



Universidad
Zaragoza

Proyecto Fin de Carrera

Desarrollo de un sistema de reconocimiento de emociones faciales en secuencias de vídeo

Autor: Antonio Loscertales Albiol
Directoras: Dra. Sandra Baldassarri
Dra. Eva Cerezo

Ingeniería Informática
Septiembre 2012



Escuela de
Ingeniería y Arquitectura
Universidad Zaragoza



Departamento de
Informática e Ingeniería
de Sistemas
Universidad Zaragoza



Sony: *¿Qué significa esta acción? Cuando estaba en el cuarto, cuando miró al otro humano cerrando un ojo. ¿Qué significa?*

Spooner: *Es un signo de confianza. Es un gesto humano. No lo entenderías.*

Yo, robot.

Agradecimientos

En primer lugar quiero agradecer a mis dos directoras: Eva y Sandra. Gracias por brindarme la oportunidad de trabajar con vosotras y de guiarme en el apasionante mundo de la computación afectiva. Gracias por vuestra ayuda incondicional, vuestra paciencia infinita y vuestro apoyo a lo largo de todo el proyecto. He trabajado realmente a gusto con vosotras.

También quiero agradecer a Isabelle por toda la ayuda ofrecida. Gracias por tus ánimos, tus ideas y tus respuestas siempre cargadas de entusiasmo y motivación. Y gracias también por arrojar optimismo cuando el reconocedor facial, los sujetos de evaluación y los clasificadores conspiraban juntos contra mí.

Gracias también a Javier Marco y a Sergio Ballano, por ayudarme a arrancar en el proyecto.

A mis compañeros de laboratorio, que me han acompañado en el camino. Gracias por compartir todas esas horas conmigo, por estar ahí (en las emociones buenas y en las malas) y gracias por acceder a ser mis conejillos de indias cuando el sistema y yo nos cansábamos de analizarme la cara.

A mis compañeros de carrera. Han sido muchas, muchas horas juntos. Muchas clases, muchas prácticas, muchos laboratorios, exámenes y horas de estudio. Pero también muchos cafés, descansos, excursiones, fiestas, viajes, *Interrailés* y *Erasmus*. Gracias por todos aquellos momentos. Y gracias por todos los que vendrán.

A mis amigos de Valderrobres que, pese a no saber muy bien lo que hago y creer que enseño a amar a las máquinas, siempre han estado ahí cuando los he necesitado.

Finalmente, mi más profunda gratitud a mi familia, que siempre me ha ayudado y me ha dado ánimos para sacar adelante mis metas, apoyándome en todas las decisiones. Aunque ahora me vaya lejos, diez mil kilómetros no son suficientes para separarnos. Muchas gracias por todo.

Desarrollo de un sistema de reconocimiento de emociones faciales en secuencias de vídeo

RESUMEN

En este Proyecto Fin de Carrera se ha desarrollado un sistema de reconocimiento automático de emociones faciales en secuencias de vídeo. El presente trabajo se enmarca dentro del proyecto nacional de investigación “*REPLIKANTS: Hacia una nueva generación de agentes de tipo humano (TIN2011-24660)*”, y de una colaboración con el grupo GRIHO de la *Universitat de Lleida*, especializado en usabilidad y con la empresa *BitBrain Technologies*, líderes en interfaces cerebro-ordenador.

Este Proyecto Fin de Carrera ha tomado como punto de partida el trabajo anterior desarrollado por el grupo GIGA Affective Lab en la clasificación facial de emociones basado en un enfoque emocional discreto y ha realizado una migración completa al espacio emocional continuo, dando lugar a un nuevo sistema de reconocimiento facial de emociones. Con la migración al nivel de descripción continuo, se han conseguido superar las limitaciones del anterior sistema, permitiendo una descripción emocional más rica y una representación más precisa de la variación de las emociones a lo largo del tiempo.

Para conseguir los objetivos, se ha realizado un estado del arte sobre los trabajos más recientes de detección emocional. Se han analizado las alternativas posibles en cada una de las fases del proceso de detección facial de emociones, seleccionando las más adecuadas para el sistema a desarrollar.

La migración del trabajo previo a un enfoque emocional continuo ha hecho necesario llevar a cabo un estudio exhaustivo de las Bases de Datos Emocionales continuas disponibles. Tras realizar el estudio, se ha seleccionado la Base de Datos Emocional que se ajustaba a los requerimientos de un sistema con enfoque emocional continuo. Con la nueva base de datos, se ha implementado y entrenado un nuevo clasificador de emociones.

Una vez realizadas dichas tareas, se ha desarrollado el nuevo sistema de reconocimiento de emociones faciales de carácter continuo, complementado por una interfaz de visualización que permite tomar un vídeo en secuencia de imágenes, seleccionar *frames* y visualizar a qué región del espacio pertenece la emoción detectada por el sistema de clasificación.

Por último, mencionar que se va a hacer uso del sistema de reconocimiento de emociones desarrollado en el “Proyecto Neurogame: Neuro-evaluación multimodal de videojuegos”, en el que el autor del trabajo ha participado activamente tanto en el análisis facial de emociones como en la definición de la metodología de evaluación.

Contenidos

CAPÍTULO 1. Introducción	1
1.1 Motivación.....	1
1.2 Contexto del Proyecto	1
1.3 Objetivos.....	2
1.4 Organización de la memoria.....	2
CAPÍTULO 2. Planteamiento del problema: reconocimiento facial automático de emociones.....	3
2.1 Introducción	3
2.2 Selección del modelo facial	4
2.3 Selección del nivel de descripción. Enfoques emocionales.....	4
2.4 Selección del mecanismo de clasificación	6
2.5 Sistemas actuales de reconocimiento facial de emociones	7
2.6 Trabajo previo	8
CAPÍTULO 3. Trabajo realizado.....	11
3.1 Esquema del sistema de reconocimiento facial de emociones en secuencias de vídeo.	11
3.2 Extracción de datos	12
3.3 Conversión de datos.....	14
3.4 Selección de una Base de Datos Emocional	17
3.5 Entrenamiento del clasificador continuo	22
3.6 Clasificación	31
3.7 Visualización	32
3.8 Validación con vídeos	33
CAPÍTULO 4. Aplicación a la evaluación emocional de videojuegos: <i>Neurogame</i>	37
4.1 Evaluación emocional de videojuegos	37
4.2 Metodología propuesta.....	38
4.3 Primeras pruebas.....	39
CAPÍTULO 5. Conclusiones y trabajo futuro	41
5.1 Conclusiones.....	41
5.2 Líneas de trabajo futuro	42
5.3 Valoración personal.....	43

ANEXO A. Gestión del Proyecto	45
A.1 Gestión del tiempo	45
A.2 Esfuerzo invertido	46
ANEXO B. Tecnología y herramientas utilizadas	49
B.1 Reconocimiento facial: <i>FaceAPI</i>	49
B.2 Herramienta de clasificación: <i>WEKA</i>	50
B.3 Captura y Procesado de Vídeo: <i>VirtualDub</i>	52
B.4 IDEs Empleados.....	53
ANEXO C. Selección de sesiones	57
C.1 El problema de la selección de los datos	57
C.2 Datos disponibles.....	58
C.3 Criterios de selección.....	60
C.4 Datos seleccionados.....	62
C.5 Problemas encontrados	64
ANEXO D. Benchmarking del clasificador	69
D.1 Introducción	69
D.2 Selección de clasificadores para las pruebas.....	69
D.3 Selección de atributos	70
D.4 Resultados de Benchmarking	72
ANEXO E. Proceso de desarrollo del Software.....	75
E.1. Sistema de reconocimiento automático de emociones faciales desarrollado	75
E.2 Proceso de desarrollo de la fase 1: extracción de datos.....	76
E.3 Proceso de desarrollo de la fase 2: conversión de datos.....	76
E.4 Proceso de desarrollo de la fase 3: clasificación	76
E.5 Proceso de desarrollo de la fase 4: visualización	79
ANEXO F. Publicación Interacción 2012.....	83
BIBLIOGRAFÍA Y DOCUMENTACIÓN.....	93
URLs de Bases de Datos Emocionales	97

ÍNDICE DE FIGURAS

Fig. 1. Proceso de reconocimiento automático de emociones faciales	3
Fig. 2. Modelos faciales. a) modelo analítico, b) modelo holístico, c) modelo híbrido.....	4
Fig. 3. Figura que muestra las 6 emociones básicas definidas por Ekman: enfado, sorpresa, tristeza, aversión, miedo y alegría.....	5
Fig. 4. Espacio dimensional activación-evaluación de Whissell [Whi89] con las 6 emociones básicas (más la neutral) resaltadas en él.....	5
Fig. 5. Esquema general del trabajo previo desarrollado en reconocimiento emocional	8
Fig. 6. Cambios necesarios en el sistema anterior para una completa migración al enfoque continuo	9
Fig. 7. Esquema general del sistema de reconocimiento facial de emociones en secuencias de vídeo	11
Fig. 8. Esquema general de la fase 1: extracción de datos.....	12
Fig. 9. 38 puntos faciales proporcionados por la tecnología FaceAPI de SeeingMachines	14
Fig. 10. Esquema general de la fase 2: conversión de datos.....	15
Fig. 11. Los ángulos [ang 01-07] a la izquierda. Las distancias faciales (dst 08-19) a la derecha... <td>16</td>	16
Fig. 12. Estructura de un fichero ARFF	17
Fig. 13. Proceso de entrenamiento del clasificador continuo	23
Fig. 14. Proceso de extracción de información de la BDE	24
Fig. 15. Distancias faciales seleccionadas A la izquierda, todas las distancias (iniciales). A la derecha, las distancias seleccionadas (finales).	25
Fig. 16. Distribución de los sujetos etiquetados en el espacio activación-evaluación.....	26
Fig. 17. Modelo 9R dividido en 9 regiones	26
Fig. 18. Salidas finales del clasificador.....	27
Fig. 19. Matriz de Confusión del clasificador seleccionado: CSC-MLP	30
Fig. 20. Ejemplo de expresiones parecidas (alegría - sorpresa) [PR00].....	30
Fig. 21. Esquema general de la fase 3: clasificación	31
Fig. 22. Esquema general de la fase 4: visualización	32
Fig. 23. Interfaz del visualizador del reconocedor de emociones faciales	33
Fig. 24. Interfaz del visualizador mostrando el <i>appex</i> de un video de la <i>MAHNOB-HCI tagging database</i> etiquetado en un nivel máximo de activación y mínimo de evaluación	34
Fig. 25. Interfaz del visualizador mostrando el <i>appex</i> de un video de un sujeto solicitado con videos que provocan alegría.....	35
Fig. 26. Diferentes canales emocionales y cognitivos del proyecto <i>Neurogame</i>	38
Fig. 27. Esquema general de una sesión.....	39
Fig. 28. Escala SAM (<i>Self Assessment Manikin</i>) utilizada para la autoevaluación emocional	39
Fig. 29. Primeros resultados del analizador facial. Usuario contento durante la evaluación emocional	40

Fig. A. 1. Evolución temporal del proyecto	45
Fig. A. 2. Distribución del tiempo para cada una de las diferentes tareas del proyecto	46
Fig. B. 1. Interfaz de la demo de reconocimiento del software <i>faceapi</i>	50
Fig. B. 2. Interfaz de <i>weka explorer</i>	51
Fig. B. 3. Gráficas de correlación. En la parte superior la correlación entre las distancias 15 y 16 con el ángulo 7 (nariz). En la parte inferior los ángulos 5 y 6 con la distancia 17 (labios)	51
Fig. B. 4. Visualización de un árbol de decisión generado en una sesión de entrenamiento.....	51
Fig. B. 5. Interfaz gráfica de <i>VirtualDub</i>	53
Fig. B. 6. Interfaz Gráfica del IDE <i>Qt Creator</i>	54
Fig. B. 7. Interfaz del entorno de desarrollo <i>Eclipse</i>	55
Fig. C. 1. Distribución de las emociones categóricas de la BDE <i>mahnob-hci tagging database</i>	58
Fig. C. 2. Distribución de las 511 sesiones en función del etiquetado en activación-evaluación..	58
Fig. C. 3. Ocurrencias por regiones de las 511 sesiones	59
Fig. C. 4. Modelo de 8-regiones escogido como salida de clasificación.....	60
Fig. C. 5. Frecuencia de aparición de los sujetos en las 511 sesiones.....	61
Fig. C. 6. Mapas de calor de las 8 emociones discretas de la bde seleccionada. Distribución en el espacio 2d activación-evaluación	63
Fig. C. 7. Información sobre las 106 sesiones seleccionadas	64
Fig. C. 8. Primeros frames de dos sesiones diferentes. A la izquierda, sujeto nº3 comenzando el vídeo con una sonrisa. A la derecha, sujeto nº 20 comenzando el vídeo mordiéndose un labio .	65
Fig. C. 9. <i>Appex</i> de una sesión etiquetada en la región 9, con máxima activación y máxima evaluación.....	66
Fig. C. 10. Expresión de miedo en 6 <i>frames</i> correspondientes a 0.5 segundos de vídeo	67
Fig. D. 1. Esquema de selección de atributos	71
Fig. D. 2. Matriz de confusión del clasificador CSC-MLP entrenado con validación cruzada y el fichero de entrada CSC-1 con 12 atributos.....	74
Fig. E. 1. Sistema de reconocimiento automático de emociones faciales desarrollado.....	75
Fig. E. 2. Diagrama de clases de la fase 3: visualización.....	77
Fig. E. 3. Prototipo del visualizador de resultados mostrando un <i>frame</i> del vídeo y el espacio dimensional.....	80
Fig. E. 4. Prototipo del visualizador de resultados mostrando la evolución emocional temporal asociada a un conjunto de <i>frames</i> seleccionados en el <i>slider</i>	81

ÍNDICE DE TABLAS

Tabla 1. Tabla comparativa de los actuales sistemas de reconocimiento facial de emociones	7
Tabla 2. Comparativa de características de la base de datos antigua y las características deseables para la nueva Base de Datos Emocional.....	19
Tabla 3. Estudio comparativo de Bases de Datos Emocionales	20
Tabla 4. Resultados generales del benchmarking	28
Tabla A. 1. Número de horas invertidas en las tareas del proyecto	47
Tabla D. 1. Resultados de clasificación obtenidos mediante validación cruzada	72
Tabla D. 2. Resultados de clasificación obtenidos mediante conjunto completo.....	72
Tabla E. 1. Análisis de requisitos de la fase 3: clasificación.....	77
Tabla E. 2. Análisis de requisitos de la fase 4: visualización	79

CAPÍTULO 1. INTRODUCCIÓN

En este primer capítulo de la memoria se describe la motivación, el contexto y los objetivos de este Proyecto Fin de Carrera. Asimismo, la última sección presenta la estructura de esta memoria y sus anexos, explicando brevemente el contenido de cada sección.

1.1 Motivación

Dentro del área de la Interacción Persona-Ordenador (IPO) la Computación Afectiva [Pic97] es un campo emergente cuyo objetivo es el desarrollo de sistemas inteligentes capaces de dotar a un ordenador de la habilidad de reconocer, interpretar y procesar las emociones humanas. Un sistema capaz de entender las emociones del usuario presenta un gran interés para el desarrollo de nuevas interfaces que sigan de manera realista el modelo de comunicación cara-a-cara humano. Por ese motivo, la detección automática de emociones en el usuario se está convirtiendo en una parte indispensable de las interfaces persona-ordenador avanzadas. En concreto, el reconocimiento facial de emociones [GP10] ha demostrado ser una de las principales herramientas para la detección de emociones en la Computación Afectiva. Sin embargo, no se ha encontrado por el momento una solución óptima de reconocimiento automático de emociones faciales, por lo que sigue siendo un problema abierto de investigación.

Este Proyecto Fin de Carrera toma como punto de partida el trabajo previo desarrollado en la clasificación facial de emociones en una tesis doctoral [Hup10] y una Tesis Fin de Máster [Ball11], y tiene como objetivo la implementación de un sistema de reconocimiento de emociones en secuencias de vídeo tomadas por una cámara web.

1.2 Contexto del Proyecto

Este Proyecto Fin de Carrera ha sido desarrollado en el seno del *GIGA Affective Lab*, grupo de investigación perteneciente al Departamento de Ingeniería Informática de Sistemas (DIIS) de la Escuela de Ingeniería y Arquitectura (EINA) de la Universidad de Zaragoza. El trabajo está encuadrado dentro de la línea de investigación del grupo en el campo de la detección y el reconocimiento automático de emociones.

En concreto, el presente trabajo se enmarca dentro del proyecto nacional de investigación “*REPLIKANTS: Hacia una nueva generación de agentes de tipo humano (TIN2011-24660)*” y de una colaboración con el grupo GRIHO de la *Universitat de Lleida* [WGri], especializado en usabilidad; y con la empresa *BitBrain Technologies* [WBTT] líderes en interfaces cerebro-ordenador (BCI), colaboración detallada en el capítulo 4 de esta memoria.

1.3 Objetivos

El objetivo de este proyecto es desarrollar un sistema de reconocimiento facial de emociones en secuencias de vídeo basado en modelos emocionales continuos.

El grupo Affective Lab ha desarrollado en los últimos años un sistema de reconocimiento facial basado en emociones discretas, inicialmente para imágenes estáticas [Hup10], y posteriormente para vídeo [HBC11]. El objetivo de este proyecto es la migración completa del sistema del enfoque emocional discreto al continuo. Para ello, será necesario abordar las siguientes tareas:

1. Realización de un estado del arte sobre los trabajos más recientes de detección emocional.
2. Estudio exhaustivo de las Bases de Datos Emocionales disponibles, y selección de la más adecuada para el entrenamiento del nuevo sistema.
3. Desarrollo de un clasificador de emociones siguiendo un enfoque emocional continuo.
4. Implementación y puesta a punto del nuevo sistema de reconocimiento facial.
5. Aplicación del sistema al reconocimiento emocional de secuencias de vídeo procedentes de evaluaciones con usuarios.

1.4 Organización de la memoria

El resto de este documento se organiza de la siguiente forma:

- El capítulo 2 plantea el problema, describe el proceso de reconocimiento facial automático de emociones y presenta el trabajo previo.
- El capítulo 3 expone todo el trabajo realizado, dividido en fases detalladas en cada subsección.
- El capítulo 4 presenta una aplicación del sistema desarrollado en el campo de evaluación de videojuegos.
- Finalmente, el capítulo 5 contiene las conclusiones y líneas de trabajo futuro.

Se incluyen los siguientes anexos:

- A. Gestión del Proyecto. Incluye detalles sobre la planificación del proyecto y el tiempo y esfuerzo invertido en el mismo.
- B. Tecnologías y herramientas utilizadas. Presenta las tecnologías y los programas que han sido utilizados en el desarrollo del proyecto.
- C. Selección de sesiones para clasificación. Incluye todo el trabajo asociado a la selección de las sesiones para entrenar al clasificador del sistema de reconocimiento de emociones.
- D. Benchmarking del clasificador. Se presenta en este anexo todo el trabajo relacionado con la selección de atributos y del clasificador.
- E. Proceso de desarrollo del software. Incluye las fases de análisis, diseño, implementación y pruebas del software desarrollado.
- F. Publicación Interacción 2012. Contiene el artículo aceptado en el XIII Congreso Internacional de IPO, a celebrar en Elche del 3 al 5 de octubre de 2012.

CAPÍTULO 2. PLANTEAMIENTO DEL PROBLEMA: RECONOCIMIENTO FACIAL AUTOMÁTICO DE EMOCIONES

En este capítulo se introduce el proceso de reconocimiento facial de emociones y las fases en las que se divide. Se discuten a continuación las alternativas existentes para llevar a cabo cada una de las fases del proceso. Se continúa presentando una tabla con los trabajos más relevantes en el área y las alternativas seleccionadas por ellos. Finalmente, se presenta el trabajo previo desarrollado en el grupo, punto partida de este Proyecto Fin de Carrera.

2.1 Introducción

Por su importancia práctica en un amplio rango de aplicaciones, el análisis automático de expresiones faciales ha ido ganando relevancia en los últimos años. Mientras que para un ser humano resulta relativamente sencillo detectar y analizar caras y expresiones faciales, el desarrollo de un sistema automático que resuelva esta misma tarea no es trivial. A continuación se presenta el esquema general de las fases del proceso que siguen los sistemas de reconocimiento de emociones en expresiones faciales (ver Fig. 1).

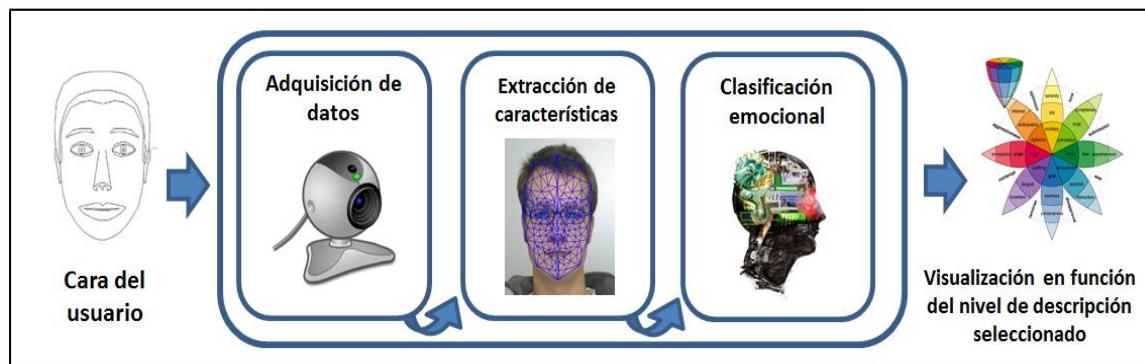


FIG. 1. PROCESO DE RECONOCIMIENTO AUTOMÁTICO DE EMOCIONES FACIALES

Para desarrollar un sistema de reconocimiento automático facial de emociones, se deben tomar decisiones sobre las diferentes alternativas existentes a la hora de llevar a cabo cada una de las fases del proceso. Las tres grandes decisiones a tomar a lo largo del desarrollo son:

1. Selección del modelo facial.
2. Selección del nivel de descripción.
3. Selección del mecanismo de clasificación.

A continuación se presentan en las siguientes secciones las diferentes alternativas existentes para cada una de las tres decisiones.

2.2 Selección del modelo facial

El primer paso en el proceso del reconocimiento facial de emociones consiste en seleccionar la entrada del sistema. Los modelos faciales son los diferentes patrones que se usan para modelar o representar el rostro. Por tanto, seleccionar un modelo facial adecuado en el reconocimiento de emociones es un punto clave, pues de dicho modelo dependerá la información que se extraiga de la cara del sujeto a analizar, siendo esta la entrada del sistema. Actualmente existen tres modelos faciales principales: (1) Modelo analítico, (2) Modelo holístico y (3) Modelo híbrido.

La representación analítica considera el rostro como un conjunto de puntos faciales. Pese a que algunos autores tienen en cuenta solamente puntos concretos del rostro [CHF06], la solución más adoptada en la literatura es introducir información sobre las variaciones, respecto a la cara neutra, de una serie de distancias y ángulos formados entre dichos puntos [HCC05] [PP06]. La principal ventaja de este modelo es su sencillez computacional, lo que lo convierte en un candidato perfecto para ser usado en tiempo real.

Por otra parte, el modelo holístico tiene en cuenta la totalidad de la cara. Los Modelos Activos de Forma [HH97] son un buen ejemplo, que parten de un modelo facial genérico ya construido (a partir de una base de datos faciales de entrenamiento) y lo proyectan sobre una imagen ajustándose al rostro que contiene. El problema de este modelo es su elevado tiempo de procesado de imagen.

Por último, el modelo híbrido es una mezcla de los dos modelos anteriores. En dicha representación, se emplea un conjunto de puntos faciales para determinar un patrón que modele el rostro. Son varios los trabajos que han optado por esta opción, entre ellos [LBA99] [ZJ05]. En la Fig. 2 pueden observarse los tres modelos faciales descritos anteriormente.

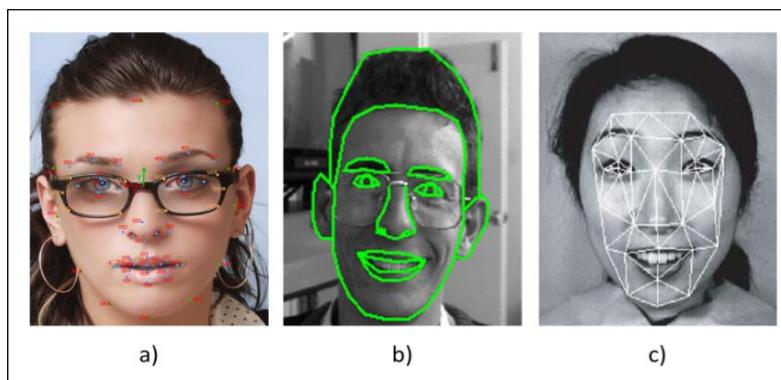


FIG. 2. MODELOS FACIALES. A) MODELO ANALÍTICO, B) MODELO HOLÍSTICO, C) MODELO HÍBRIDO.

2.3 Selección del nivel de descripción. Enfoques emocionales

Para poder realizar una correcta detección de emociones, es imprescindible definir en qué términos se describirá la información obtenida, esto es, cómo se van a representar las salidas del sistema de reconocimiento de emociones. De acuerdo con las investigaciones en psicología [GSS08], se pueden distinguir entre tres enfoques principales para modelar las emociones [GP10]: (1) enfoque categórico, (2) enfoque dimensional, (3) enfoque basado en la valoración.

El enfoque categórico o discreto se basa en la existencia de un pequeño número de emociones básicas reconocidas universalmente. Este enfoque tradicional fue promovido por Darwin [Dar98], interpretado por Tomkins [Tom62] y reforzado por Ekman [Ekm99]. De acuerdo con estos estudios, existen seis emociones básicas: alegría, tristeza, sorpresa, enfado, aversión y miedo (ver Fig. 3).



FIG. 3. FIGURA QUE MUESTRA LAS 6 EMOCIONES BÁSICAS DEFINIDAS POR EKMAN: ENFADO, SORPRESA, TRISTEZA, AVERSIÓN, MIEDO Y ALEGRÍA

El enfoque categórico supone una representación discreta donde las emociones implicadas no guardan ninguna relación entre ellas y donde es difícil reflejar el amplio rango de emociones complejas que un ser humano puede expresar.

Para superar este problema, investigadores como Whissell [Whi89], Plutchik [Plu84] o Russell [Rus80] proponen un modelo dimensional en el que las emociones no son independientes entre sí y se entrelazan de manera sistemática. El enfoque dimensional más extendido supone un espacio de dos dimensiones, cuyos ejes representan la evaluación y la activación de cada emoción. El eje de evaluación (en las abscisas) representa cómo de positiva o negativa es la emoción, mientras que el eje de activación (situado en las ordenadas) refleja el nivel de excitación de la emoción (alta o baja). En la Fig. 4 se puede observar el espacio bidimensional de Whissell, con algunas emociones, incluidas las seis emociones básicas de Ekman.

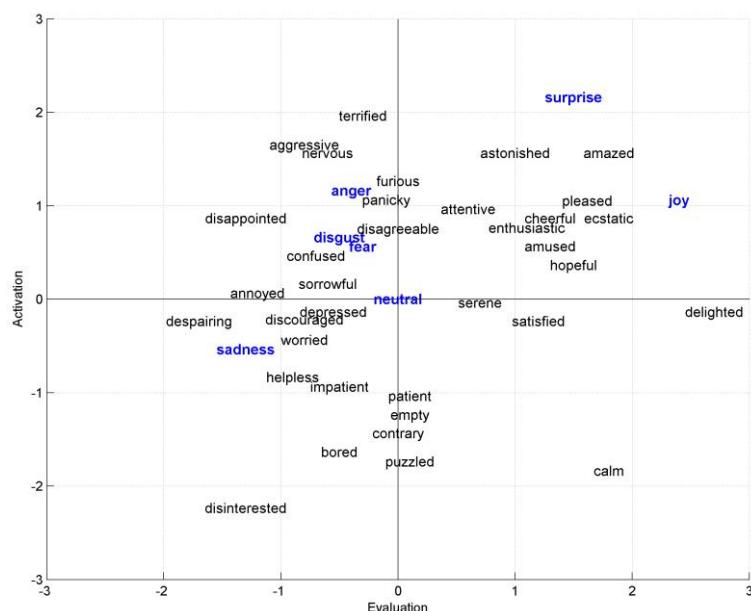


FIG. 4. ESPACIO DIMENSIONAL ACTIVACIÓN-EVALUACIÓN DE WHISSELL [WHI89]
CON LAS 6 EMOCIONES BÁSICAS (MÁS LA NEUTRAL) RESALTADAS EN ÉL.

Finalmente, autores como Scherer [SSJ01] proponen un enfoque basado en la valoración. Este modelo, que puede ser considerado como una extensión del enfoque dimensional, no se limita a considerar una serie de emociones discretas o unas pocas dimensiones afectivas. En su lugar, considera que las emociones se expresan a través de una evaluación continua y subjetiva tanto del propio estado interno del usuario como del estado del mundo exterior. En otras palabras, se centra en la variabilidad de estados emocionales, siendo posible una diferenciación de éstos modelando las diferencias individuales. Sin embargo, este enfoque se suele usar más para la generación de emociones que para la detección de éstas.

Si bien es cierto que los tres modelos son perfectamente válidos, y pese a que el enfoque categórico ha sido siempre el más utilizado a nivel de descripción, actualmente se está produciendo un movimiento hacia el enfoque dimensional en el campo de la detección de emociones. La representación dimensional proporciona una manera de describir un amplio rango de estados emocionales, así como una medida de intensidad de la emoción. Aún más importante, es capaz de representar variaciones continuas de las emociones no solamente en el espacio, sino a lo largo del tiempo.

Sin embargo, los primeros trabajos que utilizaron este enfoque se centraban más en la síntesis de caras [SSB09], procesamiento de datos [DBW07] o estudios psicológicos que en el reconocimiento emocional. De los trabajos que sí versan sobre el reconocimiento de emociones, muchos de ellos simplifican el problema a dos clases (activo vs. pasivo) [FT05] o a cuatro cuadrantes en el espacio 2D [GMK06], perdiendo parte de la información potencial que el enfoque dimensional produce.

2.4 Selección del mecanismo de clasificación

Independientemente del modelo facial y del nivel de descripción seleccionado, se debe establecer un método que clasifique las expresiones faciales del usuario. Actualmente, existen dos grupos principales de clasificadores, en función de si tienen en cuenta la información temporal de los datos faciales o no: clasificadores estáticos y clasificadores dinámicos.

Clasificadores estáticos

Son aquellos que clasifican una imagen facial o un *frame* individual de una secuencia de vídeo, sin tener en cuenta datos anteriores. Son los más usados para clasificar imágenes o *frames* de una secuencia de vídeo. Las Redes Neuronales, los Sistemas Expertos de Reglas o las Redes Bayesianas son algunos ejemplos.

Clasificadores dinámicos

Son aquellos que permiten clasificación de imágenes en función de las anteriores. La entrada de estos clasificadores es el histórico de datos de la secuencia de vídeos, introducido como una serie temporal. Los Modelos Ocultos de Markov (HMMs) son los clasificadores dinámicos más representativos.

Desde los primeros trabajos, los sistemas de clasificación estáticos han sido los utilizados en el reconocimiento automático de emociones. Sin embargo, dado que el comportamiento no verbal humano es inherentemente continuo y secuencial, muchos trabajos se han comenzado a centrar

en técnicas de clasificación dinámicas. No obstante, algunos investigadores sostienen que estos modelos dinámicos son más apropiados para un clasificador de expresiones faciales dependiente de una persona [CSG03]. De modo que si lo que se busca es un clasificador universal, las técnicas dinámicas no son una buena opción, en tanto que se debería entrenar el sistema con cada sujeto nuevo de evaluación. Además, está demostrado que la precisión en el reconocimiento de emociones aumenta considerablemente si se usan clasificadores estáticos para imágenes individuales, en comparación con el uso de modelos dinámicos alimentados con secuencias de imágenes [GP09].

Finalmente, cabe mencionar que la selección y entrenamiento del clasificador está íntimamente ligada a las Bases de Datos Emocionales (BDE), consistentes en un conjunto de imágenes de sujetos mostrando emociones y etiquetadas en uno o varios niveles de descripción que se presentan en el siguiente apartado.

2.5 Sistemas actuales de reconocimiento facial de emociones

En este apartado, se presentan los trabajos más recientes en el ámbito del reconocimiento automático de emociones, indicando para cada contribución la BDE empleada, el modelo y características faciales, su nivel de descripción y los mecanismos de clasificación usados (ver Tabla 1).

El último de ellos, desarrollado por Hupont *et al* corresponde al trabajo previo que se presenta en el siguiente apartado.

TABLA 1. TABLA COMPARATIVA DE LOS ACTUALES SISTEMAS DE RECONOCIMIENTO FACIAL DE EMOCIONES

Autores	Base de Datos	Modelo y características faciales	Mecanismo de clasificación	Nivel de Descripción
Ray y Chakrabarti [RC12]	Propia	Analítico. 81 puntos faciales	Estático (SVM)	Continuo: Evaluación y Activación.
Nicolau, Gunes y Pantic [NGP11]	<i>SAL Database</i>	Analítico. 20 puntos faciales y 4 puntos en los hombros	Estático (SVR)	Continuo: Evaluación y Activación
Valstar <i>et al</i> [VJM11]	<i>MMI Facial Expression Database</i>	Analítico. Basado en 4 expresiones faciales	Estático (SVM)	Discreto: 5 emociones (enfado, miedo, alegría, alivio y tristeza)
Gurpreet y Baljit Singh [GSB11]	<i>Cohn-Kanade Database</i>	Sistema FACS	Estático (Árbol de Decisión)	Discreto: 4 emociones (alegría, tristeza, enfado y estado neutral)
Hupont <i>et al</i> [HCB11]	<i>MMI Facial Expression Database</i>	Analítico. 7 distancias faciales	Estático (RIPPER, MLP, NB, C4.5, SVM)	Discreto: 6 emociones básicas y estado neutral.

2.6 Trabajo previo

Se presenta en este punto el trabajo previo desarrollado por el grupo de investigación, enmarcado en una tesis doctoral [Hup10] y un trabajo fin de máster [Ball11], que son el punto de partida de este trabajo. Para ello, se presenta el esquema seguido por el sistema previo de reconocimiento de emociones y se discuten las modificaciones necesarias para llevar a cabo la migración completa del enfoque discreto al continuo.

2.6.1 Sistema previo de reconocimiento de emociones

El trabajo previo en reconocimiento de emociones ha dado lugar a un sistema de detección multimodal [HBC11] (no solamente facial) en el que intervienen varios canales para el análisis emocional, realizando una clasificación emocional discreta. Consta de cinco etapas bien diferenciadas divididas en cinco programas, donde los datos iniciales de entrada se van transformando en cada etapa hasta la visualización de resultados (ver Fig. 5)

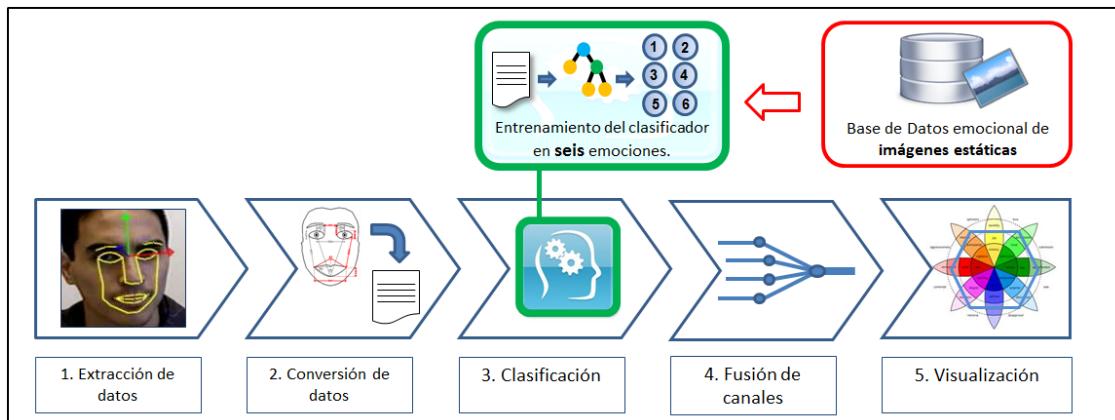


FIG. 5. ESQUEMA GENERAL DEL TRABAJO PREVIO DESARROLLADO EN RECONOCIMIENTO EMOCIONAL

En la primera fase se extraen los datos del vídeo con un reconocedor facial automático y se almacenan en un archivo, para cada *frame*, las coordenadas de los puntos faciales. La siguiente fase toma las coordenadas faciales y las convierte en las distancias que alimentarán al clasificador. Durante la fase de clasificación, el sistema toma las distancias de cada *frame* y las etiqueta en una de las seis emociones básicas de Ekman (más una emoción neutra) según el enfoque categórico. La siguiente fase es la encargada de fusionar otros datos emocionales, de manera que otros canales como voz o texto, simultáneamente extraídos con las expresiones faciales, tengan también su peso en la detección del estado emocional del usuario. Finalmente, la última fase visualiza *frame a frame* la evolución emocional del usuario. Cabe añadir que en la fase 3, el sistema de clasificación es alimentado con un clasificador previamente entrenado con el material de una base de datos de imágenes etiquetada de acuerdo con el enfoque emocional discreto.

Los dos trabajos previos hacen uso de un modelo facial analítico, utilizan clasificadores estáticos y tienen un nivel de descripción categórico. La diferencia entre la Tesis Doctoral realizada y el Trabajo Fin de Máster radica en que este último realizaba la extracción de datos de manera

automática, mientras que en el trabajo desarrollado en la Tesis Doctoral, la extracción de características faciales era completamente manual. Sin embargo, sus principales limitaciones residen en que en ambos trabajos el clasificador ha sido entrenado con imágenes de Bases de Datos Emocionales actuadas y discretas (ver sección 3.4 de esta memoria).

2.6.2 *Migración completa del discreto al continuo*

Como se ha visto anteriormente el enfoque categórico tiene limitaciones que se pueden resolver pasando al enfoque continuo, que permite una descripción emocional más rica y una representación adecuada de las variaciones emocionales a lo largo del tiempo.

El objeto de este proyecto ha sido llevar a cabo la migración completa del sistema de reconocimiento facial del enfoque discreto al continuo. Ello ha implicado modificar algunas fases del sistema anterior y desarrollar algunas nuevas (ver Fig. 6).

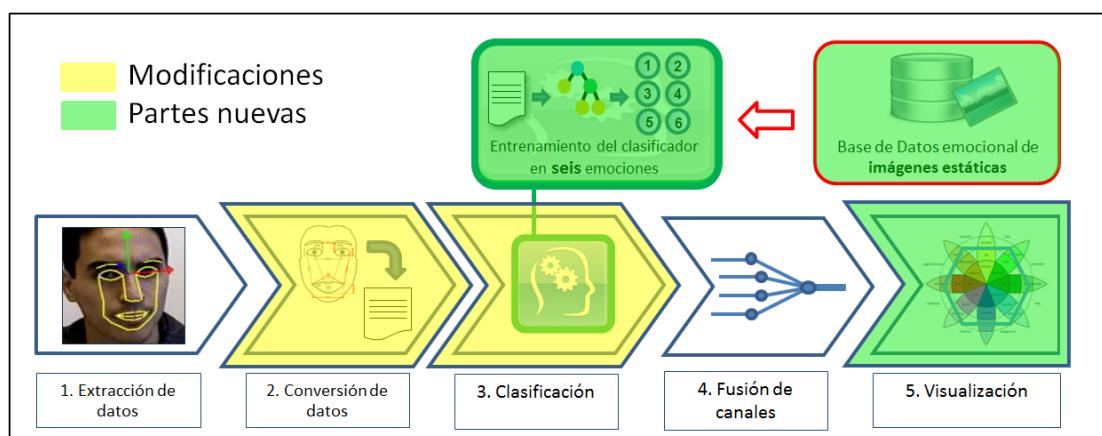


FIG. 6. CAMBIOS NECESARIOS EN EL SISTEMA ANTERIOR PARA UNA COMPLETA MIGRACIÓN AL ENFOQUE CONTINUO

Como puede observarse en la Fig. 6, es necesario introducir importantes cambios en el sistema anterior. Concretamente:

- Modificación de los módulos 2 y 3 correspondientes a la conversión de datos faciales y clasificación de éstos.
- Selección de una nueva Base de Datos emocional de vídeos etiquetada en términos acordes al nivel de descripción continuo.
- Desarrollo de un nuevo sistema de clasificación estático, alimentado con los datos procesados de la nueva Base de Datos Emocional.
- Desarrollo de un visualizador de resultados en el espacio dimensional.

En la fase de extracción de datos, no es necesario realizar mayores cambios. En cuanto al módulo de fusión, éste estaba ya preparado para el trabajo en continuo. Al ser el objetivo de este proyecto el reconocimiento facial, se ha dejado de lado en este trabajo.

Una vez realizados los cambios y desarrolladas las nuevas partes, se integra todo en un nuevo sistema de reconocimiento de emociones con un enfoque emocional completamente continuo, tal como se detalla en el capítulo siguiente

CAPÍTULO 3. TRABAJO REALIZADO

Este capítulo presenta todo el trabajo realizado durante el desarrollo del Proyecto Fin de Carrera y está estructurado en varios apartados. El primero de ellos muestra un esquema general del sistema implementado, con las diferentes partes que lo componen. A continuación cada apartado detalla el trabajo desarrollado para cada una de las partes. El último punto se encarga de explicar la validación del sistema mediante videos obtenidos de diferentes fuentes.

3.1 Esquema del sistema de reconocimiento facial de emociones en secuencias de vídeo.

Una vez introducido el trabajo previo en el capítulo anterior, en la Fig. 7 se presenta el esquema general del sistema:

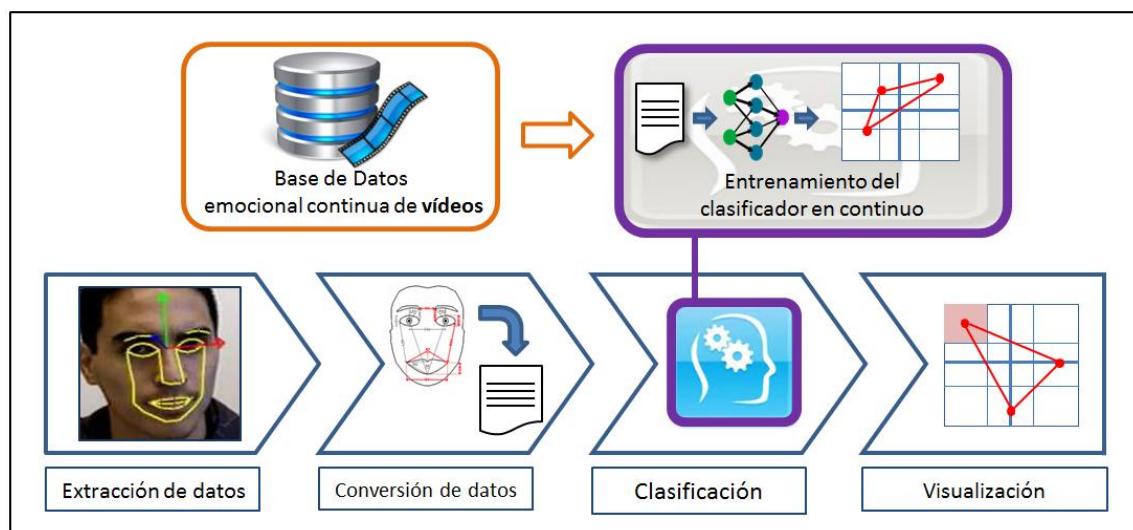


FIG. 7. EQUEMA GENERAL DEL SISTEMA DE RECONOCIMIENTO FACIAL DE EMOCIONES EN SECUENCIAS DE VÍDEO

Este nuevo sistema utiliza un modelo facial analítico (extracción de puntos faciales), y utiliza mecanismos de clasificación estáticos. Sin embargo, a diferencia de los dos trabajos previos presentados en el capítulo anterior, el nivel de descripción es continuo o dimensional. Esto implica realizar cambios en las dos últimas fases (clasificación y visualización), así como la selección de una Base de Datos Emocional y el entrenamiento del nuevo clasificador en continuo.

Como se puede observar en la Fig. 7, el proceso general consta de cuatro fases bien diferenciadas que siguen un proceso secuencial. Cada una de ellas corresponde a un programa diferente, cuyos resultados se almacenan en ficheros y sirven como entrada al programa correspondiente a la siguiente fase.

Cada una de las fases cumple una función en el proceso de reconocimiento emocional. Concretamente:

- La fase 1, extracción de datos, se encarga de extraer la información facial de los vídeos en forma de puntos para cada *frame* y almacenarlos en un fichero. Esta fase está detallada en el apartado 3.2.
- La fase 2 se encarga de la conversión de datos. Este módulo toma el fichero de puntos faciales producidos por la fase 1 y los convierte en distancias y ángulos faciales almacenados en un fichero que alimentará a la siguiente fase de clasificación. Los detalles de la fase 2 se encuentran recogidos en la sección 3.3.
- La fase 3 (sección 3.6) se encarga de tomar el fichero creado por la fase 2 y clasificar cada *frame* en el espacio emocional continuo, almacenando los resultados en un fichero. Para poder llevar a cabo la clasificación emocional de un *frame* es necesario haber entrenado previamente el clasificador (sección 3.5) con una Base de Datos Emocional adecuada (sección 3.4).
- Finalmente, la fase 4 se encarga de tomar el fichero creado en la fase anterior y visualizar el estado emocional del usuario en el espacio dimensional. Los detalles del proceso de esta última fase se encuentran en la sección 3.7 de esta memoria.

3.2 Extracción de datos

La extracción de datos es la primera de las fases en el sistema de reconocimiento de emociones faciales. El objetivo de esta etapa es convertir el vídeo de entrada donde aparece el usuario al que se quiere realizar la evaluación emocional, en un fichero que contenga las coordenadas de los puntos faciales extraídos para cada *frame* del vídeo. El esquema detallado del proceso de extracción se puede observar en la Fig. 8.



FIG. 8. ESQUEMA GENERAL DE LA FASE 1: EXTRACCIÓN DE DATOS

A continuación se explica en detalle el trabajo llevado a cabo en cada parte de esta fase.

3.2.1 División del vídeo en frames

Para realizar una extracción de datos de una secuencia de vídeo, es necesario primero convertir el vídeo en una secuencia de imágenes o *frames*. Esto es debido a que los clasificadores estáticos clasifican las distancias y ángulos faciales asociados a un *frame*, sin tener en cuenta la relación de un *frame* con el siguiente.

Para realizar esta tarea, se ha utilizado el software *VirtualDub* (más información en el Anexo B), una herramienta que permite dividir un vídeo en una secuencia de imágenes en función de varios parámetros. En este caso, se tienen en cuenta las siguientes restricciones:

1. La resolución de salida de la imagen debe ser la misma que la resolución del vídeo.
2. La medida de frecuencia FPS (Fotogramas Por Segundo o *Frames Per Second*) debe ser un valor comprendido entre los 30 y los 60 FPS. Estos valores han sido seleccionados para poder captar aquellas emociones que ocurren muy deprisa (sorpresa o aversión) y que podrían perderse con una frecuencia menor.
3. Durante los primeros 3 segundos del vídeo el sujeto de evaluación debe mostrar una expresión relajada y neutra. Es muy importante que así sea dado que los *frames* asociados a los primeros segundos se utilizan para establecer la cara neutra del sujeto y poder medir luego todas las emociones que exprese durante el vídeo en relación a su cara neutra.

Una vez realizada la división del vídeo y con todas las imágenes almacenadas en una carpeta, se procede a la extracción de puntos faciales de cada una de las imágenes.

3.2.2 Extracción de puntos faciales mediante FaceAPI

La tecnología usada para el reconocimiento automático es *FaceAPI*, un software comercial de la empresa *SeeingMachines* [FASM]. Esta herramienta realiza un seguimiento de la cara proporcionando para cada *frame* una serie de puntos faciales llamados *landmarks*. En esta herramienta de pago el número de *landmarks* que se extraen varía en función del tipo de licencia que se posea. La licencia adquirida por el *GIGA Affective Lab* solamente permite proporcionar 38 puntos interiores del rostro, que pueden observarse en la Fig. 9.

Los puntos faciales obtenidos se presentan en dos formatos de coordenadas distintos:

1. *Pixel Coordinate Frame*. Coordenadas en el espacio 2D.
2. *World Coordinate Frame*. Coordenadas en el espacio 3D.

En este trabajo se ha decidido emplear el sistema de coordenadas World Coordinate Frame (3D) para representar los puntos faciales, ya que es mucho más preciso que el sistema *Pixel Coordinate Frame* en tanto que tiene una dimensión más en la representación. Hay que tener en cuenta, que durante el reconocimiento facial, lo más normal es que el sujeto esté ubicado delante de la cámara mirando hacia ella, pero en algunas ocasiones ocurre que el sujeto cambia de posición, moviendo su plano facial y cambiando la orientación de su mirada. En estas ocasiones la dimensión extra de profundidad que proporciona el *World Coordinate Frame* es imprescindible para no cometer errores de proyección. Por tanto, teniendo en cuenta las tres

dimensiones, es posible asegurar un reconocimiento mucho más robusto y con menos errores que si sólo se usasen las dimensiones ofrecidas por el sistema de coordenadas de píxeles.

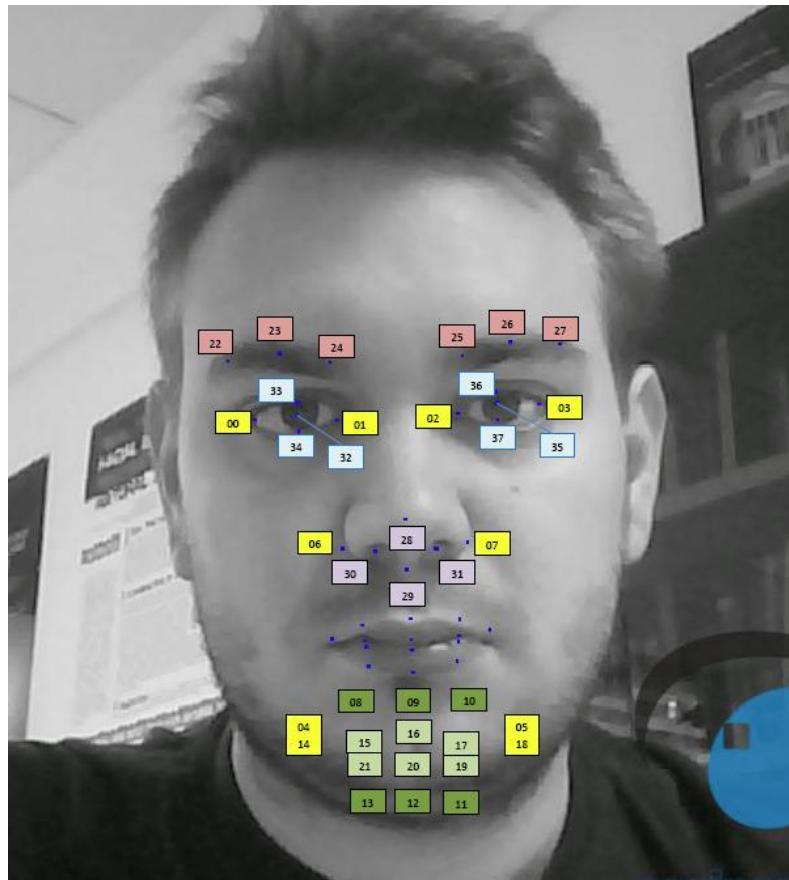


FIG. 9. 38 PUNTOS FACIALES PROPORCIONADOS POR LA TECNOLOGÍA FACEAPI DE SEEINGMACHINES

3.2.3 Escritura de los puntos faciales en fichero

Una vez que se han extraído los puntos faciales del rostro para cada *frame*, se debe escribir en un fichero para que la siguiente fase de conversión de datos pueda usarlos como entrada. Para ello, para cada *frame* se almacena: (1) una estampilla que indica la posición del *frame* en la secuencia de imágenes; (2) lista con las coordenadas 3D de cada uno de los 38 puntos.

3.3 Conversión de datos

La segunda fase en el sistema de reconocimiento de emociones faciales corresponde a la conversión de datos. El objetivo de esta fase es obtener, para cada *frame*, las variaciones de las distancias y ángulos (obtenidos de los 38 puntos faciales) respecto de la cara neutra. Para ello, se toma el fichero de datos de puntos faciales obtenido en la fase anterior como entrada. La salida es un fichero que contiene las distancias y ángulos que alimentarán al clasificador en la fase 3. La Fig. 10 muestra el esquema detallado del proceso de conversión de datos.

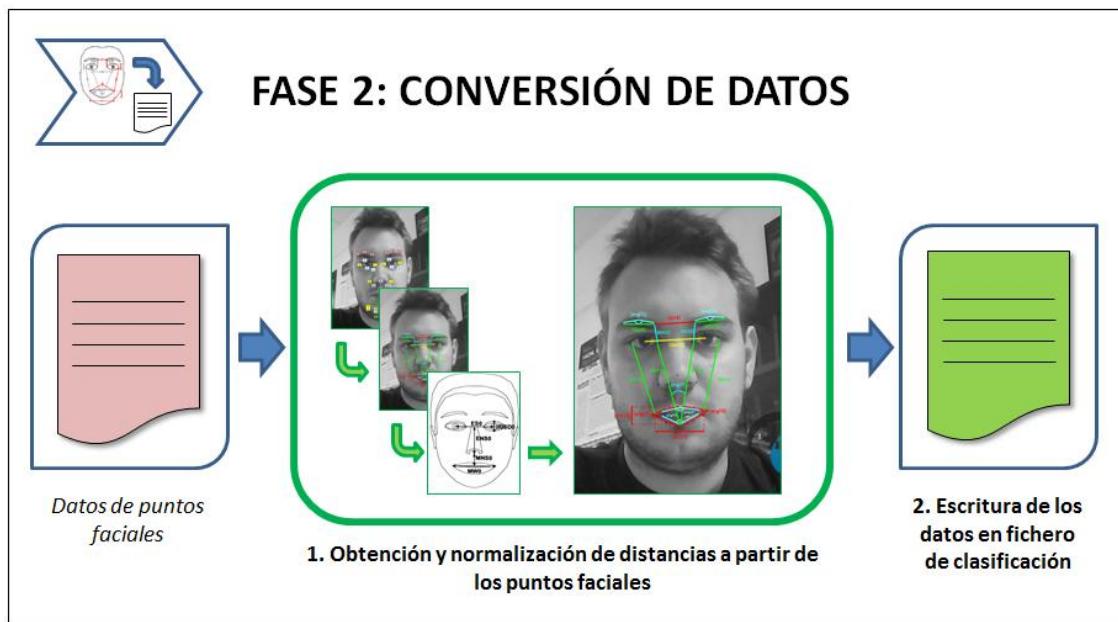


FIG. 10. ESQUEMA GENERAL DE LA FASE 2: CONVERSIÓN DE DATOS

A continuación se detallan los puntos de la fase 2 de conversión de datos.

3.3.1 Elaboración y normalización de distancias

Distancias y ángulos empleados

Como ya se ha comentado en el capítulo 2 de esta memoria, la representación analítica considera el rostro como un conjunto de puntos faciales. La solución que se adopta en el presente trabajo consiste en convertir primero los puntos faciales obtenidos en cada *frame* a un conjunto de distancias y ángulos faciales, y obtener sus variaciones respecto de las distancias y ángulos faciales de la cara neutra del sujeto.

Se obtienen 19 ángulos y distancias faciales a partir de los 38 *landmarks* expuestos en el apartado anterior. Estos 19 ángulos y distancias han sido seleccionados teniendo en cuenta los trabajos previos [Hup10][Ball11] (no obstante, al clasificador final no le entrarán todas estas distancias, como se verá más adelante).

Las distancias pueden observarse en la Fig. 11, divididas en ángulos y rectas. A continuación se explica cómo se han obtenido tanto los ángulos como las distancias faciales:

Para obtener el ángulo entre dos vectores en tres dimensiones, primero se calcula los vectores u, v y posteriormente el ángulo θ determinado por sus direcciones, que viene dado por la expresión:

$$\theta = \sec \frac{u \cdot v}{|u| \cdot |v|}$$

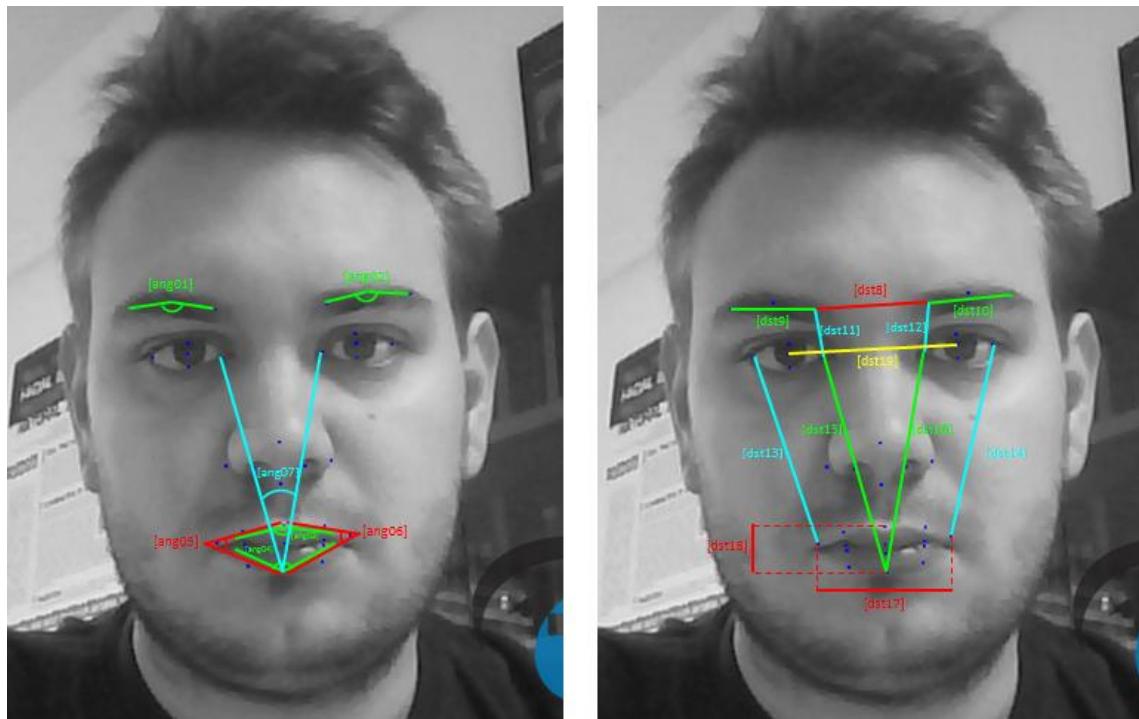


FIG. 11. LOS ÁNGULOS [ANG 01-07] A LA IZQUIERDA. LAS DISTANCIAS FACIALES (DST 08-19) A LA DERECHA.

La obtención de las distancias faciales entre dos puntos a, b en el espacio 3D, se hace utilizando la expresión de distancia euclídea:

$$d = \sqrt{(a.x - b.x)^2 + (a.y - b.y)^2 + (a.z - b.z)^2}$$

Normalización de distancias

Los valores de las distancias calculadas deben ser consistentes independientemente de la escala de la imagen, la distancia a la cámara, etc. Por ello, todas las distancias deben ser normalizadas respecto a la FAPU “ESO” del estándar MPEG4 [MPEG4], una distancia constante en el rostro e invariante con respecto a la expresión facial mostrada. El empleo de ángulos proporciona ya de por sí valores invariantes, evitando el esfuerzo de normalización.

Para normalizar las distancias deberá tenerse la distancia ESO (distancia nº 19 en la Fig. 11) del sujeto con expresión neutra, de modo que solamente es necesario normalizar las distancias 08 a 18. Esto se consigue al principio del reconocimiento durante los primeros *frames* del vídeo, donde es requisito que el sujeto mantenga una expresión facial neutra. Estos primeros segundos de cara neutra permiten hallar, para cada distancia, un valor medio de dicha distancia o ángulo en expresión neutra (el valor medio es necesario dado que tomar las distancias de un único *frame* sería arriesgarse a que justamente ese *frame* no representara correctamente una expresión neutra) Una vez obtenida dicha información, se normaliza y se calcula la variación de cada distancia y ángulo mediante las expresiones que se exponen a continuación.

Cálculo de la variación de los ángulos (distancias 1 a 7) respecto a la cara neutra. Obsérvese que los ángulos no precisan de normalización:

$$\Delta\theta_i \text{actual} = \theta_i \text{actual} - \theta_i \text{neutro}; \quad 1 \leq i \leq 7$$

Normalización y cálculo de la variación de las distancias faciales (distancias 8 a 18) respecto a la cara neutra:

$$\Delta d_i \text{actual} = \frac{d_i \text{actual} - ESO \text{actual}}{d_i \text{neutro} - ESO \text{neutro}}; \quad 8 \leq i \leq 18$$

Mediante la normalización de las variaciones se garantiza que los datos pasados al clasificador son consistentes y robustos, y que la fisionomía del sujeto o la distancia a la cámara no influirán en la salida del clasificador emocional.

3.3.2 Escritura del fichero de clasificación

La salida del programa de conversión de datos es un fichero que contiene, para cada *frame*, la variación de las distancias faciales de dicho *frame* respecto a las distancias de la cara neutra. El fichero es de tipo ARFF (*Atribute-Relation File Format*) [ARFF], y está organizado en tres secciones. En la primera de ellas aparecerá el nombre del fichero y la información necesaria para describir su contenido. En la segunda sección se indican las distancias a tener en cuenta en la clasificación, también llamados atributos. Por último, la tercera sección contiene toda la información de los *frames*, de manera que en cada línea se encuentran cada una de las variaciones de las distancias calculadas. La Fig. 12 refleja la estructura de un fichero ARFF.

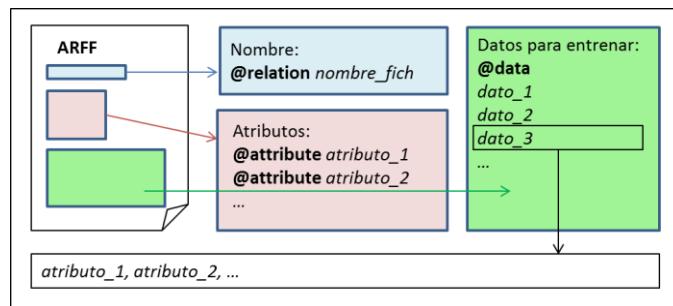


FIG. 12. ESTRUCTURA DE UN FICHERO ARFF

3.4 Selección de una Base de Datos Emocional

Como se ha comentado, es necesario entrenar el clasificador, por lo que es necesario una Base de Datos Emocional. Los resultados del sistema de reconocimiento automático de emociones están estrechamente ligados con la calidad de la base de datos finalmente seleccionada, por lo que es importante la selección de la misma. Es por esto que se ha llevado a cabo un estudio exhaustivo de las Bases de Datos Emocionales existentes en la actualidad, estudio que se detalla a continuación.

3.4.1 Características de las Bases de Datos Emocionales

Una Base de Datos Emocional (BDE) es un conjunto de archivos presentado en uno o varios formatos que contienen información emocional sobre los sujetos que han sido grabados para su elaboración. Esta información emocional se puede presentar en varios canales en forma de expresiones faciales, voz, gestos corporales o incluso señales fisiológicas. Generalmente, los archivos de una BDE suelen estar anotados, proporcionando información útil como el tipo de emoción sentida por el sujeto, información sobre éste o el método de elicitation de emociones empleado.

A continuación se presentan las características a analizar en una BDE de cara a compararla con otras:

- **Formato:** es el formato en el que se presentan los archivos que contienen la información emocional. El formato puede ser vídeo, audio, imagen, texto, archivos específicos a un canal (señales fisiológicas) o cualquier combinación de éstos.
- **Modalidad:** indica qué canales de análisis emocional intervienen (e.g. expresiones faciales, lenguaje corporal, voz, señales fisiológicas... etc). Cuando una BDE contiene varios canales, se dice que es una BDE multimodal.
- **Tamaño:** número de muestras con información útil que contiene la base de datos.
- **Nº de sujetos:** en este campo se especifica el número de sujetos de los que se ha tomado información emocional para la elaboración de la base de datos.
- **Nivel de descripción:** indica el enfoque emocional en el que está anotada la base de datos, así como los distintos tipos de emociones que se evalúan en ésta.
- **Universalidad:** se dice que una base de datos emocional es universal si contiene archivos que provengan de sujetos de diferente edad, fisionomía, sexo u origen étnico.
- **Método de elicitation:** hace referencia al modo en que se extrae la información emocional de los sujetos. La elicitation se puede hacer de tres maneras: natural, donde el sujeto expresa sus emociones de manera espontánea; inducida, donde se lleva al sujeto a sentir la emoción buscada; y posado, donde el sujeto es un actor que está fingiendo la emoción.
- **Tipo de anotación:** indica la autoría de la anotación de los archivos. Las anotaciones pueden ser llevadas a cabo por anotadores expertos, por los sujetos evaluados o por ambas partes.
- **Accesibilidad:** indica la disponibilidad de la base de datos. Puede estar disponible para todos los públicos, para personal de investigación bajo pedido al autor o pudiendo ser de pago.
- **Lenguaje:** el idioma en el que está anotado la Base de Datos Emocional

3.4.2 Aspectos deseables en una Base de Datos emocional

Se quiere crear un nuevo sistema de reconocimiento de emociones realizando una migración del enfoque emocional discreto al enfoque continuo. Por ello, la selección una nueva Base de Datos emocional adecuada se convierte en un aspecto de vital importancia, pues de ella dependerá el clasificador y el nivel de descripción del sistema.

En la Tabla 2 se presenta la comparativa entre las características de la Base de Datos empleada para entrenar el sistema de clasificación anterior y las características deseables que debería tener la nueva BDE.

**TABLA 2. COMPARATIVA DE CARACTERÍSTICAS DE LA BASE DE DATOS ANTIGUA
Y LAS CARACTERÍSTICAS DESEABLES PARA LA NUEVA BASE DE DATOS EMOCIONAL**

Nombre	Anterior: <i>MMI Database</i>	Futura: características deseables
Formato	Imágenes y vídeos	Varios formatos
Modalidad	Expresiones faciales	Multimodal
Tamaño	1280	Mayor número posible
Nº Sujetos	43	Mayor número posible
Contenido emocional	Enfoque categórico. 6 emociones básicas	Enfoque continuo. Mayor nº posible de dimensiones
Universalidad	Si	Sí
Método de elicitation	Fingido	Natural
Tipo de anotación	Anotación realizada por expertos	Anotación realizada por el sujeto y por expertos
Accesibilidad	Bajo petición	Disponible o Bajo Petición
Lenguaje	Inglés	Indiferente

Por tanto, y a la vista de la comparación entre la antigua base de datos y las características deseables de la nueva a seleccionar, la diferencia fundamental es el contenido emocional, por lo que se buscan aquellas que tengan vídeos etiquetados con un enfoque dimensional. El método de elicitation debe ser natural, no fingido, debido a que se persigue el desarrollo de un clasificador entrenado con expresiones naturales para que posteriormente se puedan clasificar expresiones del mismo tipo, en situaciones cotidianas. También es conveniente que los vídeos estén etiquetados por los propios usuarios además de los anotadores expertos, pues nadie mejor que ellos conoce las emociones sentidas. Además, es deseable que la Base de Datos sea Multimodal y contenga no sólo información de expresiones faciales. De este modo, y de cara líneas de trabajo futuro, se podrá realizar una fusión de la información obtenida de cada canal para enriquecer el reconocimiento emocional.

3.4.3 Estudio comparativo de las actuales Bases de Datos Emocionales

En este trabajo, se ha realizado un estudio exhaustivo de las Bases de Datos Emocionales existentes en la literatura. Para ello, en algunas ocasiones se ha tenido que contactar con el equipo de investigación que ha desarrollado la BDE para obtenerla mediante accesos a servidores o incluso envío de los datos en formato DVD. En otras ocasiones la adecuación de la información de la base de datos se podía probar mediante las muestras que se ofrecían desde los diferentes sitios web. La Tabla 3 refleja el estudio comparativo de 25 las Bases de Datos Emocionales, en la cual se han tenido en cuenta las características antes mencionadas.

Teniendo en cuenta los aspectos que interesan para el nuevo clasificador, se ha elegido la base de datos multimodal *MAHNOB-HCI Tagging Database*, que se describe en detalle en el siguiente apartado.

TABLA 3. ESTUDIO COMPARATIVO DE BASES DE DATOS EMOCIONALES

Nombre	Formato	Modalidad	Tamaño	Nº Sujetos	Enfoque emocional	Univers.	Método de elicitación	Anotación hecha por	Acces.	Lenguaje	Ref.
MAHNOB-HCI Tagging Database	Vídeo, Audio, Texto	Facial, Voz, Mirada, Fisiológico	500 sesiones	27 (15H - 12M)	Continuo: 4 dimensiones Discreto: 8 emociones	Sí	Natural	Sujetos	Bajo demanda	Inglés	[SLPP12] [LS11] [BDE-1]
HUMAINE Database	Vídeo	Facial, Corporal	40 sesiones	40 (23H - 17M)	Categórico: 6 emociones	No	Natural	Expertos	Disponible	Inglés, hebreo, francés.	[DCSC07] [BDE-2]
Belfast Naturalistic Database	Vídeo	Facial, Corporal	125 sesiones	125 (31H - 94M)	Categórico: 6 emociones	No	Natural	Expertos	Disponible	Inglés	[DCSC07] [BDE-3]
SAL database	Vídeo	Facial	80 sesiones	20	Continuo: 1 dimensión	Si	Inducido	Expertos	Bajo demanda	Inglés, griego.	[BDE-4]
ORESTEIA database	Audio, Texto	Voz, Fisiológico	29 sesiones	29	Categórico: 6 emociones	No	Inducido	Expertos	Bajo demanda	Inglés y griego	[BDE-5]
XM2VTSDB MM Face Database	Video, Imagen, Audio, 3D	Facial, Voz, Corporal	Desconocido	295	Sin etiquetar	Sí	Posado	Sin anotar	De pago	Inglés	[BDE-6]
The AR Face Database	Imagen	Facial	3.016 imágenes	116 (63H - 53H)	Categórico: 4 emociones	Sí	Posado	Expertos	Bajo demanda	Inglés	[BDE-7]
CVL Face Database	Imagen	Facial	756 imágenes	116 (108H - 6M)	Categórico: 1 emoción	No	Posado	Expertos	Bajo demanda	Inglés	[BDE-8]
The Psychological Image Collection at Stirling	Imagen	Facial	100 imágenes	90	Categórico: 3 emociones	Sí	Posado	Expertos	Bajo demanda	Inglés	[BDE-9]
The Japanese Female Facial Expression (JAFFE) Database	Imagen	Facial	70 imágenes	10 (0H - 10M)	Categórico: 6 emociones	No	Posado	Expertos	Bajo demanda	Inglés	[BDE-10]
CMU PIE Database	Imagen	Facial	40.000 imágenes	68	Categórico: 4 emociones	Sí	Posado	Expertos	Bajo demanda	Inglés	[SBB02] [BDE-11]

Nombre	Formato	Modalidad	Tamaño	Nº Sujetos	Enfoque emocional	Univers.	Método de elicitación	Anotación hecha por	Acces.	Lenguaje	Ref
Indian Institute of Technology Kanpur Database	Imagen	Facial	240 imágenes	40 (20H - 20M)	Categórico: 6 emociones	No	Posado	Expertos	Disponible	Inglés	[BDE-12]
The Yale Face Database	Imagen	Facial	165 imágenes	15	Categórico: 3 emociones	No	Posado	Expertos	Bajo demanda	Inglés	[BDE-13]
Caltech Frontal Face DB	Imagen	Facial	450 imágenes	27	Sin etiquetar	No	Posado	Sin anotar	Disponible	Inglés	[BDE-14]
HumanScan Bioid Face DB	Imagen	Facial	1500 imágenes	23	Sin etiquetar	No	Posado	Sin anotar	Disponible	Inglés	[BDE-15]
Oulu University Physics-Based Face Database	Imagen	Facial	1200 imágenes	120	Sin etiquetar	Sí	Posado	Sin anotar	De pago	Inglés	[BDE-16]
The Bosphorus Database	Imagen	Facial	4666 imágenes	105	Sin etiquetar	Sí	Posado	Sin anotar	Bajo demanda	Inglés	[BDE-17]
The Database of Faces AT&T	Imagen	Facial	400 imágenes	40	Sin etiquetar	Sí	Posado	Sin anotar	Disponible	Inglés	[BDE-18]
The Yale Face Database B	Imagen	Facial	5760 imágenes	10	Sin etiquetar	No	Posado	Sin anotar	Bajo demanda	Inglés	[BDE-19]
FABO Database	Vídeo	Facial, Corporal	1094 vídeos	23	Categórico: 10 emociones	Sí	Posado	Sujetos y expertos	Bajo demanda	Inglés	[GP06] [BDE-20]
Semaine	Vídeo	Facial, Corporal	50 sesiones	20	Continuo: 5 dimensiones Categórico: 6 emociones	Sí	Natural	Sujetos	Disponible	Inglés	[MVP10] [BDE-21]
MMI Database	Vídeo, Imagen	Facial	740 imágenes y 848 sesiones	19	Categórico: 6 emociones	Sí	Posado	Expertos	Bajo demanda	Inglés	[PVR05] [BDE-22]
MUG Database	Video	Facial	1462 sesiones	86 (51H - 35M)	Categórico: 13 emociones	No	Natural	Expertos	Bajo demanda	Inglés	[BDE-23]
Cohn Kanade Database	Video	Facial, Corporal	500 sesiones	100	Categórico: 6 emociones	Sí	Posado	Expertos	Bajo demanda	Inglés	[KTC00] [BDE-24]
Extended Cohn Kanade Database	Video	Facial, Corporal	610 sesiones	127	Categórico: 6 emociones	Sí	Posado	Expertos	Bajo demanda	Inglés	[LCK10] [BDE-25]

3.4.4 MAHNOB-HCI Tagging Database

La *MAHNOB-HCI Tagging Database* es una BDE multimodal grabada en respuesta a estímulos afectivos con el propósito de investigar el reconocimiento de emociones y el etiquetado implícito (auto-etiquetado de la emoción sentida) de los sujetos [SLPP12].

Se ha seleccionado esta BDE por ser una de las más recientes y completas existentes en la literatura. Posee más de 3000 sesiones de vídeos (aunque actualmente están sólo 511 sesiones completas disponibles) en las que aparecen 30 sujetos diferentes. Todas las sesiones están debidamente etiquetadas por los propios sujetos en ambos enfoques discreto y continuo, y la información viene presentada en un formato *XML* completamente organizado. Asimismo, la calidad de los vídeos es muy buena (*HD 720p*) y el reconocedor facial ha respondido satisfactoriamente a las pruebas realizadas durante el estudio. Adicionalmente, la *MAHNOB-HCI Tagging DB* es una base de datos multimodal (i.e. contiene otro tipo de información emocional además de vídeos faciales), que tiene también audio grabado por el sujeto de evaluación y señales fisiológicas. Entre éstas últimas, se destacan el volumen de presión sanguínea, el pulso respiratorio, electroencefalograma, electrocardiograma o conductancia de la piel.

La anotación en continuo consiste en un valor numérico que va desde 1 hasta 9 para cada una de las 4 dimensiones en las que la BDE está etiquetada: activación, evaluación, dominancia y predictibilidad. La dimensión de activación se refiere al grado de excitación o apatía de la emoción y va desde emociones poco activas como el aburrimiento hasta emociones de máxima activación como la ira o la sorpresa. La dimensión de evaluación se refiere a la forma negativa o positiva de la emoción, oscilando desde sentimientos desagradables como tristeza hasta sentimientos agradables como la alegría. La dimensión de dominancia se refiere al grado de dominio que tiene el sujeto sobre la situación (i.e. si la situación supera al sujeto que está sintiendo la emoción o por el contrario es éste quien controla la situación). Por último, la dimensión de predictibilidad se refiere al grado en que la emoción era esperada. Esto es, un valor bajo de predictibilidad indica que la emoción era completamente inesperada mientras que un valor alto indica todo lo contrario.

3.5 Entrenamiento del clasificador continuo

Una vez vista la base de datos que se va a emplear y antes de entrar dentro de la fase de clasificación, es imprescindible entrenar al clasificador. El entrenamiento se realiza con el software de clasificación WEKA (ver Anexo B para más detalles). Los objetivos de esta etapa son los siguientes (ver Fig. 13):

1. Extraer la información facial de la BDE seleccionada.
2. Seleccionar entradas, salidas y clasificador.
 - 2.1 Seleccionar qué ángulos y distancias (atributos) serán las entradas del clasificador.
 - 2.2 Definir cómo será la salida del clasificador.
 - 2.3 Seleccionar un método de clasificación de entre los disponibles en WEKA.
3. Extraer la información de los resultados obtenidos, necesaria para la clasificación posterior.

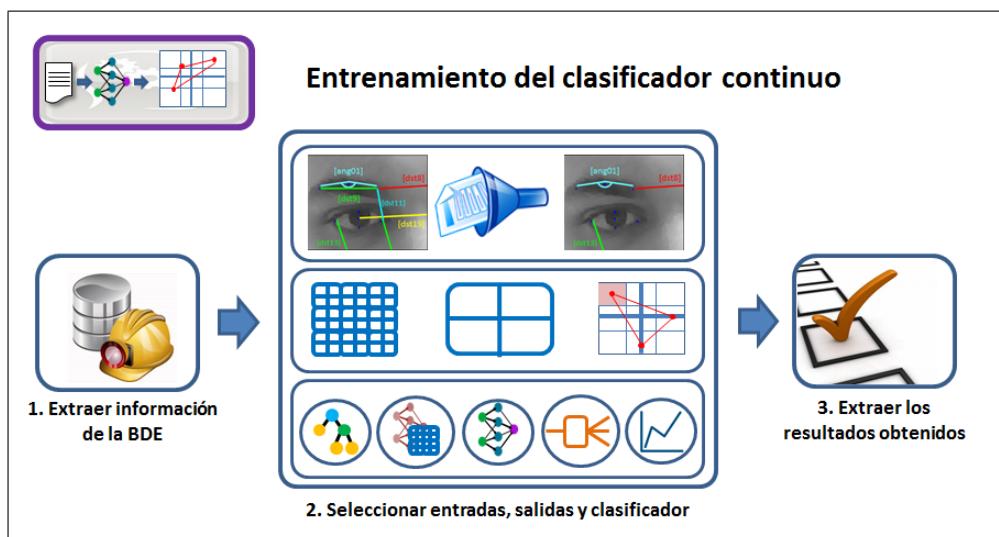


FIG. 13. PROCESO DE ENTRENAMIENTO DEL CLASIFICADOR CONTINUO

3.5.1 Extracción de la información de la BDE

La cantidad de datos disponible en la MAHNOB-HCI Tagging Database es enorme, de modo que antes de ver qué se hace con los datos, se explica brevemente cómo están dispuestos.

Los datos están disponibles *online* una vez conseguido el permiso de los autores y se pueden descargar en forma de sesiones. Una sesión recoge toda la información disponible correspondiente a la grabación de una toma de datos. Por toma de datos, se entiende la recogida de información emocional de los distintos sensores colocados en el sujeto durante la muestra de un vídeo, siendo éste el estímulo que provoca las reacciones emocionales en el sujeto.

Existen dos tipos principales de información en una sesión: (1) información descriptiva de las sesiones; (2) material multimedia de las sesiones. El primer tipo de información viene en formato *XML* [*WXML*] y se encarga de recoger los datos que identifican la sesión y el sujeto, además de las anotaciones emocionales realizadas por éste. El segundo tipo de información son los archivos grabados por los diferentes sensores, estando disponibles para cada sesión 6 vídeos faciales, un fichero que contiene toda la información de los sensores fisiológicos y un archivo de audio correspondiente al micrófono empleado en las sesiones.

El objetivo de la extracción es convertir sesiones de la BDE en conjuntos de distancias que servirán de entrenamiento al clasificador. Para ello, se seguirá el proceso que muestra la Fig. 14, que consta de 4 etapas:

1. En la primera etapa se extraen 38 puntos faciales de cada *frame* del vídeo, de manera análoga al proceso de extracción de datos del sistema general.
2. La segunda etapa toma los puntos de cada *frame* individual y los convierte en distancias, tal y como se hace en la fase de conversión de datos del sistema general.
3. La tercera etapa consiste en seleccionar manualmente *aquel* *frame* que contenga el pico de la emoción, también llamado *appex*. La selección del *appex* es manual, por lo que es

necesario ver el vídeo previamente y seleccionar después el *frame*. Esta es probablemente la fase más costosa, en tanto que no se puede automatizar.

4. Por último, una vez se sepa el frame que contiene el *appex* de la emoción, se copiarán las distancias obtenidas de éste a un fichero.

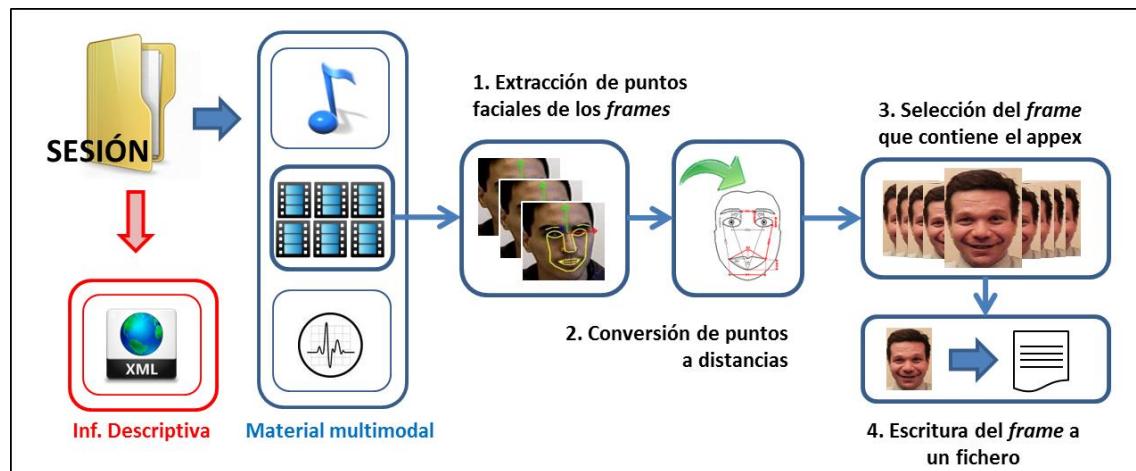


FIG. 14. PROCESO DE EXTRACCIÓN DE INFORMACIÓN DE LA BDE

El proceso explicado previamente debe realizarse para cada una de las sesiones. Además, se ha tomado la decisión de incluir solamente un frame por sesión, de modo que en el archivo final de datos de entrenamiento no se encuentren dos de ellos que provengan de la misma sesión. Aunque es cierto que aumentar el número de datos es beneficioso para el entrenamiento, se estaría creando dependencia entre dos pares de datos, resultando en el empeoramiento de la clasificación.

De las 511 sesiones disponibles, se han procesado sólo 250 completas debido a la carga de trabajo que supone cada una de ellas. De esas 250 sesiones, finalmente se han tenido en cuenta 106 de ellas para obtener un balance de sesiones en todas las regiones y contribuir a la mejora del clasificador. Para un mayor detalle sobre la selección de las sesiones, véase el Anexo C.

3.5.2 Selección y filtro de entradas

Como ya se ha explicado en la primera fase del sistema de reconocimiento de emociones, las entradas del sistema son las variaciones, respecto a la cara neutra, de un conjunto de distancias y ángulos entre los 38 puntos faciales obtenidos. Con la idea de determinar qué entradas o atributos son las más influyentes para la emoción a predecir, se ha llevado a cabo un estudio de la matriz de correlación y diagramas de dispersión entre las distancias obtenidas de la BDE utilizada.

Finalmente de las 19 distancias iniciales (ver Fig. 11), se han suprimido 7 de ellas quedando la entrada del clasificador con 12 distancias, por ser éstas las más influyentes y las que mejores resultados de clasificación obtenían (ver Anexo D). En la Fig. 15 pueden observarse a la izquierda todas las distancias iniciales, y a la derecha las distancias finales.

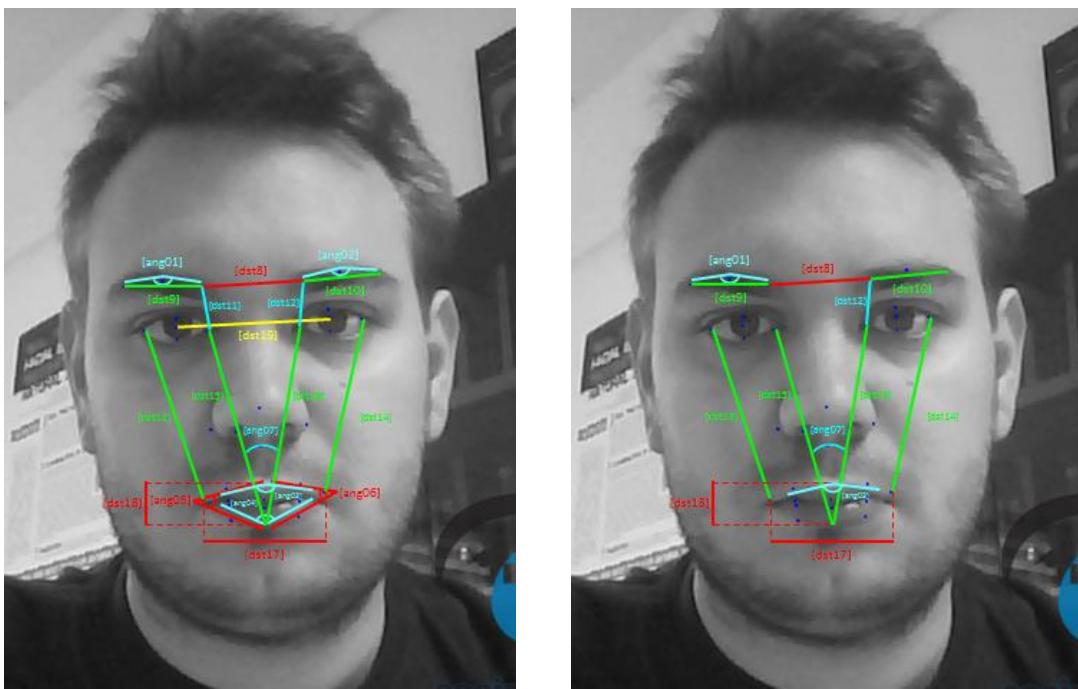


FIG. 15. DISTANCIAS FACIALES SELECCIONADAS

3.5.3 Selección de salidas

Una vez definidas las entradas del clasificador, deben definirse cuáles serán las salidas.

De las 4 dimensiones disponibles en la *MAHNOB-HCI Tagging Database*: activación, evaluación, dominancia y predictibilidad, se han elegido las dos primeras para fijar el nivel de descripción del sistema, por compatibilidad con el método de fusión multimodal implementado previamente [HBC11], y por estar basado en los modelos continuos más relevantes en la literatura [Whi89][Rus80].

Cada una de las sesiones, está etiquetada por el sujeto con un valor numérico de 1 a 9 en las dos dimensiones de interés (activación y evaluación), de manera que en una situación ideal la salida debería pertenecer a un espacio 2D formado por los distintos valores de activación y evaluación (9x9), 81regiones en total, que son consideradas como salidas. Sin embargo, cuando se entrena un sistema de clasificación y se quiere obtener un correcto funcionamiento, debe responderse siempre a la pregunta de si hay suficientes instancias (datos) de cada clase para que el clasificador pueda aprender correctamente. En el clasificador continuo que se trata de desarrollar, dicha pregunta se traduce a comprobar si existe un número suficiente de datos para cada una de las coordenadas del espacio evaluación-activación. Desafortunadamente, las 511 sesiones de las que se dispone no están repartidas homogéneamente en el espacio continuo, de modo que se tienen algunas regiones con muy pocos datos, estando algunas completamente vacías y sin datos para entrenar, como puede observarse en la Fig. 16.

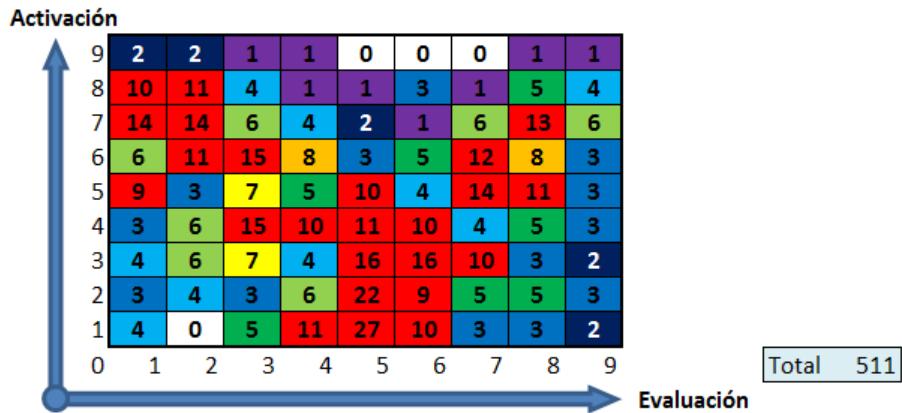


FIG. 16. DISTRIBUCIÓN DE LOS SUJETOS ETIQUETADOS EN EL ESPACIO ACTIVACIÓN-EVALUACIÓN

Por tanto, debemos reducir esas 81 clases a un número más pequeño en el que se dispongan de datos de entrenamiento para el clasificador. Se propone a continuación un modelo que consta de 9 clases diferentes, quedando 3 valores y de activación y 3 valores de evaluación, resultado de agrupar en 9 zonas las 81 anteriores, como muestra la Fig. 17.

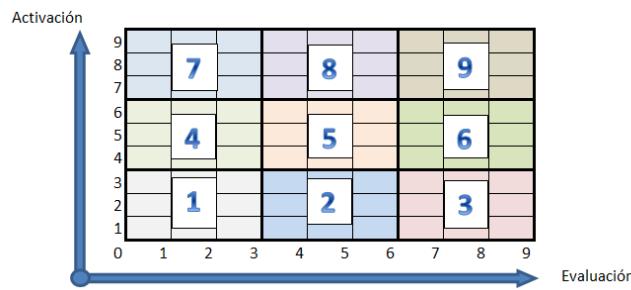


FIG. 17. MODELO 9R DIVIDIDO EN 9 REGIONES

Esta salida del clasificador (9 clases, una por región) es una versión simplificada del modelo ideal que contendría 81 clases. Sin embargo, esta discretización es necesaria por un doble motivo: en primer lugar, se incrementa el número de instancias disponibles para cada una de las 9 clases; y en segundo lugar, minimizamos el porcentaje de error (porcentaje de instancias mal clasificadas). Obviamente, el número de regiones en las que se divide el espacio está relacionado con el porcentaje de error que se obtendrá: cuantas más clases haya, mayor será el error obtenido. Por lo tanto, se trata de conseguir un compromiso entre un espacio suficientemente dividido y el error que obtendrá el clasificador.

Adicionalmente, se ha decidido omitir una de las 9 clases (concretamente la región 3), resultando el modelo final en 8 clases. Esta decisión ha sido fundada en los trabajos de Whissell [Whi89], que sostiene que la región 3 del espacio continuo 2D activación-evaluación apenas contiene emociones, como puede verse en la Fig. 4 del Capítulo 2. Las regiones que finalmente compondrán la salida del clasificador pueden verse en la Fig. 18.

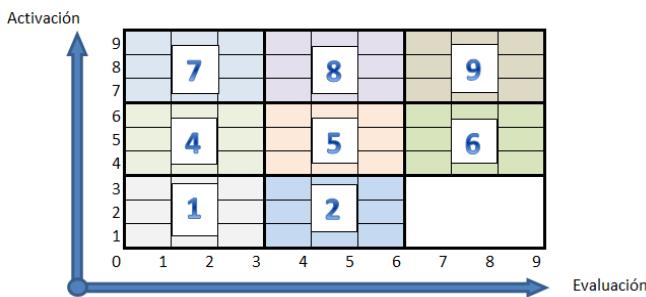


FIG. 18. SALIDAS FINALES DEL CLASIFICADOR

No obstante, y para facilitar la numeración de las regiones, se mantendrán los números omitiendo simplemente la tercera región.

3.5.4 Selección del clasificador

El objetivo de esta etapa es seleccionar un clasificador de entre los disponibles en el software de clasificación para integrarlo en el sistema final. Se proponen inicialmente algunos clasificadores y se realizan un conjunto de pruebas con ellos (llamado *benchmarking*), en las que cuales se selecciona la mejor configuración para el sistema. La herramienta empleada para realizar las pruebas de clasificación es el software WEKA (*Waikato Environment for Knowledge Analysis*) de la universidad de Waikato, Nueva Zelanda [WEKA].

Clasificadores disponibles

Se va a dividir el *benchmarking* en función de los clasificadores disponibles. La herramienta WEKA proporciona un conjunto de algoritmos de aprendizaje automático para tareas de clasificación. De esta colección, se han tenido en cuenta 5 clasificadores: RIPPER, Multilayer Perceptron (MLP), Naive Bayes, C4.5 y Clasificador Sensible al Coste (CSC). Esta selección se basa en el uso generalizado de estos clasificadores en el reconocimiento de emociones y en el trabajo previo desarrollado en el grupo.

- **RIPPER:** *Repeated Incremental Pruning to Produce Error Reduction*. Es un algoritmo basado en reglas. Se tiene en cuenta este clasificador por su simplicidad y su buen rendimiento.
- **MLP:** *Multilayer Perceptron*. Es la red neuronal más usada para tareas de clasificación. Su utilidad radica en la similitud con ciertas redes neuronales biológicas del cerebro humano, lo cual resulta muy útil para este dominio de trabajo.
- **Naive Bayes:** es un clasificador probabilístico basado en el teorema de Bayes y algunas hipótesis simplificadoras adicionales. Es a causa de estas simplificaciones, que se suelen resumir en la hipótesis de independencia entre las variables predictoras, que recibe el apelativo de ingenuo (*naive*). Sin embargo, puede entrenarse muy eficientemente en un ambiente de aprendizaje supervisado.

- C4.5: es también un clasificador basado en reglas, pero se usa para generar árboles de decisión. Su buen rendimiento se debe al uso del concepto de entropía en el ámbito de la información para establecer mecanismos de clasificación.
- CSC-MLP: Clasificador Sensible al Coste con MLP. Este método de clasificación tiene en cuenta una matriz de coste asociada dada por el usuario y un clasificador. En dicha matriz el usuario establece los pesos de cada clase para el clasificador internamente ponderar las instancias de entrenamiento. Se usará el *Multilayer Perceptron* como clasificador asociado.

Benchmarking

Para realizar las pruebas de clasificación se entrena con el fichero que contiene los datos de las sesiones procesadas y varias configuraciones de distancias. Se probarán diferentes parámetros para cada clasificador en función de sus características.

Uno de los problemas del *benchmarking* radica en la forma de evaluar los resultados de clasificación. Si bien es cierto que la bondad de un clasificador se mide por su tasa de acierto, el método estadístico que se suele emplear para la valoración de resultados es de tipo *holdout*: reserva un conjunto de datos para el entrenamiento del método y otro para su validación. Sin embargo, hoy en día existen otras formas mucho más sofisticadas, eficientes y realistas de evaluar las tasas de acierto de los clasificadores. De modo que los resultados han sido calculados mediante un test de validación cruzada con 10 particiones (*10-fold cross-validation test*) sobre los datos de entrada [WCV]. Este tipo de pruebas miden la capacidad del clasificador para auto-adaptarse a nuevos datos, diferentes de los usados para su entrenamiento. La validación cruzada es una práctica estadística consistente en partir la muestra total de datos disponibles en k subconjuntos, de tal modo que uno de ellos se reserva para la validación del esquema de aprendizaje, mientras que el resto de subconjuntos se usan para el entrenamiento del clasificador (el proceso es repetido k veces). De este modo, se obtienen unas medidas mucho más realistas del comportamiento del clasificador. La Tabla 4 refleja los resultados finales de *benchmarking* distinguiendo entre aquellos obtenidos con *10-fold cross-validation test* y sin *cross-validation*.

TABLA 4. RESULTADOS GENERALES DEL BENCHMARKING

CLASIFICADORES	Método de validación	
	10-fold cross validation	Sin cross-validation
RIPPER	26%	46%
NAIVE BAYES	29%	47%
C4.5	23%	83%
MLP	30%	72%
CSC-MLP	33%	79%

Como puede observarse, las tasas de acierto en la tabla son mucho mayores para las pruebas realizadas sin *cross-validation*. No obstante, se toman como válidas las tasas obtenidas mediante la técnica de *10-fold cross-validation* en tanto que ofrecen una visión mucho más realista del comportamiento del clasificador. El Anexo D recoge todo el estudio de Benchmarking en profundidad.

A la vista de los resultados, se observa que están todos más o menos entre el 20% y el 40%, si bien es verdad que tanto MLP como CSC-MLP se destacan sobre los demás. En el apartado siguiente se discuten las decisiones tomadas como resultado de toda la fase de entrenamiento.

Resultados de entrenamiento

Vistos los resultados de los 5 clasificadores, se ha decidido hacer una primera criba y quedarse con los clasificadores que utilizan redes neuronales (MLP y CSC-MLP). La selección no está solamente justificada a que obtengan mejores resultados que los demás, ya que las diferencias no son muy grandes teniendo en cuenta los datos empleados. La justificación reside también en el hecho de que las redes neuronales tienen una excelente capacidad para extraer información de datos complicados o imprecisos y pueden obtener y detectar patrones que son pasados por alto por humanos u otras técnicas de clasificación. Además, cabe mencionar sus útiles cualidades tales como:

1. Alta adaptabilidad a conjuntos de datos que van cambiando.
2. No realizan ninguna suposición con respecto a funciones de densidad u otra información estadística subyacente en comparación a otros modelos probabilistas [SJC96].
3. Se usan comúnmente en la literatura para reconocimiento de emociones.

Una vez decidido que el clasificador debe utilizar una red neuronal, la decisión es seleccionar uno de los dos clasificadores que hacen uso de ella: MLP vs CSC-MLP. Finalmente, se ha optado por elegir CSC-MLP (Clasificador Sensible al Coste con *MultiLayer Perceptron*). La decisión de seleccionar CSC-MLP frente a MLP se debe a que permite condicionar la salida del clasificador en función de las necesidades deseadas a través de una matriz de coste. Esto es especialmente útil si se quiere indicar al clasificador que no es grave confundir las regiones 6 ó 9 (ambas situadas en el cuadrante superior derecho del espacio dimensional), pero que es un problema grave que clasifique un dato en la región 1 cuando debería estar en la 9 (zonas opuestas en el espacio dimensional).

Por lo tanto, pese que un porcentaje del 33% puede parecer bajo, si se observa en la Fig. 19 la matriz de confusión del clasificador seleccionado (matriz cuadrada de tamaño 8 qué indica cómo se han clasificado las instancias) se comprueba que muchas instancias mal clasificadas que contribuyen a disminuir el porcentaje de acierto, están en realidad clasificando en una zona adyacente. De modo que, aunque sólo haya clasificado correctamente el 33% de las instancias, la mayoría de los fallos no son graves (extremos opuestos) y podrían ser tomados como aciertos para algunas emociones que ocupan un amplio rango en el espacio dimensional, como alegría o sorpresa.

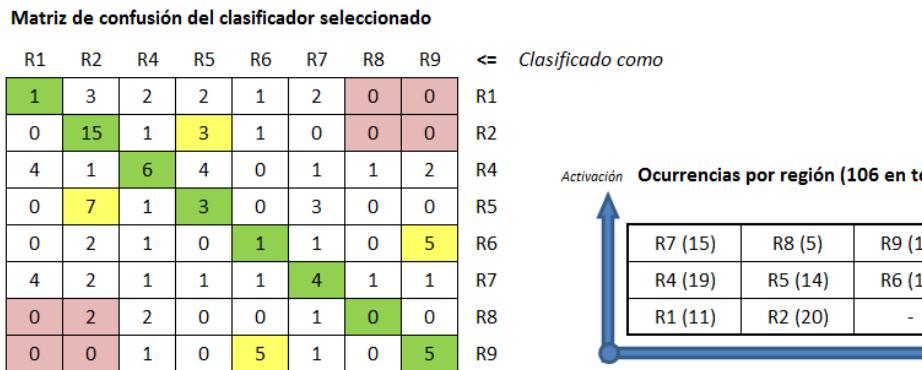


FIG. 19. MATRIZ DE CONFUSIÓN DEL CLASIFICADOR SELECCIONADO: CSC-MLP

La matriz de confusión es una matriz cuadrada de 8 filas y 8 columnas, debido a las 8 clases de salida. Cada fila indica en qué clases se han clasificado los datos procedentes de la región asociada a esa fila. Por ejemplo, la fila 2 (con índice R2 a la derecha de la matriz) indica en dónde se han clasificado los 20 datos asociados a la Región 2: 15 de ellos se han clasificado como R2 (por tanto, clasificados correctamente), 3 de ellos se han clasificado como R5, y los dos restantes se han clasificado en las regiones 2 y 4 (uno en cada una).

De este modo, los números en la diagonal principal, marcados en verde, indican las instancias clasificadas correctamente (un tercio del total). Si se analizan las zonas marcadas en rojo, se puede comprobar que no se ha clasificado prácticamente ninguna instancia allí. Eso significa que el clasificador no confunde las clases 9 y 8 con las clases 1 y 2 (muy importante en la clasificación en continuo). Por otra parte, los números marcados en amarillo indican aquellas instancias clasificadas incorrectamente, pero que son errores aceptables. Por ejemplo, el clasificador confunde las clases 6 y 9: clasifica 5 instancias que deberían ser R6 como R9 y viceversa. Dado que dichas regiones se encuentran en las zonas de alta activación y evaluación, sus emociones asociadas (alegría y sorpresa), a veces se pueden confundir como muestra la Fig. 20.



FIG. 20. EJEMPLO DE EXPRESIONES PARECIDAS (ALEGRÍA - SORPRESA) [PRO0].

3.5.5 Extracción de resultados

Seleccionados el clasificador, y las entradas y salidas de éste, el siguiente paso es encapsular el sistema de clasificación seleccionado en un archivo que pueda alimentar a la fase 3 de clasificación. Para ello, se han utilizado las funciones de exportación de WEKA, que permiten guardar el modelo y la matriz de confusión entrenados, para usarlos posteriormente. Concretamente, el clasificador se ha encapsulado junto con su matriz de confusión en un archivo “.emo”, mediante una pequeña aplicación desarrollada conjuntamente con el programa principal de clasificación de la fase 3, que se ve en la sección siguiente.

3.6 Clasificación

Una vez tratados los apartados de selección de BDE (punto 3.4) y de entrenamiento de clasificador continuo (punto 3.5), y detalladas las fases de extracción de datos (fase 1, punto 3.2) y la fase de conversión de datos (fase 2, punto 3.3) se retoma en este punto la fase tercera del sistema de reconocimiento de emociones faciales: la clasificación. El objetivo de esta fase es clasificar los ángulos y distancias obtenidos en la segunda fase a un nivel de descripción emocional continuo. De este modo, para cada frame se obtiene una salida que indica el estado emocional etiquetado en una de las 8 regiones del espacio dimensional. El proceso de esta fase se puede observar en la Fig. 21.

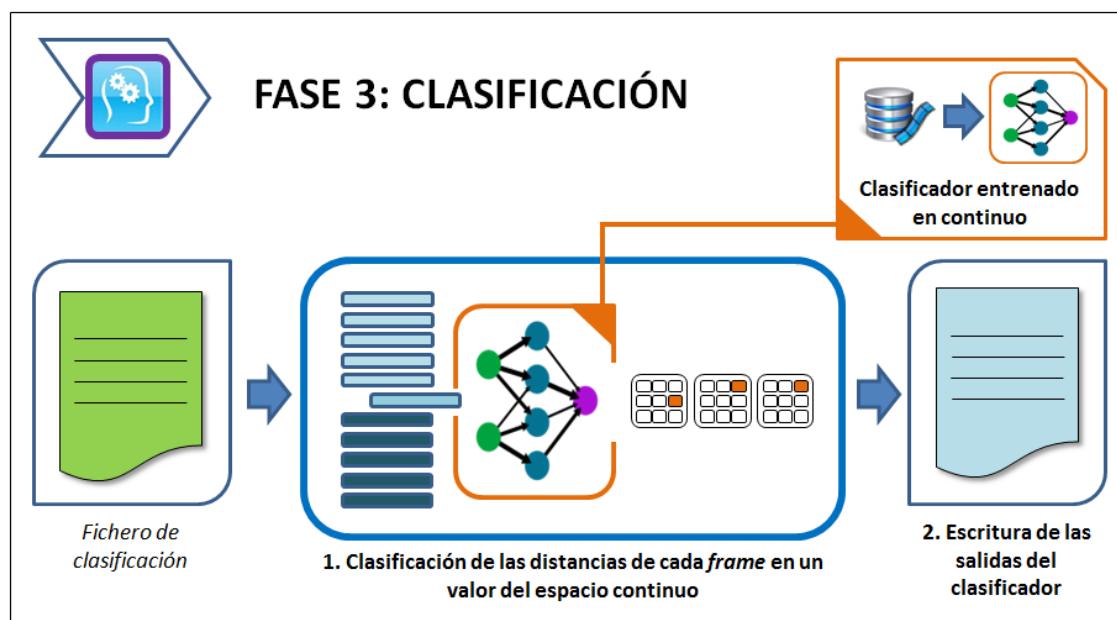


FIG. 21. ESQUEMA GENERAL DE LA FASE 3: CLASIFICACIÓN

3.6.1 Clasificación en el espacio continuo

Para llevar a cabo el objetivo de esta fase, el programa toma el fichero ARFF de clasificación proporcionado por la fase 2 y a partir de cada una de las instancias del fichero obtiene una de las ocho etiquetas emocionales del espacio dimensional activación-evaluación.

El programa hace uso de las librerías WEKA para el IDE Eclipse, de modo que el clasificador previamente entrenado y presentado en el punto 3.5.4 de esta memoria, puede encapsularse en un archivo y utilizarse en esta fase para realizar la clasificación.

3.6.2 Escritura de las salidas del clasificador

Una vez se clasifican todos los *frames* del vídeo a analizar, se debe escribir en un fichero la salida obtenida con una de las nueve regiones del espacio continuo final. Para ello, se almacena para cada *frame*: (1) el código de tiempo del vídeo; (2) la región emocional asociada a dicho frame.

3.7 Visualización

Finalmente, la fase 4 del sistema de reconocimiento facial de emociones corresponde a la fase de visualización. El objetivo de esta última fase es permitir la presentación de los resultados del reconocedor de una manera clara y en la que sea posible contrastar el vídeo analizado con el resultado. El esquema del proceso de esta fase se muestra en la Fig. 22.

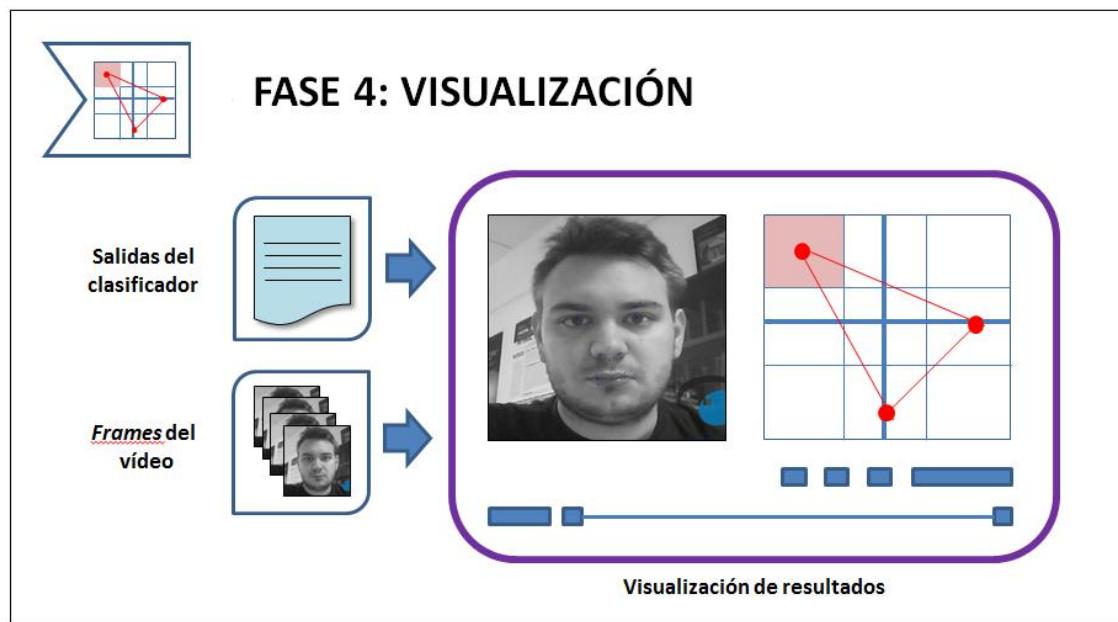


FIG. 22. ESQUEMA GENERAL DE LA FASE 4: VISUALIZACIÓN

El programa de visualización se ha desarrollado en C++ con librerías gráficas Qt y el IDE *QtCreator*, que facilitan la creación de formularios, botones y ventanas de diálogo con el uso del ratón. El programa utiliza como entrada el fichero de resultados de clasificación obtenido en la fase previa, además de una carpeta que contiene los *frames* del vídeo a visualizar.

Para cada frame, se obtiene su estampilla temporal y su etiqueta en el espacio 2D de activación-evaluación, de modo que el usuario puede visualizar para cada fotograma del vídeo la emoción clasificada en el espacio de activación-evaluación. Para más detalle sobre el visualizador implementado, consultar Anexo E: Proceso de desarrollo del software. El aspecto de la interfaz se puede ver en la Fig. 23.

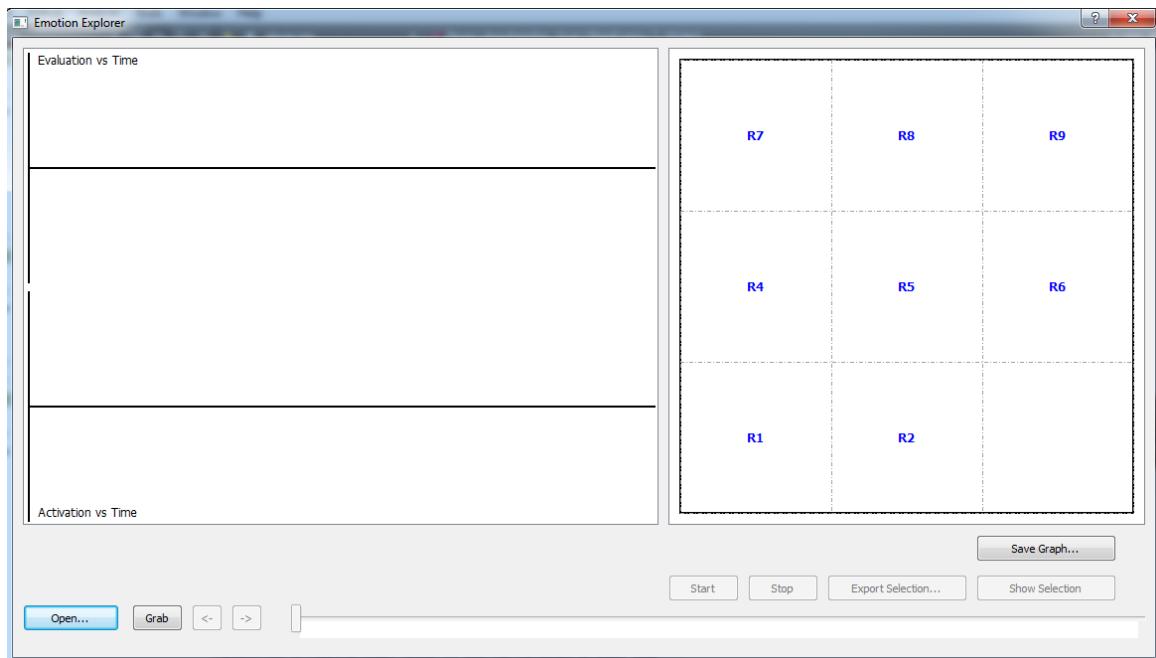


FIG. 23. INTERFAZ DEL VISUALIZADOR DEL RECONOCEDOR DE EMOCIONES FACIALES

3.8 Validación con vídeos

Una vez desarrollado por completo en el sistema de reconocimiento de emociones faciales en un nivel de descripción continuo, queda pendiente la validación del mismo. En esta sección se aplica, siguiendo las fases explicadas previamente, el sistema desarrollado a diferentes vídeos procedentes de evaluaciones con usuarios para comprobar el correcto funcionamiento del sistema y así visualizar los resultados. En primer lugar, se aplicará el sistema a vídeos procedentes de la BDE *MAHNOB-HCI Tagging Database*. En segundo lugar, a los vídeos de entrada procedentes de sesiones grabadas en el laboratorio.

3.8.1 Reconocimiento de emociones en secuencias de vídeo de la BDE utilizada

En primer lugar se ha probado el funcionamiento del sistema sobre algunos de los vídeos de la BDE *MAHNOB-HCI Tagging Database*. De este modo, se puede comprobar si el estado emocional que muestra el sistema como salida se corresponde a la etiqueta emocional con la que el sujeto había etiquetado el vídeo.

El proceso a seguir es el mismo que el proceso presentado en el capítulo 4.

1. Se extraen los datos del vídeo de la BDE.
2. Se convierten los datos a las 12 distancias seleccionadas.
3. Se clasifican dichos datos con el clasificador entrenado en continuo.
4. Se ven los datos en el visualizador.

Por último, se realiza una comparación entre el etiquetado de la sesión y el resultado mostrado por el visualizador. La Fig. 24 muestra la interfaz del visualizador del sistema cuya entrada es un ejemplo de vídeo de la BDE etiquetado con máxima activación y mínima evaluación.

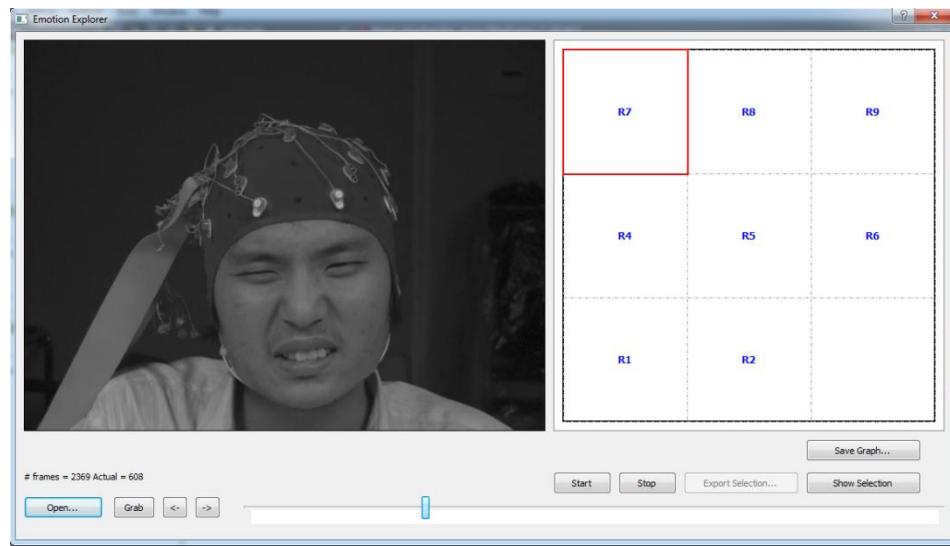


FIG. 24. INTERFAZ DEL VISUALIZADOR MOSTRANDO EL APPEX DE UN VIDEO DE LA **MAHNOB-HCI TAGGING DATABASE**
ETIQUETADO EN UN NIVEL MÁXIMO DE ACTIVACIÓN Y MÍNIMO DE EVALUACIÓN

Como se puede ver en la Fig. 24, la interfaz muestra a la izquierda el frame que contiene el *appex* (pico de la emoción), y la derecha el espacio emocional continuo de 8 regiones considerado como nivel de descripción. El *appex* aparece etiquetado en la región 7 (máxima activación, mínima evaluación), de modo que concuerda lo etiquetado por el sujeto de evaluación y la salida del sistema de reconocimiento automático facial de emociones. Es más, si se observa detenidamente al sujeto, se verá que muestra una cara de aversión o asco, correspondiente a dicha región del espacio continuo.

3.8.2 Reconocimiento de emociones en secuencias de vídeo grabadas en el laboratorio

En segundo lugar se ha probado el funcionamiento del sistema sobre vídeos grabados en el laboratorio. Esto es, se han utilizado sujetos de prueba a los que se ha grabado las reacciones emocionales tras haberles elicitado emociones mediante vídeos. Se comprueba que la etiqueta emocional que muestra el sistema de salida es acorde al tipo de emoción que hace sentir el vídeo (e.g un vídeo alegre hace sentir al sujeto una etiqueta con alta evaluación).

El proceso a seguir es idéntico al proceso seguido en el anterior punto, con la salvedad de que el vídeo ahora ha sido grabado en el laboratorio y el etiquetado ha sido realizado por el sujeto grabado.

La Fig. 25 muestra el visualizador del sistema cuya entrada es un vídeo grabado en el laboratorio con un sujeto de evaluación que ha sido elicitado a sentir emociones de alegría (alta activación, alta evaluación).

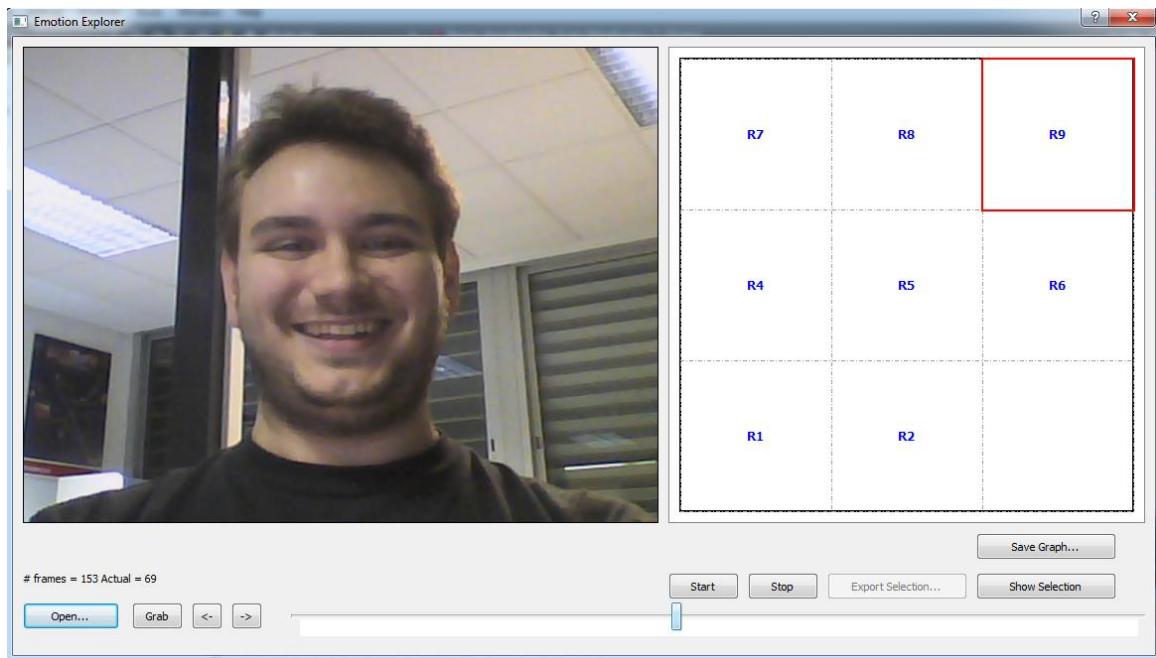


FIG. 25. INTERFAZ DEL VISUALIZADOR MOSTRANDO EL APPEX DE UN VÍDEO DE UN SUJETO ELICITADO CON VÍDEOS QUE PROVOCAN ALEGRÍA

Como se puede observar en la Fig. 25, la salida del sistema para el *frame* que contiene el pico de la emoción es la región 9 (zona de más alta activación y evaluación). En este caso, no existen datos para comprobar si la salida es correcta o no (ya que son vídeos espontáneos no etiquetados grabados en el laboratorio). Sin embargo, la cara del sujeto de evaluación muestra una emoción de alegría que se corresponde con la región 9 del espacio dimensional continuo.

CAPÍTULO 4. APLICACIÓN A LA EVALUACIÓN EMOCIONAL DE VIDEOJUEGOS: *NEUROGAME*

Este capítulo expone una aplicación práctica del sistema desarrollado. En la primera sección del capítulo se describen los objetivos de la colaboración “*Neurogame: Neuro-evaluación multimodal de videojuegos*”, en el que el autor de esta memoria ha participado activamente. A continuación se explica brevemente la metodología propuesta para realizar las evaluaciones, y en la última sección se muestran los primeros resultados obtenidos. Finalmente, el Anexo F contiene la publicación enviada y aceptada en el XIII Congreso Internacional de Interacción Persona-Ordenador, del que el autor de esta memoria es primer firmante.

4.1 Evaluación emocional de videojuegos

Una de las muchas aplicaciones de la detección de emociones reside en la evaluación de la Experiencia de Usuario (UX) en los videojuegos. En dicho contexto, la evaluación hedónica (¿qué sentimos al jugar al videojuego?) completa las informaciones clásicas más orientadas a una evaluación pragmática. Este enfoque permite obtener una visión más amplia dirigida a la mejora de la aceptación de videojuegos, experiencia interactiva y retorno económico en la fase de desarrollo del producto [GGG11]. Además, los estados emocionales influencian procesos cognitivos como la capacidad de atención y memorización, el rendimiento del usuario y, de esta forma, su valoración del producto [BN02][Nor02].

El objetivo del proyecto *Neurogame* de evaluación de videojuegos es profundizar en la comprensión de la medición y el análisis de información asociada a la evaluación hedónica, en forma de respuestas emocionales y cognitivas, para obtener una perspectiva más detallada y precisa de la Experiencia de Usuario. En dicho proyecto participa el *Affective Lab* de la Universidad de Zaragoza con el trabajo desarrollado en este PFC, el grupo GRIHO [WGRI] de la *Universitat de Lleida*, la empresa *BitBrain Technologies* [WBTT] y el Departamento de Psicología y Sociología de la Universidad de Zaragoza [WDPS].

Para realizar la detección emocional y cognitiva, se usarán todo tipo de sensores durante la evaluación que se encargarán de cubrir los diferentes canales mediante los cuales se puede adquirir datos. Dentro de este equipo multidisciplinar la misión del grupo de investigación *GIGA Affective Lab* es la detección automática de emociones en el canal facial, como muestra la Fig. 26.

La participación del autor de este proyecto en *Neurogame* ha sido doble: por un lado desarrollando el sistema de reconocimiento facial continuo necesario para el análisis facial de las evaluaciones y, por otro, participando en la definición de la metodología a seguir en la evaluación. Además, ha sido el responsable de generar la publicación de investigación asociada a dicho proyecto.

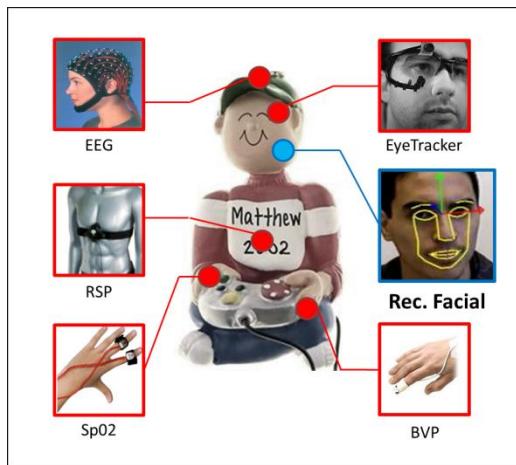


FIG. 26. DIFERENTES CANALES EMOCIONALES Y COGNITIVOS DEL PROYECTO *NEUROGAME*.

A continuación se explica brevemente dicha metodología (en el Anexo F se ha incluido la publicación que la detalla).

4.2 Metodología propuesta

En este apartado se propone una metodología general de evaluación multimodal y emocional de videojuegos, definiendo el objeto y los sujetos de estudio, el nivel de descripción de salida emocional y el proceso de evaluación a seguir.

Objeto y sujetos de estudio

Se seleccionarán uno o varios videojuegos, elegidos por sus similitudes en cuenta a género, historia y protagonistas. En cuanto a los sujetos de estudio, el tamaño muestral del experimento será de unos 20 sujetos balanceando el número de hombres/mujeres en la medida de lo posible.

Nivel de descripción emocional

El proyecto *Neurogame*, pese a involucrar diversos canales de adquisición de datos (fisiológicos y faciales), usa el mismo proceso que los sistemas de reconocimiento de afecto descritos en el capítulo 2. Es por tanto necesario decidir un nivel de descripción emocional que defina en qué términos será descrita la información obtenida, esto es, cómo van a ser representadas las salidas del sistema de reconocimiento de emociones. Se ha acordado utilizar el nivel de descripción continuo en el espacio dimensional, el mismo utilizado para el desarrollo de reconocimiento de emociones faciales del presente trabajo.

Proceso de evaluación

La evaluación está estructurada en una o más sesiones. Todas las sesiones constan, además de una primera fase de configuración inicial, de dos partes diferenciadas: calibración y evaluación (ver Fig. 27). En ambas se presentan varios estímulos al usuario en una o varias modalidades (imágenes, videos, melodías...). El objetivo de la calibración es estimular a la persona en las dos condiciones extremas de las dimensiones emocionales utilizadas. De este modo, se construye un

conjunto de referencias necesarias para evaluar siguientes respuestas emocionales del usuario. Por otro lado, la parte de evaluación consiste en la presentación de estímulos a los sujetos de evaluación para obtener mediciones de sus estados cognitivos y emocionales. Durante la evaluación, se considera una o más categorías del juego (e.g. personajes o escenario ambiental). Para cada una de estas categorías se presentan una serie de muestras, que consisten en un descanso, un estímulo presentado y una pequeña encuesta en la que el sujeto de evaluación se autoevalúa emocionalmente mediante una escala SAM [BL94] (ver Fig. 28).

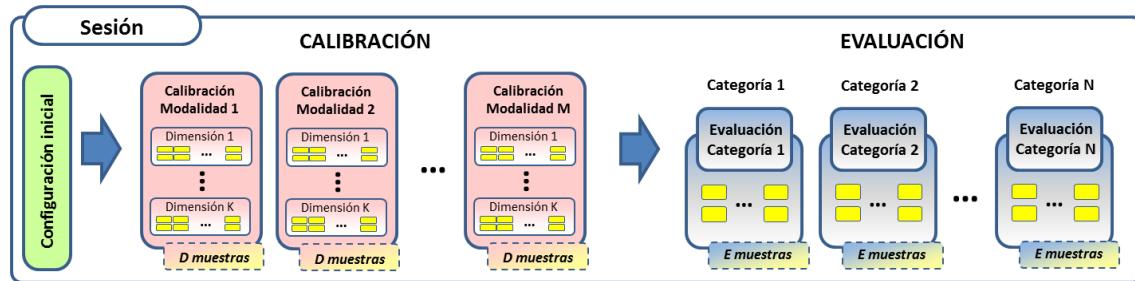


FIG. 27. ESQUEMA GENERAL DE UNA SESIÓN

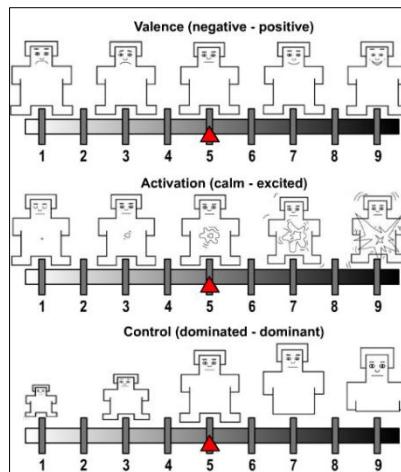


FIG. 28. ESCALA SAM (SELF ASSESSMENT MANIKIN) UTILIZADA PARA LA AUTOEVALUACIÓN EMOCIONAL

4.3 Primeras pruebas

Se ha realizado una primera prueba experimental con 3 sujetos, cuyos datos obtenidos están siendo analizados. Gracias a ellos, se ha podido comprobar la adecuación de los algoritmos de clasificación con datos fisiológicos, y al mismo tiempo realizar pruebas con el nuevo sistema desarrollado de detección de emociones faciales en continuo.

Dispositivos de captura

Para una correcta toma de datos faciales en las sesiones de evaluación, es necesario preparar y comprobar que los dispositivos de captura graban vídeos válidos que sirven como entrada al reconocedor facial. De este modo, las videocámaras necesitan para su correcta recogida de datos unas determinadas condiciones de iluminación, posición y frecuencia. Por ello, se han realizado ensayos y pruebas variando entre diferentes posiciones y niveles de iluminación, hasta

conseguir una configuración adecuada. En el experimento se ha utilizado una videocámara frontal especializada en tecnología para visión artificial y aplicaciones de visión por computador, que permite realizar grabaciones a 60 FPS. Gracias a ello, se ha podido realizar un correcto seguimiento facial que posibilita la captura de cualquier expresión facial, por muy rápida que sea.

Análisis facial

Los primeros análisis están siendo satisfactorios (ver Fig. 29). Se ha podido comprobar la captura satisfactoria por los diferentes sensores de los diversos estados emocionales inducidos en los usuarios dispuestos durante la evaluación. Cabe destacar es la reafirmación de la importancia de la utilización de sensores complementarios (tanto fisiológicos como no fisiológicos), que suplan los fallos en la clasificación emocional de una modalidad. Así, los sensores fisiológicos son capaces de recoger/detectar información emocional incluso cuando el usuario no la está mostrando explícitamente mediante sus expresiones faciales. Y los faciales pueden ayudar a discernir entre estados emocionales que den lugar a respuestas fisiológicas similares.

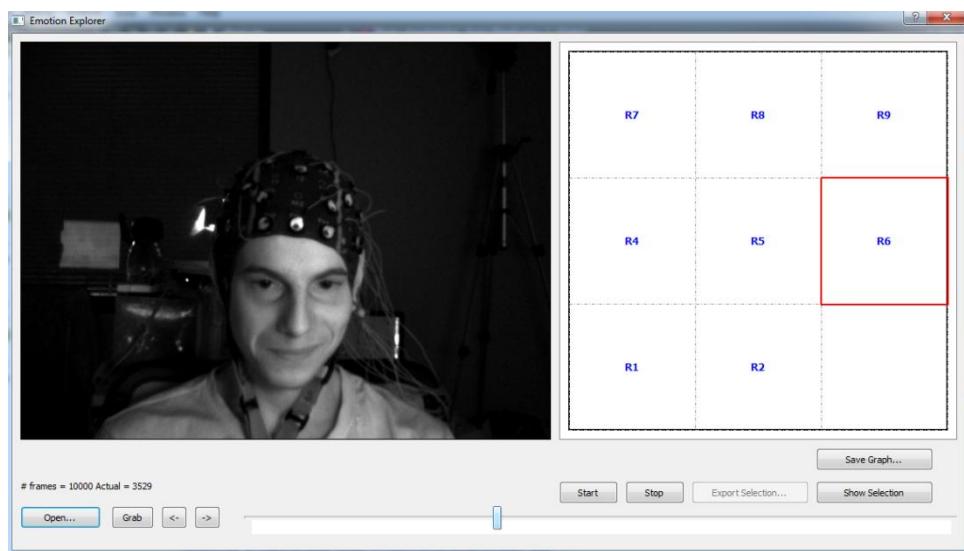


FIG. 29. PRIMEROS RESULTADOS DEL ANALIZADOR FACIAL. USUARIO CONTENTO DURANTE LA EVALUACIÓN EMOCIONAL

De los 3 sujetos con los que se realizó el experimento, solamente ha sido posible realizar análisis de manera satisfactoria con el tercero de ellos. Esto es debido a que en las dos primeras evaluaciones los sujetos llevaban equipado un *Eye-Trakcer* en forma de gafas con cámaras. Esta tecnología era invasiva en la cara del sujeto en tanto que cambiaba el contorno de su rostro, además de lanzar reflejos a la cámara que grababa. Por lo tanto, solo se han podido analizar resultados en aquel sujeto que no utilizó la tecnología *Eye-Tracker*.

Esta primera prueba experimental ha permitido validar la viabilidad de la metodología propuesta. Queda pendiente llevar a cabo una validación completa con varios sujetos de evaluación.

CAPÍTULO 5. CONCLUSIONES Y TRABAJO FUTURO

En este último capítulo se presentan las conclusiones finales del Proyecto Fin de Carrera, indicando el cumplimiento de los objetivos definidos en el Capítulo 1. Asimismo, se incluye una sección de líneas de trabajo futuro donde se indican posibles mejoras para el sistema de reconocimiento de emociones faciales, así como las nuevas líneas de investigación abiertas en la detección automática de emociones en continuo. Finalmente, en la última sección se expone una valoración personal del trabajo desarrollado.

5.1 Conclusiones

Una vez finalizado el proyecto, es posible afirmar que se han alcanzado todos los objetivos propuestos y que se ha implementado satisfactoriamente un sistema de reconocimiento de emociones en secuencias de vídeo. El sistema desarrollado permite ver la evolución del estado emocional de un sujeto grabado en un vídeo en un nivel de descripción completamente continuo, siendo ésta la mayor aportación del presente proyecto.

La evolución temporal del proyecto puede verse detalladamente en el Anexo A, y los objetivos cumplidos durante la misma son los siguientes:

1. Se ha realizado un estado del arte del reconocimiento automático facial.
2. Se ha realizado un estudio exhaustivo de las Bases de Datos Emocionales disponibles en la literatura y se ha llevado a cabo la selección de la más adecuada para el sistema.
3. Se ha desarrollado un clasificador que toma características faciales como entrada y devuelve una salida en las dos principales dimensiones del espacio emocional continuo.
4. Se ha implementado y puesto a punto el nuevo sistema de reconocimiento de emociones en secuencias de video con el clasificador en continuo.
5. Se ha definido una aplicación práctica del sistema. Gracias a una colaboración existente con el grupo GRIHO de la *Universitat de Lleida* [WGri] y con la empresa *BitBrain Technologies* [WBTT], se han podido realizar unas primeras evaluaciones emocionales de videojuegos, donde el sistema desarrollado ha desempeñado la función de extraer la información emocional del canal facial. Si bien inicialmente se pensaba contar con las primeras evaluaciones para este trabajo, el retraso del proyecto *Neurogame* ha hecho que solamente se haya dispuesto de los datos de las primeras evaluaciones de prueba.

Se concluye por tanto que la migración del enfoque discreto al dimensional se ha realizado satisfactoriamente, permitiendo al grupo de investigación *GI GA Affective Lab* disponer de un reconocedor automático de emociones en un nivel de descripción completamente continuo. Además de participar en la definición de la metodología y puesta a punto, y realización de las primeras pruebas, el autor de este trabajo ha sido el encargado de preparar y enviar un artículo de investigación a un congreso relevante de Interacción presentando *Neurogame*, publicación que ha sido aceptada.

5.2 Líneas de trabajo futuro

El desarrollo de este Proyecto Fin de Carrera, abre nuevas puertas y líneas de investigación con las que seguir trabajando. Concretamente, como trabajo futuro y mejoras posibles cabe citar las siguientes:

1. **Sistema de reconocimiento de emociones en tiempo real.** El sistema desarrollado trabaja *offline*. Esto es, los vídeos se capturan en una primera fase de toma de datos, y el resultado en forma de estado emocional se obtiene tras la aplicación de un proceso de 4 fases en las que se van transformando los datos de entrada. Trabajar en tiempo real es posible dado que el sistema actual no ofrece ningún obstáculo para ello, de modo que una primera mejora del sistema sería convertir su proceso *offline* a uno de tiempo real (por supuesto, ateniéndose siempre a las restricciones del procesador que la tecnología de reconocimiento automático requiere).
2. **Mejora del reconocedor facial.** Una de las limitaciones del sistema desarrollado reside en la precisión del reconocedor facial *FaceAPI*. Si bien es cierto que este software está preparado para trabajar en tiempo real, en algunas ocasiones no toma los puntos faciales correctamente, implicando un reconocimiento erróneo de toda la secuencia de imágenes. Como mejora, se propone la aplicación de otras alternativas mencionadas en el Anexo B, como *Microsoft Kinect* [MKin] u *OpenCV* [OpCV]. De este modo, no se dependería en tanta medida del correcto funcionamiento del reconocedor facial, dado que estas dos alternativas propuestas permiten modificar el código mejorándolo y adecuándolo a las circunstancias en las que se trabaje.
3. **Mejora del clasificador.** Las tasas de acierto obtenidas por el clasificador son un primer paso en el campo de la detección emocional continua, pero sería deseable minimizar los errores y conseguir una colección más completa de datos. Como mejora inmediata, y debido a la adaptabilidad que las redes neuronales dotan al clasificador, éste puede mejorarse aumentando los datos de entrenamiento creando así un sistema de clasificación más fiable, robusto y universal (si bien es cierto que hay que invertir muchas horas de trabajo). Para ello, se pueden usar otras BDE donde conseguir material etiquetado en continuo o descargar las sesiones de la *MAHNOB-HCI Tagging Database* [BDE-01] que los autores siguen etiquetando actualmente.
4. **Aumentar las clases de salida.** La mejora anterior implica un aumento considerable del número de datos, suprimiendo el problema inicial que forzó a la división del espacio de 81 regiones a 9. Así, podría dividirse un poco más el espacio dimensional resultando en una salida de 16 o incluso 25 regiones. De este modo, y poco a poco, se conseguiría un clasificador más preciso que explotase aún más los beneficios del enfoque emocional continuo.
5. **Neurogame.** Una vez llevada a cabo la evaluación con usuarios con la metodología propuesta habrá que llevar a cabo el análisis exhaustivo de dichos datos, tarea que por razones temporales no ha sido posible incluir en el presente trabajo.

5.3 Valoración personal

La experiencia del desarrollo de este proyecto me ha resultado muy positiva y enriquecedora. Gracias a él me he visto inmerso en un trabajo de investigación, y más concretamente, en el apasionante mundo de la Computación Afetiva.

He podido aplicar directamente conocimientos adquiridos en varias asignaturas de la carrera: de carácter más general como asignaturas relacionadas con el proceso de desarrollo de software y proyectos; o más específicos, como conocimientos adquiridos en asignaturas de Inteligencia Artificial o Estadística. Asimismo, he visto como el trabajo realizado y las destrezas adquiridas durante estos años de carrera han dado sus frutos, proporcionándome la adaptabilidad necesaria para desempeñar con eficacia nuevas tareas que hacen uso de tecnologías recientes o herramientas desconocidas.

Por otra parte, la gran carga de investigación de este proyecto ha implicado una primera fase de aprendizaje y estudio del estado del arte en el que había momentos en los que parecía que nunca se avanzaba. No obstante, dicha fase previa ha demostrado ser extremadamente enriquecedora y necesaria para el desarrollo del proyecto, proporcionándome un espíritu crítico con el trabajo de los demás y con el mío propio. Asimismo, la realización de este proyecto me ha brindado la oportunidad de escribir un artículo en un congreso internacional, permitiéndome tomar contacto directo con el mundo de la investigación y ofreciéndome una visión más amplia de las opciones que tiene uno al acabar su carrera universitaria.

En definitiva, personalmente considero que ha sido una experiencia altamente positiva y satisfactoria, permitiendo desarrollar mis aptitudes como ingeniero y completando mi formación.

BIBLIOGRAFÍA Y DOCUMENTACIÓN

- [ARFF] Attribute-Relation File Format. Página de la Universidad de Waikato, <http://www.cs.waikato.ac.nz/ml/weka/arff.html> (último acceso 09/07/2012)
- [Ball11] S. Ballano. *"Evaluación de un Sistema Multimodal de Reconocimiento de Emociones"*. Universidad de Zaragoza, 2011. Tesis Fin de Máster.
- [BL94] Bradley, M.M., Lang, P.J. 1994. "Measuring emotion: The self-assessment manikin and the semantic differential". *Journal of Behavior Therapy and Experimental Psychiatry* 25, 49-59. 1994.
- [BN02] Brave, S., Nass, C. "Emotion in human-computer interaction". The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications (chap. 4) 81-96. J.Jacko & A. Sears (Eds.) Hillsdale, NJ: Lawrence Erlbaum Associates. 2002.
- [CHF06] Y. Hang, C. Hu, R. Feris, M. Turk. "Manifold based analysis of facial expression". *J. Image & Vision Computing*, Vol. 24, nº 6, pp. 605-614, 2006.
- [CSG03] I. Cohen, N. Sebe, A. Garg, L.S. Chen, T.S. Huang. "Facial expression recognition from video sequences: temporal and static modelling". *J. Computer Vision and Image Understanding*, Vol. 11, pp. 160–187. 2003.
- [Dar98] C. Darwin. *The expression of the emotions in man and animals (3rd Edition)*. New York: Oxford University Press.
- [DBW07] Y. Du, W. Bi, T. Wang, Y. Zhang, H. Ai, "Distributing expressional faces in 2-D emotional space" *In Proceedings of the 6th ACM International Conference on Image and Video Retrieval*, 2007, pp. 395-400. 2007.
- [DCSC07] E. Douglas-Cowie, R. Cowie, I. Sneddon, C. Cox, O. Lowry, M. McRorie, J.-C. Martin, L. Devillers, S. Abrilian, A. Batliner, N. Amir, K. Karpouzis. "The HUMAINE database: addressing the collection and annotation of naturalistic and induced emotional data". *In Proceedings of the 2nd International Conference on Affective Computing and Intelligent Interaction (ACII '07), vol. 4738 of Lecture Notes in Computer Science, pp. 488–500, Lisbon, Portugal*. September 2007.
- [EIDE] Eclipse Integrated Development Environment. <http://www.eclipse.org/> (ultimo acceso el 22/0/2012)
- [Ekm99] P. Ekman. *Facial Expression, The Handbook of Cognition and Emotion*. John Wiley et Sons, 1999.
- [ELop] Edouard López. Documento comparativo de funcionalidades entre los software Weka y R. Url: <http://edouard-lopez.com/fac/ICPS%20-%20S7/Data%20Mining/ComparingWekaAndR.pdf> (último acceso el 22/07/2012)
- [FASM] Software FaceAPI, SeeingMachines. Url: <http://www.seeingmachines.com/product/faceapi/> (último acceso 26/06/2012).

- [FT05] N. Fragopanagos, J.G. Taylor. "Emotion recognition in human computer interaction" *Neural Networks*, 18(4), pp. 389-405. 2005.
- [GGG11] González Sánchez, J.L., Gil Iranzo, R.M, Gutiérrez Vela, F. L. 2011. "Enriqueciendo la Evaluación en Videojuegos". *Actas del XII Congreso Internacional de Interacción Persona-Ordenador. Interacción 2011* (Lisboa, Portugal, 2-6 de Septiembre). Néstor Garay y Julio Abascal (Eds.), 273-282.
- [GMK06] G. Garidakis, L. Malatesta, L. Kessous, N. Amir, A. Paouzaio, K. Karpouzis. "Modeling naturalistic affective states via facial and vocal expressions recognition". In *Proceedings of the 8th international conference on Multimodal interfaces (ICMI 2006). Banff, Alberta, Canada*, pp. 146-154. 2006
- [GP06] H. Gunes, M. Piccardi. "A bimodal face and body gesture database for automatic analysis of human nonverbal affective behavior". In *International Conference on Pattern Recognition (ICPR), volume 1, pages 1148–1153*. 2006.
- [GP09] H.Gunes , M.Piccardi. "Automatic temporal segment detection and affect recognition from face and body display". *IEEE Transactions on Systems, Man and Cybernetics – Part B* 39(1) pp 64-84. 2009.
- [GP10] H.Gunes, M.Pantic. "Automatic, dimensional and continuous emotion recognition". *Int Journal of Synthetic Emotions (IJSE)* 1(1), 68-99. 2010.
- [HBC11] I. Hupont, S. Ballano, E. Cerezo, S. Baldassarri. "Scalable multimodal fusion for continuous affect sensing". In *Proceedings of IEEE Workshop on Affective Computational Intelligence - WACI (Paris)* pp 1-8. April 2011.
- [HCC05] Z. Hammal, L. Couvreur, A. Caplier, M. Rombaut, "Facial Expressions Recognition Based on The Belief Theory: Comparison with Diferent Classifiers". In *Proc. 13th International Conference on Image Analysis and Processing (ICIAP), Italy*. 2005.
- [HH97] C.L. Huang, Y.M. Huang, "Facial Expression Recognition Using Model-Based Feature Extraction and Action Parameters Classification", *J. Visual Comm. and Image Representation*, 8(3), pp. 278-290. 1997.
- [Hup10] I. Hupont, 2010. "Affective computing: emotional facial sensing and multimodal fusion". Universidad de Zaragoza, 2010. Tesis Doctoral.
- [KCT00] T. Kanade, J. Cohn, Y. Tian. "Comprehensive database for facial expression analysis". In *Proceedings of International Conference on Face and Gesture Recognition*, pp 46–53. March, 2000.
- [LBA99] M.J. Lyons, J. Budynek, S. Akamatsu, "Automatic Classification of Single Facial Images", *IEEE Trans. Pattern Analysis and Machine Intelligence*, 21(12), pp. 1357-1362. 1999.
- [LCK10] P. Lucey, J.F Cohn, T. Kanade, J. Saragih, Z. Ambadar, I. Matthews. "The extended Cohn-Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression". In *Computer vision and Pattern Recognition Workshop on Human-Communicative Behavior*, 2010.
- [LS11] J. Lichtenauer, M. Soleymani. "MAHNOB-HCI-TAGGING Database Manual". London, 2011.

- [MKin] Microsoft Kinect. Url: <http://www.microsoft.com/en-us/kinectforwindows/> (último acceso 26/06/2012).
- [MPEG4] ISO/IEC 14496 – MPEG-4 International Standard, Moving Picture Experts Group. Url: www.cselt.it/mpeg (último acceso 26/06/2012).
- [MVP10] G. McKeown, M. Valstar, M. Pantic, R. Cowie." The SEMAINE corpus of emotionally coloured character interactions". *In Proc. IEEE Int. Conf. Multimedia & Expo* pp. 1–6. 2010.
- [Nor02] Norman, D. "Emotion and Design: Attractive things work better". *Interactions Magazine*, 9(4), pp 36-42. 2002.
- [OpCV] Librerías OpenCV. Url: <http://opencv.willowgarage.com/wiki/> (último acceso 09/07/2012).
- [Pic97] R. Picard, 1997. *Affective Computing*. MIT Press, Cambridge.
- [Plu84] Plutchik, R. "Emotions: A general psychoevolutionary theory". *In K.Scherer & P. Ekman (Eds.), Approaches to emotion (pp. 197-219)*. Hillsdale, NJ: Lawreacen Erlbaum Associates. 1984
- [PP06] M. Pantic,I. Patras, "Dynamics of facial expression: Recognition of facial actions and their temporal segments from face profile image sequences". *IEEE Trans. On Systems, Man and Cybernetics - Part B*, 36(2), pp. 433-449. 2006.
- [PR00] M. Pantic, L.J.M. Rothkrantz, "Automatic Analysis of Facial Expressions: The State of the Art". *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 22(12), pp. 1424–1445. 2000.
- [PVR05] M. Pantic, M. Valstar, R. Rademaker, L. Maat. "Fully automatic Facial Action Recognition in Spontaneous Behavior". *In Proceedings of the International Conference on Multimedia and Expo*, pages 317–321, 2005.
- [Rmin] Rapid Miner. Url: <http://rapid-i.com/content/view/281/225/lang,en/> (último acceso el 22/07/2012).
- [RPro] Software R. Url: <http://www.r-project.org/> (último acceso el 22/07/2012).
- [Rus80] Russell, J. A. "A circumplex model of affect". *Journal of Personality and Social Psychology*, 39, pp 1161-1178. 1980.
- [SBB02] T. Sim, S. Baker, M. Bsat, "The CMU Pose, Illumination, and Expression (PIE) Database," *In Proc. IEEE Int'l Conf. Automatic Face and Gesture Recognition*. May 2002.
- [SJC96] M. C. Su, W. F. Jean, H. T. Chang., "A Static Hand Gesture Recognition System Using a Composite Neural Network." *Fifth IEEE Int. Conf. on Fuzzy Systems*, pp. 786-792, New Orleans, U.S.A. (NSC85-2213-E-032-009). 1996.
- [SLPP12] M.Soleymani, J.Lichtenauer, T.Pun, M.Pantic "A Multi-Modal Database for Affect recognition and Implicit Tagging," *IEEE Trans. Affective Computing*, 3(1), pp. 42-55. Jan.-Mar, 2012.
- [SSB09] N. Stoiber, R Seguier, G. Breton, "Automatic design of a control interface for a synthetic face" *In Proceedings of the 13th International Conference on Intelligent User Interfaces*, pp. 207-216. 2009
- [SSJ01] Scherer, K. R., Schorr, A. Johnstone, T. *Appraisal processes in emotion: Theory, methods, research*. New York: Oxford University press. 2001.

- [Tom62] Tomkins, S. S. *Affect, imagery, consciousness: Vol 1. The positive affects*. New York: Springer, 1962.
- [VDUB] Software Virtual Dub. Url: <http://www.virtualdub.org/> (ultimo acceso el 25/07/2012).
- [WBTT] Página web de la empresa BBT (BitBrain Technologies). Url: <http://bitbrain.es/> (último acceso 19/07/2012).
- [WCV] Wikipedia: Validación Cruzada. [http://en.wikipedia.org/wiki/Cross-validation_\(statistics\)](http://en.wikipedia.org/wiki/Cross-validation_(statistics)) (último acceso 10/07/2012).
- [WDoc] Documentación de Weka. Url: http://www.cs.waikato.ac.nz/ml/weka/index_documentation.html (último acceso el 22/06/2012).
- [WDPS] Departamento de Psicología y Sociología de universidad de Zaragoza. Url: <http://psico-socio.unizar.es/> (último acceso 22/06/2012)
- [WEKA] Waikato Environment for Knowledge Analysis. Url: <http://www.cs.waikato.ac.nz/ml/weka/index.html> (último acceso 09/07/2012).
- [WGNU] GNU. Página de GNU en Wikipedia. Url: <http://es.wikipedia.org/wiki/GNU> (último acceso 22/06/2012).
- [WGri] Página web del grupo GRIHO (Grupo de Investigación en Interacción Persona Ordenador e Integración de Datos). Url: http://griho.udl.cat/es/inici.html?__locale=es (último acceso 19/07/2012).
- [Whi89] Whissell, C. M. "The dictionary of affect in language". In R. Plutchik & H. Kellerman (Ed.). *Emotion: Theory, research and experience. The measurement of emotions. Vol 4, pp 113-131.* New York: Academic Press. 1989.
- [WXML] Estándar XML del W3C: Url: <http://www.w3.org/XML/> (último acceso 19/07/2012).
- [ZJ05] Y. Zhang , Q. Ji, "Active and dynamic information fusion for facial expression understanding from image sequence". *IEEE Trans. Pattern Analysis & Machine Intelligence*, 27(5), pp. 699-714. 2005.

URLS DE BASES DE DATOS EMOCIONALES

- [BDE-01] Página web de la BDE MAHNOB-HCI Tagging Database. Url: <http://mahnob-db.eu/> (último acceso 22/06/2012).
- [BDE-02] Página web de la BDE Humaine Database. Url: <http://sspnet.eu/2009/12/humaine-database> (último acceso 22/06/2012).
- [BDE-03] Página web de la BDE Belfast Naturalistic Database. Url: <http://belfast-naturalistic-db.sspnet.eu> (último acceso 22/06/2012).
- [BDE-04] Página web de la BDE SAL Database. Url: <http://www.image.ntua.gr/ermis> (último acceso 22/06/2012).
- [BDE-05] Página web de la BDE Oreiestea Database. Url: <http://manolito.image.ece.ntua.gr/oresteia/index.html> (último acceso 22/06/2012).
- [BDE-06] Página web de la BDE XM2VTSDB Database. Url: <http://www.ee.surrey.ac.uk/CVSSP/xm2vtsdb> (último acceso 22/06/2012).
- [BDE-07] Página web de la BDE The AR Face Database. Url: <http://www2.ece.ohio-state.edu/~aleix/ARdatabase.html> (último acceso 22/06/2012).
- [BDE-08] Página web de la BDE CVL Face Database. Url: <http://www.lrv.fri.uni-lj.si/facedb.html> (último acceso 22/06/2012).
- [BDE-09] Página web de la BDE The Psychological Image Collection at Stirling Database. Url: <http://pics.psych.stir.ac.uk> (último acceso 22/06/2012).
- [BDE-10] Página web de la BDE The Japanese Femal Facial Expression (JAFFE) Database. Url: <http://www.kasrl.org/jaffe.html> (último acceso 22/06/2012).
- [BDE-11] Página web de la BDE CMUPIE (Pose, Illumination and Expression) Database. Url: http://www.ri.cmu.edu/research_project_detail.html?project_id=418&menu_id=261 (último acceso 22/06/2012).
- [BDE-12] Página web de la BDE Indian Institute of Technology Kanpur Database. Url: <http://vis-www.cs.umass.edu/~vidit/IndianFaceDatabase> (último acceso 22/06/2012).
- [BDE-13] Página web de la BDE The Yale Face Database. Url: <http://cvc.yale.edu/projects/yalefaces/yalefaces.html> (último acceso 22/06/2012).
- [BDE-14] Página web de la BDE Caltech Frontal Face Database. Url: <http://www.vision.caltech.edu/html-files/archive.html> (último acceso 22/06/2012).
- [BDE-15] Página web de la Human Scan BiOID Face Database. Url: <https://www.bioid.com/download-center/software/bioid-face-database.html> (último acceso 22/06/2012).
- [BDE-16] Página web de la BDE Oulu University Physics-Based Face Database. Url: <http://www.ee.oulu.fi/research/imag/color/pbfd.html> (último acceso 22/06/2012).
- [BDE-17] Página web de la BDE The Bosphorus Database. Url: <http://bosphorus.ee.boun.edu.tr/Home.aspx> (último acceso 22/06/2012).

- [BDE-18] Página web de la BDE The Database of Faces AT&T. Url: <http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html> (último acceso 22/06/2012).
- [BDE-19] Página web de la BDE The Yale Face Database B. Url: <http://cvc.yale.edu/projects/yalefacesB/yalefacesB.html> (último acceso 22/06/2012).
- [BDE-20] Página web de la BDE FABO Database. Url: <http://research.it.uts.edu.au/cvrg/FABO.htm> (último acceso 22/06/2012).
- [BDE-21] Página web de la BDE Semaine Database. Url: <http://semaine-db.eu> (último acceso 22/06/2012).
- [BDE-22] Página web de la BDE MMI Database. Url: <http://www.mmifacedb.com/accounts/register/> (último acceso 22/06/2012).
- [BDE-23] Página web de la BDE MUG Database. Url: <http://mug.ee.auth.gr/fed/> (último acceso 22/06/2012).
- [BDE-24] Página web de la BDE Cohn Kanade Database. Url: http://www.ri.cmu.edu/research_project_detail.html?project_id=421&menu_id=261 (último acceso 22/06/2012).
- [BDE-25] Página web de la BDE Extended Cohn Kanade Database. Url: http://www.pitt.edu/~jeffcohn/CVPR2010_CK+2.pdf (último acceso 22/06/2012).

