*.

**Tabla 7.** Pruebas en *Jaffe* con 6 expresiones y tamaño de la cara de 300x300

Expresión detectada → Expresión analizada ↓	Happy	Disgust	Anger	Surprise	Sadness	Fear
Happy	80,6	6,4	3,2	0,0	3,2	6,4
Disgust	0,0	82,1	3,5	0,0	7,1	7,1
Anger	3,3	20,0	70,0	0,0	6,6	0,0
Surprise	10,0	6,6	0,0	53,3	3,3	26,6
Sadness	6,4	12,9	6,4	0,0	48,3	25,8
Fear	15,6	28,1	3,1	9,3	9,3	34,3

### 4.4 Realización de pruebas con imágenes obtenidas de la cámara del ordenador

Se han hecho también pruebas sobre una base de datos creada a partir de imágenes tomadas con la cámara del ordenador para comprobar su funcionamiento en el mundo real. Actualmente es una base de datos poco relevante ya que cuenta solo con imágenes de 8 sujetos y se pretende ampliar el número de sujetos de esta base de datos en un futuro. Los resultados obtenidos se muestran en la **Tabla 8**. Las matrices representativas utilizadas para la comparación han sido las obtenidas con la base de datos *CohnKanade+*.

Se puede observar que en la expresión *happy* se obtiene un porcentaje de acierto muy alto, superior incluso al obtenido en las pruebas realizadas en la base de datos *CohnKanade+*. Sin embargo en el resto de expresiones se obtienen porcentajes inferiores, ya que los sujetos analizados de la cámara del ordenador expresan estas emociones de forma diferente a los sujetos de *CohnKanade+*.

**Tabla 8.** Pruebas en imágenes de la cámara del ordenador con 6 expresiones y tamaño de la cara de 300x300

Expresión detectada → Expresión analizada ↓	Happy	Disgust	Anger	Surprise	Sadness	Fear
Happy	85,7	0,0	0,0	14,3	0,0	0,0
Disgust	42,8	42,8	0,0	0,0	0,0	14,3
Anger	0,0	33,3	16,7	16,7	33,3	0,0
Surprise	0,0	14,3	14,3	42,8	0,0	28,6
Sadness	0,0	14,3	42,8	14,3	28,6	0,0
Fear	25,0	25,0	0,0	0,0	25,0	25,0

#### 4.5 Comparación de resultados con estudios similares

Como se ha indicado en el estado del arte hay una gran cantidad de métodos diferentes para la detección de emociones. El problema es la dificultad de comparar los resultados obtenidos de cada uno de estos métodos ya que cada uno usa una base de datos diferente o solo una parte de una base de datos, por lo que las comparaciones no serían muy precisas, ya que varían significativamente los resultados en función de la base de datos utilizada.

Sin embargo, se puede concluir que con los *papers* analizados, los mejores resultados y los más utilizados por los grupos de investigación se obtienen en los métodos de LBP (*local binary pattern*), SRC (*sparse representation classifier*), *Gabor Wavelets* y el uso de redes neuronales y SVM para la clasificación de la expresión.

Se han comparado los resultados obtenidos con los de los *papers* encontrados que se basan en la detección de puntos de la cara para la clasificación de la emoción con el uso de la base de datos *CohnKanade+*. Se ha comparado con el método desarrollado en [5] que detalla un método de reconocimiento de emociones extrayendo 8 o 68 puntos de la cara y posteriormente clasificando la emoción en función de las distancias entre éstos, utilizando *CohnKanade+*. Posteriormente también se ha comparado con [6] que utiliza la segmentación y la detección de bordes de *Canny*, como el método implementado durante el desarrollo de estas prácticas, para la extracción de características, pero sólo detecta 4 expresiones (el método implementado en estas prácticas detecta 6 expresiones). El último de los métodos comparados ha sido el *paper* que explica la creación de la base de datos *CohnKanade+* ya que presenta porcentajes de aciertos del análisis entero de ésta con el método *FACES* [30].

Se muestran los porcentajes de acierto de cada método en la **Tabla 9** para poder realizar la comparación entre éstos. Se puede observar que en todos los métodos los mejores porcentajes de acierto, se obtienen en *happy* y *surprise*. Nuestra implementación mejora significativamente los resultados de [6], ya que se obtienen porcentajes de acierto similares o superiores y detecta mayor número de expresiones que [6]. Se destaca que los métodos desarrollados en [5] y [30] tienen porcentajes de acierto superiores a los nuestros pero los peores porcentajes de acierto los obtienen en las mismas expresiones con las que nosotros obtenemos los menores porcentajes: *sadness*, *fear* y *anger*.

Con todo esto se puede concluir que nuestro método tiene todavía mucho margen de mejora, pero que se ha conseguido realizar un método bastante fiable que sea un buen punto de partida para futuras investigaciones en este ámbito en nuestro laboratorio.

**Tabla 9.** Comparación de resultados

Expresión analizada	Método desarrollado	Resultados de [5]	Resultados de [6]	Resultados De [30]
Happy	85,1	97,1	95,2	100,0
Disgust	75,4	88,2	-	94,7
Anger	47,7	75,6	50,0	75,0
Surprise	91,3	98,7	95	96,0
Sadness	40,7	89,3	33,3	68,0
Fear	50	76,0	-	65,2

## 5. Conclusiones

Gracias al estudio realizado del estado del arte se ha conseguido el primer objetivo propuesto: profundizar y entender la mayoría de las herramientas que existen en la actualidad del análisis de las expresiones faciales.

Otro de los objetivos iniciales era analizar y optimizar el método propuesto en el Proyecto Fin de Carrera [1] debido a que no proporciona resultados óptimos y detecta un rango limitado de expresiones. Para ello en el nuevo algoritmo desarrollado se procede a la localización de 12 puntos de interés en la cara del sujeto a analizar mediante la separación de la cara en regiones, seguido de un pre-procesado y una segmentación más precisa. Después se compara la distribución de estos puntos con una serie de matrices representativas correspondientes a 6 expresiones (*happy*, *disgust*, *anger*, *surprise*, *sadness*, y *fear*) ya que era la forma más eficiente en cuanto al tiempo disponible y proporcionaba resultados satisfactorios: se pasa de tan solo detectar 3 expresiones (*surprise*, *anger* y *sadness*) a las 6 mencionadas anteriormente. A su vez, los porcentajes de acierto obtenidos en la implementación de [1] han sido mejorados notablemente, como se observa en la Tabla 8.

**Tabla 10.** Comparación de resultados entre el algoritmo de partida y el algoritmo final

PORCENTAJES DE ACIERTO	Happy	Disgust	Anger	Surprise	Sadness	Fear
Algoritmo inicial con base de datos <i>Jaffe</i>	53,3	36,6	-	-	-	56,6
Algoritmo final con base de datos <i>Jaffe</i>	80,6	82,1	70,0	53,3	48,3	34,3
Algoritmo final sobre con de datos <i>CohnKanade+</i>	85,1	75,4	47,4	91,3	40,7	50

Se puede observar que los peores porcentajes obtenidos tras la realización de pruebas sobre CK+ son en las expresiones *sadness*, *anger* y *fear*, ya que son las más parecidas de esta base de datos. Sin embargo en la base de datos *Jaffe* los peores porcentajes se obtienen además de en *sadness* y *fear*, por las mismas razones que en la anterior base de datos, en *surprise* también, debido a que en esta base de datos los sujetos expresan sorpresa de una forma diferente, ya que no abren la boca y muchas expresiones que etiquetadas como *surprise* son clasificadas como *fear*.

## 6. Trabajo futuro

Aunque los resultados obtenidos para las expresiones analizadas son satisfactorios, el sistema aún tiene margen de mejora. Es razonable pensar que perfeccionando la detección de los puntos de la cara se obtendrán resultados más sólidos en la clasificación de las expresiones, ya que en determinados sujetos con arrugas, sombras o inclinaciones de la cara, los puntos se detectan de manera errónea. Una forma de hacerlo puede ser mejorar el sistema de pre-procesado de la imagen para que sea capaz de corregir la orientación de la cara; así como condiciones de iluminación variables que puedan empeorar la segmentación. Otra mejora puede ser el aumentar el número de puntos extraídos de la cara y realizar pruebas para comprobar la veracidad de la relevancia del aumento de puntos. Además se podrían añadir a la matriz de representativa de clasificación de la expresión más columnas. Así en vez de las cuatro descritas, se podría añadir por ejemplo una con información de las distancias entre puntos de la cara para proporcionar más información que fuese de utilidad para la clasificación de la emoción.

Además se podría estudiar la posibilidad de aplicación de este método a la detección de micro-expresiones y expresiones espontáneas, ya que es en este ámbito donde se centrará la investigación en un futuro.

Por otro lado la implementación de otro método para clasificar las emociones, como SVM, daría lugar a una comparación más exhaustiva y serviría para mejorar resultados y determinar el método óptimo.

Por último se destaca que aunque el método implementado trabaja muy rápido, se podría intentar mejorar la eficacia del método para conseguir que trabajase en tiempo real y poder ser utilizado con vídeo.

Cabe señalar que aunque este estudio se centra en la clasificación de la expresión facial, ésta no es la única fuente de información a la hora de interpretar las emociones. La combinación de estos métodos con información contextual o proveniente de otras formas de expresión, como la voz, permitiría discriminar de forma más efectiva estas emociones.

## 7. Referencias

- [1] R. N. Rojas Bello, "Identificación de características relevantes para reconocimiento de emociones en el rostro," 2009.
- [2] 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.
- [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] J. Ahlberg, "CANDIDE-3 - An Updated Parameterised Face," vol. 1, pp. 1–16, 2001.
- [8] M. Lyons and S. Akamatsu, "Coding Facial Expressions with Gabor Wavelets," pp. 200–205, 1998.
- [9] S. W. Chew, P. Lucey, S. Lucey, J. Saragih, J. F. Cohn, I. Matthews, and S. Sridharan, "In the Pursuit of Effective Affective Computing: The Relationship Between Features and Registration.," *IEEE Trans. Syst. Man. Cybern. B. Cybern.*, vol. 42, no. 4, pp. 1006–1016, May 2012.
- [10] S. Zhang, X. Zhao, and B. Lei, "Robust facial expression recognition via compressive sensing.," *Sensors (Basel)*, vol. 12, no. 3, pp. 3747–61, Jan. 2012.
- [11] M. A. R., "Facial Expression Recognition Using SVM Classification in Perceptual Color Space," vol. 2, no. June, pp. 363–368, 2013.
- [12] S. Wan and J. K. Aggarwal, "Spontaneous facial expression recognition: A robust metric learning approach," *Pattern Recognit.*, vol. 47, no. 5, pp. 1859–1868, May 2014.
- [13] S. Lucey, A. B. Ashraf, and J. F. Cohn, "Recognition through AAM Representations of the Face," 2005.
- [14] P. Wang, F. Barrett, E. Martin, M. Milonova, R. E. Gur, R. C. Gur, C. Kohler, and R. Verma, "Automated video-based facial expression analysis of neuropsychiatric disorders.," *J. Neurosci. Methods*, vol. 168, no. 1, pp. 224–38, Mar. 2008.
- [15] 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.
- [16] 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.

- [17] H. Meng, B. Romera-Paredes, and N. Bianchi-Berthouze, "Emotion recognition by two view SVM\_2K classifier on dynamic facial expression features," *Face Gesture 2011*, pp. 854–859, Mar. 2011.
- [18] R. Mehta, J. Yuan, and K. Egiazarian, "Face recognition using scale-adaptive directional and textural features," *Pattern Recognit.*, vol. 47, no. 5, pp. 1846–1858, May 2014.
- [19] D. Chakrabarti and D. Dutta, "Facial Expression Recognition Using Eigenspaces," *Procedia Technol.*, vol. 10, pp. 755–761, Jan. 2013.
- [20] B. Seddik, H. Maâmatou, S. Gazzah, T. Chateau, N. Essoukri, and B. Amara, "Unsupervised Facial Expressions Recognition and Avatar Reconstruction from Kinect," pp. 1–6, 2013.
- [21] M. Taner Eskil and K. S. Benli, "Facial expression recognition based on anatomy," *Comput. Vis. Image Underst.*, vol. 119, pp. 1–14, Feb. 2014.
- [22] K. Yurtkan and H. Demirel, "Feature selection for improved 3D facial expression recognition," *Pattern Recognit. Lett.*, vol. 38, pp. 26–33, Mar. 2014.
- [23] C. Tanchotsrinon, S. Phimoltares, S. Maneeroj, and A. Virtual, "FACIAL EXPRESSION RECOGNITION USING GRAPH-BASED FEATURES," 2011.
- [24] R. R. Londhe and V. P. Pawar, "Analysis of Facial Expression and Recognition Based On Statistical Approach," no. 2, pp. 391–394, 2012.
- [25] V. Bettadapura, "Face Expression Recognition and Analysis : The State of the Art," pp. 1–27, 2012.
- [26] S. Consultant and I. F. E. Databases, "A SURVEY ON FACIAL EXPRESSION DATABASES," vol. 2, no. 10, pp. 5158–5174, 2010.
- [27] N. U. Khan, "A Comparative Analysis of Facial Expression Recognition Techniques," pp. 1262–1268, 2012.
- [28] P. Viola, O. M. Way, and M. J. Jones, "Robust Real-Time Face Detection," vol. 57, no. 2, pp. 137–154, 2004.
- [29] B. Seddik, H. Maâmatou, S. Gazzah, T. Chateau, N. Essoukri, and B. Amara, "Unsupervised Facial Expressions Recognition and Avatar Reconstruction from Kinect," in *10th International Multi-Conference on Systems, Signals & Devices (SSD)*, 2013, pp. 1–6.
- [30] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, I. Matthews, and F. Ave, "The Extended Cohn-Kanade Dataset ( CK + ): A complete dataset for action unit and emotion-specified expression," no. July, 2010.
- [31] J. G. Michael J. Lyons, Miyuki Kamachi, "Japanese Female Facial Expressions (JAFFE), Database of digital images," 1997.