Escuela Técnica Superior de Ingenieros

Universidad de Sevilla

Evaluación de la Depresión Mediante el Seguimiento de Características Faciales

Resumen

Margara Tejera Padilla

Tutores:

Aníbal Ollero Baturone

Universidad de Sevilla

Fernando de la Torre Frade

Carnegie Mellon University

Junio 2009

Resumen

Las expresiones faciales ha sido siempre de particular interés para la comunidad psicológica, puesto éstas están muy asociadas con la emoción. Avances en los campos de la visión por computador y el aprendizaje automático y la realización de grandes esfuerzos en la adquisición de datos han hecho que el estudio automático de las expresiones faciales en secuencias de vídeo sea cada vez más posible. Basándose en este contexto, este proyecto aplicará técnicas novedosas como los Active Appearance Models (Modelos Activos de Apariencia) y Support Vectors Machines (Máquinas de Soporte Vectorial), con el objetivo de desarrollar la primera aproximación de un software capaz de evaluar la depresión a partir de la información de forma de la cara de los pacientes. En particular, se presentarán experimentos de clasificación desarrollados sobre la base de datos Spectrum – un conjunto de secuencias de vídeo en los que aparecen sujetos deprimidos y no deprimidos. Además, se discutirán distintos métodos de alineamiento y de extracción de características.

Índice general

Capítulo 1

Introducción

La expresión facial trae consigo una gran cantidad de información: tanto el estado de ánimo como otras estructuras más sutiles, aunque de importante significado. La configuración y el movimiento de las características faciales pueden aportar información valiosa para entender a las personas. Sin embargo, la complicada naturaleza de las expresiones faciales (hay más de 20 músculos dedicados), y por ende de su dinámica, hace imposible estudiarla empleando únicamente herramientas estadísticas.

El aprendizaje automático ofrece una solución a este tipo de problemas en los que la estadística tradicional no es suficiente. Aunque la capacidad de adquisición de datos sigue superando con creces la capacidad para extraer información de ellos, la importancia de este campo de conocimiento ha crecido mucho durante las últimas dos décadas. La actual demanda de conocimiento en minería de datos y de aprendizaje automático por parte de la industria ha resultado en un gran crecimiento de las becas concedidas en este área. A pesar de ser un campo desarrollado muy recientemente y con gran cantidad de problemas abiertos, se han conseguido ya importantes logros: el reconocimiento del habla es considerado ya un problema prácticamente resuelto (hay muchas aplicaciones comerciales que consiguen resultados muy satisfactorios [15]), y se han realizado grandes avances en el campo del reconocimiento facial, en el que se han desarrollado sistemas avanzados con un rendimiento muy prometedor. El análisis automático de la expresión facial es quizás el siguiente paso.

Los primeros estudios publicados sobre el análisis automático de expresiones faciales datan de finales de los setenta, pero no es hasta los noventa que se consiguen resultados prácticos. Los avances en la visión por computador, especialmente en el reconocimiento y seguimiento de caras, y el incremento de la capacidad de cómputo de los ordenadores han hecho posible gran cantidad de estudios; el problema, sin

embargo, está muy lejos de considerarse resuelto [12].

1.1. Motivación

El interés en el estudio de las expresiones faciales ha crecido durante las últimas dos décadas, quizás debido a que los avances en el aprendizaje máquina y la visión por computador han hecho que su estudio automático en secuencias de vídeo sea cada vez más posible.

Estudios recientes han comenzado a explorar seriamente la aplicabilidad del aprendizaje máquina y la visión por computador en el estudio de la actividad facial, y en una amplia gama de aplicaciones: depresión, engaño, dolor, efecto de las drogas, estado de alerta de los conductores, etc. [1, 17, 14, 9].

La depresión es, sin duda, uno de los grandes problemas de la sociedad actual. Según la Organización Mundial de la Salud, se estima que, en un momento dado, el 5-10 % de la población sufre problemas de depresión identificables y necesita tratamiento psiquiátrico o ayuda psicológica [13].

El test de Hamilton es ampliamente usado para la evaluación de la depresión y se combina con el estudio de enfermedades no necesariamente relacionadas con la psicología, como la fibromialgia o el alzheimer [10, 7]. Este test consiste en una entrevista entre un paciente y un terapeuta, siguiendo un cuestionario de 17 ó 25 preguntas. Aunque este cuestionario está normalizado, la intensidad de la depresión del paciente se basa en la puntuación que el terapeuta da a cada una de las preguntas, no siendo pues completamente objetiva.

Conseguir un sistema capaz de evaluar la depresión automáticamente a partir de secuencias de vídeo/audio de los sujetos supondría un gran avance tanto en el campo de la psicología como en el de la visión por computador, y es pues la principal motivación de este proyecto. Por una parte, facilitaría la tarea de los terapeutas, que complementarían su decisión con aquella dada por el método de evaluación automática. Por otra parte, demostraría que estudiar y entender el comportamiento humano automáticamente es viable mediante el empleo de técnicas de visión por computador y de aprendizaje automático. Por último, sentaría precedente para el estudio automático de otras enfermedades, tales como el autismo.

En conclusión, este proyecto es una primera aproximación de un sistema de evaluación automática de la depresión, basado en un procedimiento de clasificación de dos clases: *deprimido* y *no deprimido*. Ya que los experimentos de clasificación se centrarán en partes específicas de la cara, como cejas, ojos y boca, será posible

determinar cuáles de estas zonas son más discriminantes a la hora de evaluar la depresión. Por tanto, puede decirse que otra de las motivaciones de este proyecto conseguir un mejor entendimiento de las expresiones faciales y del comportamiento humano relacionados con la depresión.

1.2. Objetivos

La base de datos Spectrum es la fuente de información de este proyecto. Consiste en una colección de secuencias de vídeo de pacientes depresivos, estando la imagen centrada en la cara de los mismos (para más detalles, ver la Sección 3.2).

Un método de seguimiento facial basado en *Active Appearance Models* (Modelos Activos de Apariencia) es aplicado a cada una de las sesiones de vídeo, de las que se extraen un conjunto de rasgos faciales. Una vez estas características son extraídas, tras aplicarles un método de reducción de dimensionalidad serán empleadas como entrada a un clasificador *Support Vector Machine* (Máquina de Soporte Vectorial). Este paso nos dirige hacia el principal objetivo de este proyecto: conseguir determinar automáticamente si una persona está deprimida o no deprimida. En otras palabras, el objetivo de este proyecto es desarrollar una primera aproximación de un software para la evaluación de la depresión, capaz de clasificar entre sujetos deprimidos y no deprimidos mediante el empleo de técnicas novedosas de visión por computador y de aprendizaje máquina.

Es fácil pensar que la dificultad de este proyecto se basa en la construcción del clasificador que nos permite separar la clase deprimida de la no deprimida. Sin embargo, la principal complejidad con la que este proyecto cuenta es la de encontrar las características de la cara que son relevantes para la evaluación de la depresión. Es por ello que será esencial para el éxito del mismo el tener en cuenta previos trabajos de investigación sobre el comportamiento humano, así como el usar la propia intuición, algo realmente importante cuando se trata de desarrollar nuevas teorías y algoritmos.

Vale la pena recalcar que este proyecto no tiene como objetivo el mejorar los métodos de visión computador y de aprendizaje automático empleados (por ejemplo, *Active Appearance Models* y *Support Vector Machines*), sino que tratará únicamente de explotar las técnicas actuales y más punteras para así adaptarlas a nuestro propósito.

1.3. Estructura del proyecto

Esta memoria se estructura en tres partes principales: los datos, la clasificación y los resultados. Antes de entrar en profundidad en estos tres puntos, el Capítulo 2 se centrará en introducir los principales conceptos asociados a la depresión, tal como su definición, el proceso de evaluación de la depresión y el criterio de severidad.

1.3.1. Datos

El punto de partida de los datos usados en este proyecto es la base de datos Spectrum, comprendida por un conjunto de secuencias de vídeo en las que sujetos depresivos mantienen una entrevista con un terapeuta (ver la Sección 3.2). Sin embargo, tener los datos adecuados no es suficiente. Un método de seguimiento de características faciales basado en *Active Appearance Models* (AAMs) se aplica a cada una de estas sesiones de vídeo, permitiendo así la extracción de información sobre el comportamiento de la cara de los pacientes. Finalmente, y dado que este método de seguimiento no es infalible (normalmente debido a que se producen oclusiones de la cara), se llevará a cabo un procedimiento de "limpieza", obteniendo de esta forma la versión final de los datos que se usarán para la clasificación.

1.3.2. Clasificación

El proceso de clasificación se detalla en los Capítulos 4 y 5. El Capítulo 4 presenta el clasificador *Support Vector Machine* (SVM), así como las medidas que se usarán para evaluar los resultados de la clasificación.

En el Capítulo 5 se explica en detalle el procedimiento de extracción de características. En particular, se trata de un procedimiento basado en tres capas de extracción: el nivel fotograma, el nivel de grupo y el nivel global. Esta estructura en capas resultará en una reducción del tamaño de los datos, algo esencial dada la gran cantidad de datos manejada en este proyecto. Además, se presentarán un nuevo tipo de características faciales, a las que llamaremos *coeficientes de forma*. Estos se basan en un Análisis de Componentes Principales (*Principal Component Analysis*) y su interés reside en que son capaces de representar el comportamiento de la cara como conjunto.

1.3.3. Resultados

Los experimentos de clasificación se realizarán utilizando dos tipos de características diferentes: los *landmarks* resultantes del seguimiento de la cara mediante AAMs y los coeficientes de forma. Además, también se emplearán dos algoritmos de clasificación diferentes, ambos basados en un método de *K-fold cross validation* (validación cruzada con K subconjuntos). El primero emplea un esquema muy similar al *leave-one-out*, mientras que el segundo puede considerarse como un *4-fold* (empleo de 4 subconjuntos). Ambas técnicas se modifican ligeramente para poder solventar el hecho de que la base de datos Spectrum contiene más de una sesión para el mismo sujeto e incluir sesiones del mismo sujeto tanto en el conjunto de entrenamiento como en el de validación conllevaría la presencia de datos altamente correlados en ambos conjuntos.

El rendimiento de los experimentos vendrá dado por las curvas de Característica de Funcionamiento del Receptor (*Receiver Operating Characteristic curves* o curvas ROC) y por un conjunto de valores escalares, tales como el parámetro F, la exactitud de la clasificación y el área bajo la curva ROC.

1.4. Marco de desarrollo del proyecto

Este proyecto se ha desarrollado en el Human Sensing Lab, laboratorio de investigación centrado en la visión por computador y el aprendizaje automático de la Carnegie Mellon University y liderado por el Dr. Fernando de la Torre, en colaboración con el Affect Analysis Group de la University of Pittsburgh, dirigido por el psicólogo Dr. Jeffrey Cohn. Puesto que este proyecto es fundamentalmente interdisciplinario, no habría sido posible llevarlo a cabo sin la continua interacción entre estos dos grupos de trabajo.

Capítulo 2

Conceptos básicos de la depresión

2.1. Definición de depresión

La depresión es un trastorno mental común, caracterizado por la tristeza, la pérdida de interés y placer, sentimientos de culpabilidad, baja autoestima, desórdenes alimenticios y del sueño, baja energía y poca concentración.

Estos problemas pueden convertirse en crónicos o recurrentes, impidiendo al individuo desarrollar una vida normal. Aunque la mayor parte de los casos pueden ser tratados y curados con medicación y psicoterapia, los casos más severos pueden llevar incluso al suicidio.

2.2. Evaluación de la depresión

2.2.1. Visión general

La evaluación y diagnóstico de la depresión debe ser dirigido por un terapeuta experto, quien debe tener en cuenta las circunstancias actuales de la persona, su historial biográfico y sus síntomas actuales. La parte más importante es el examen del estado mental del paciente, el cual consiste en una evaluación de su estado de ánimo y del contenido de sus pensamientos, en particular de aquellos con temática de desesperanza y pesimismo, autolesión y suicido, y la ausencia de pensamientos o planes positivos.

La puntuación en una escala de valores no es suficiente para diagnosticar la depresión, pero sí provee una indicación de la intensidad de los síntomas durante un cierto período de tiempo. Existen varias escalas de valores utilizadas para este propósito, siendo la más célebre la llamada escala de Hamilton para la depresión

(*Hamilton Rate Scale for Depression* o HRSD), presentada a continuación.

2.2.2. La escala de Hamilton para la depresión

La escala de Hamilton para la depresión fue creada por Max Hamilton en 1960. Está pensada para su aplicación en pacientes con desórdenes afectivos de tipo depresivo ya diagnosticados. Se emplea para cuantificar los resultados a partir de una entrevista, durante la cual el entrevistador debe usar toda la información que tiene a sus disposición para ayudar al paciente y realizar la evaluación final.

La entrevista consiste en un cuestionario de 17 preguntas, llamadas el test de Hamilton (ver Apéndice A de la memoria completa), en la que se discuten los siguientes temas:

- **Estado de ánimo depresivo.** El indicador más importante de este tipo de estado de ánimo es la tendencia del paciente a llorar, siempre teniendo en cuenta su formación cultural.
- **Suicidio.** Un intento de suicido es dado la mayor puntuación. En ocasiones es complicado diferenciar entre un intento real de suicidio y un intento demostrativo, al que se le daría una puntuación más baja.
- **Trabajo y pérdida de interés.** La dificultad para el trabajo y la falta de interés en hobbies y actividades sociales se incluyen en este punto. Al paciente que ha dejado el trabajo debido a su enfermedad depresiva se le da la máxima puntuación.
- **Retardo.** Este punto es muy difícil de evaluar. A un paciente que se niega a hablar se le daría la mayor puntuación, pero éste sería poco adecuado para serle aplicada la escala de Hamilton. Otros comportamientos menos severos necesitan mucha paciencia y cuidado para ser puntuados, aunque puede llegar a conseguirse.
- **Agitación.** Ésta se define como inquietud asociada con ansiedad.
- **Síntomas gastrointestinales.** Ocurren en conexión tanto de la ansiedad como de la depresión.
- **Síntomas somáticos generales.** En la depresión estos son imprecisos y poco definidos, siendo por tanto extremadamente difícil conseguir un descripción satisfactoria de ellos por parte del paciente.

- **Hipocondría.** Es fácil de evaluar cuando está presente de forma obvia, siendo más dificultoso cuando se trata de preocupaciones hipocondríacas moderadas.
- **Perspicacia.** Debe ser considerada en relación con el grado de educación y conocimiento del paciente. Es muy importante distinguir entre un paciente con poca perspicacia, de otro que simplemente se niega a admitir que tiene una enfermedad mental.
- **Pérdida de peso.** Idealmente esta pérdida se mediría en libras o kilos, pero muy pocos pacientes conocen su peso normal y continúan un seguimiento del mismo. Ha sido pues necesario emplear una escala únicamente de tres puntos.

Durante la entrevista, el terapeuta debe puntuar cada pregunta. La suma de las puntuaciones de las 17 preguntas del test constituye la llamada puntuación de Hamilton.

2.2.3. Intensidad de la depresión

Una vez que ha sido determinada la puntuación de Hamilton del paciente, es necesario decidir cuál es la gravedad de sus síntomas de depresión. Cuanto mayor sea la puntuación de Hamilton, más deprimido está el paciente.

Para los psicólogos, definir un umbral fijo para la puntuación de Hamilton resulta una tarea difícil. Este umbral permitiría definir la frontera entre una persona deprimida y otra que, aunque con leves síntomas, ya ha remitido.

El objetivo de este proyecto es desarrollar un clasificador capaz de separar las sesiones (secuencias de vídeo) en las que aparecen pacientes deprimidos, de aquellas en las que aparece una persona sana. Para construir el clasificador se deben llevar a cabo dos tareas fundamentales: el entrenamiento y la validación. Para la realización de ambas es indispensable contar con el *ground truth*, que consiste en una serie de etiquetas asociadas a cada una de las sesiones de vídeo. Por ejemplo, un 1 si se trata de una persona deprimida y un -1 en otro caso.

Para poder asignar estas etiquetas *deprimido* y *no deprimido* a cada una de las secuencias de vídeo de la base de datos Spectrum, debe tomarse una decisión acerca de la relación entre la puntuación de Hamilton y la intensidad de la depresión de la persona en cuestión. En particular, y con el propósito de usarlo únicamente en el ámbito de este proyecto, definimos el siguiente criterio:

- **Sesiones *deprimidas*:** Serán aquellas con una puntuación de Hamilton mayor o igual a 15.

- **Sesiones *no deprimidas*:** Aquellas con una puntuación de Hamilton menor o igual a 7.

Capítulo 3

La base de datos Spectrum

3.1. Introducción

El capítulo anterior se ha dedicado a introducir los conceptos fundamentales asociados a la depresión, tales como su definición y cómo se evalúa. Este capítulo se centrará en los datos contenidos en la base de datos Spectrum. Esto incluirá la explicación de cómo fueron adquiridos, procesados y finalmente usados, en este mismo orden. Además, se discutirán algunas de las técnicas de visión por computador aplicadas en este proyecto, siendo el ejemplo más importante el empleo de *Active Appearance Models* (AAMs) para la interpretación de las imágenes de las secuencias de vídeo.

3.2. Colección de datos

El objetivo de este trabajo es analizar y comprender el comportamiento de los pacientes deprimidos. Para ello, es esencial contar con un conjunto de datos para analizar, como lo es la base de datos Spectrum. Esta base de datos consiste en una serie de vídeos en el que aparecen pacientes deprimidos mientras mantienen una entrevista con un terapeuta, a lo largo de la cual se les es realizado el test de Hamilton.

Durante el proceso de adquisición de datos (aproximadamente dos años), los pacientes fueron medicados o recibían terapia psicológica, por lo que se espera su mejoría a lo largo del tiempo. Para realizar el seguimiento de la gravedad de la depresión, los pacientes fueron entrevistados una vez a la semana por un terapeuta, durante la cual se les realizó el test de Hamilton. Sin embargo, no todas estas en-

trevistas fueron grabadas, sino únicamente aquellas pertenecientes a las semanas 1, 7, 13 y 21 desde el inicio del tratamiento del paciente, constituyendo así la base de datos Spectrum.

Además de las secuencias de vídeo grabadas, se recogió información personal sobre el paciente. En concreto, su edad, sexo, estado civil, etnia y raza. Esta información no se usará directamente en los experimentos de clasificación, aunque si nos ayudará a entender la población estadística con la que estamos tratando.

Teóricamente, al final del proceso de adquisición de datos, cada sujeto debe haber sido grabado en cuatro sesiones de entrevista. Sin embargo, algunos de ellos dejaron el tratamiento antes de que todas las grabaciones se completaran, o simplemente no se presentaron a la sesión correspondiente a una semana en concreto, lo cual no implicó que no se les incluyera en la base de datos.

El número total de sesiones y sujetos presentes en la versión *limpia* de la base de datos Spectrum se muestra en la Tabla 3.1. Se organizan en sesiones 1, 2, 3 y 4, que se corresponden con las semanas 1, 7, 13 y 21 (respectivamente) en las que los pacientes se sometieron al test de Hamilton.

# sujetos	**Ses. 1**	**Ses. 2**	**Ses. 3**	**Ses. 4**	**Total # sessions**
52	48	39	33	29	149

Tabla 3.1: Número total de sujetos y sesiones de la base de datos Spectrum.

3.3. Procesamiento: seguimiento de características faciales

3.3.1. Visión general

Una vez todas las grabaciones de los sujetos han terminado, ya estamos listos para usar las secuencias de vídeo para nuestros propósitos de clasificación. Sin embargo, es claro que para ello no es posible utilizar toda la información contenida en cada uno de los fotogramas de las secuencias. Será pues necesario solventar este problema empleando técnicas de visión por computador. Estas técnicas nos permiten centrarnos en la información importante de los fotogramas, que en el caso que nos ocupa se trata de la cara de los pacientes. En particular, las características faciales

de los pacientes se extraerán aplicando *Active Appearance Models* (AAMs), presentados en esta sección. Además, se hará una breve descripción de la PCA (*Principal Component Analysis*), base de los AAMs.

3.3.2. Principal Component Analysis

La PCA, también conocida como la transformación Karhunen-Loeve o la transformación Hotelling, es una pieza fundamental en la construcción del modelo usando AAMs [6, 4]. En esencia, es una herramienta estadística que permite una representación compacta de datos de grandes dimensiones. La literatura sobre el tema es muy extensa; sólo presentaremos aquí un pequeño repaso.

La PCA permite encontrar una representación de baja dimensión de los datos que a la vez minimiza el error cuadrático de reconstrucción. Dado un conjunto de n muestras $\boldsymbol{x}_i \in \Re^d$ (asumamos, sin pérdida de generalidad, datos de media cero) organizados en una matriz $\mathbf{X} \in \Re^{d \times n}$, PCA encuentra una matriz $\mathbf{B} \in \Re^{d \times d'}$ tal que $\left\|\mathbf{X} - \mathbf{B}\mathbf{B}^\top \mathbf{X}\right\|_2$ es mínimo, sujeto a que $\mathbf{B}$ sea ortogonal y diagonalizando la matriz de covarianza $cov\left(\mathbf{B}^\top \mathbf{X}\right)$, con $d' < d$.

La matriz $\mathbf{C} = \mathbf{B}^\top \mathbf{X}$ con $\mathbf{C} \in \Re^{d' \times n}$ es una representación de menor dimensión de la matriz de datos original. Las variables subyacentes de la nueva representación están incorreladas; si la distribución del espacio original es Gaussiano, entonces las variables son independientes. Las columnas de $\mathbf{B}$ son las direcciones de máxima varianza en el espacio original (típicamente ordenadas), y son generalmente llamadas modos de variación. Estos describen una relación multivariable entre las variables que descompone la variabilidad encontrada en los datos en componentes incorrelados (independientes, si se asume gaussianidad).

Cuando es implementada, la PCA se resuelve normalmente utilizando una descomposición de autovectores de la matriz de covarianza, guardando aquellos autovectores asociados con los d' autovalores mayores. Otra posibilidad es aplicar *Singular Value Decomposition* (descomposición en valores singulares):

$$\begin{aligned} \mathbf{X} &= \mathbf{U}_{d \times d} \mathbf{D}_{d \times n} \mathbf{V}^\top_{n \times n} \\ \mathbf{B} &= \mathbf{U}_{1,\ldots,d \times 1,\ldots,d'} \end{aligned} \tag{3.1}$$

El valor de los autovalores conservados (y descartados) está directamente relacionado con la cantidad de información conservada, considerada normalmente como la energía asociada a los datos. La energía total viene descrita por el sumatorio

de todos los autovalores. Quedándonos con el número de autovalores que permiten describir un cierto porcentaje de los datos, los datos quedan reducidos a una representación que muestra la información más relevante de los mismos. Cuando usamos *Singular Value Decomposition* (SVD), los *singular values* o valores singulares vienen dados por la raíz cuadrada de los autovalores de $cov\,(\mathbf{X})$. En términos generales:

$$cov\,(\mathbf{X}) \quad = \mathbf{X}\mathbf{X}^\top = \quad \mathbf{U}\mathbf{D}\mathbf{V}^\top\mathbf{V}\mathbf{D}^\top\mathbf{U}^\top = \mathbf{U}\mathbf{D}\mathbf{D}^\top\mathbf{U}^\top \tag{3.2}$$

3.3.3. Seguimiento de características faciales mediante AAMs

3.3.3.1. Visión general

Los *Active Appearance Models* (AAMs) son una técnica basada en modelos empleada para la interpretación de imágenes en las que aparecen objetos variables. Según Cootes *et al* [5], un *Active Appearance Model* contiene un modelo estadístico de la forma y de la apariencia, en niveles de grises, del objeto de interés, lo cual puede generalizarse a casi cualquier otro ejemplo válido. Sin embargo, los métodos de modelos activos, tales como los AAMs, son particularmente populares en el campo del seguimiento de caras, por lo que nos centraremos en él para su descripción.

El método AAM puede resumirse en dos fases:

- **Fase de entrenamiento.** En esta fase se aprende la relación entre los desplazamientos de los parámetros del modelo y los errores residuales inducidos entre una imagen de entrenamiento y un ejemplo sintetizado del modelo. Finalmente, se construye un modelo que describe el objeto a seguir.

- **Fase de ajuste.** Una vez se ha obtenido el modelo, para ajustarlo con una imagen se miden los errores residuales actuales y se usa el modelo para predecir los cambios en los parámetros actuales, dando como resultado un mejor ajuste. En términos generales, un buen ajuste se consigue tras una serie de iteraciones.

3.3.3.2. Modelo de apariencia

Los modelos son generados combinando un modelo de variación de forma con otro de variación de apariencia. Se requiere para ello tener un conjunto de imágenes etiquetadas, en las que los *landmarks* marcan el objeto de interés. Estos *landmarks* deben colocarse en las posiciones que permitan describir mejor el objeto; en nuestro caso, al ser éste una cara, los *landmarks* se situarían en los ojos, las cejas, la nariz,

la boca y el contorno. En particular, se usará el modelo de forma mostrado en la Figura 3.1, constituido por 66 *landmarks*.

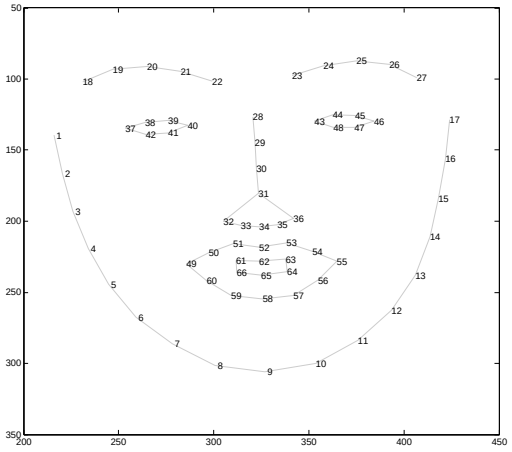

Figura 3.1: *Landmarks* empleados en el AAM.

3.3.3.3. Construcción del modelo

El Modelo Activo de Apariencia se construye a partir de una serie de ejemplos de entrenamiento, es decir, imágenes de caras donde los *landmarks* se encuentran marcados. Las formas son alineadas entre sí empleando un procedimiento iterativo de alineamiento para eliminar las componentes de movimiento rígido, como las traslaciones, rotaciones y cambios de escala [5].

Las coordenadas x-y de los puntos se ordenan en una matriz de datos. En el caso de nuestro modelo AAM, la dimensión de cada uno de los ejemplos es $d = 2 \times 66$, es decir, las coordenadas 2D de los 66 *lanmarks*. Tras esto, se aplica la PCA, conservándose un cierto porcentaje de la energía de los datos.

Los triángulos de textura de cada una de las imágenes se *warpean* en la forma media, normalmente ya determinada. Esto resulta en un conjunto de imágenes *warpeadas* en un marco de referencia común. Los pixels de estas imágenes se organizan en vectores columna, correspondiendo cada una de las filas a un cierto valor de pixel. El espacio definido por las columnas de esta matrix describe un rango de posibles variaciones de textura del objeto. PCA vuelve entonces a aplicarse, resultando en un modelo lineal de apariencia.

Por tanto, toda cara podrá ser representada como una forma media más una

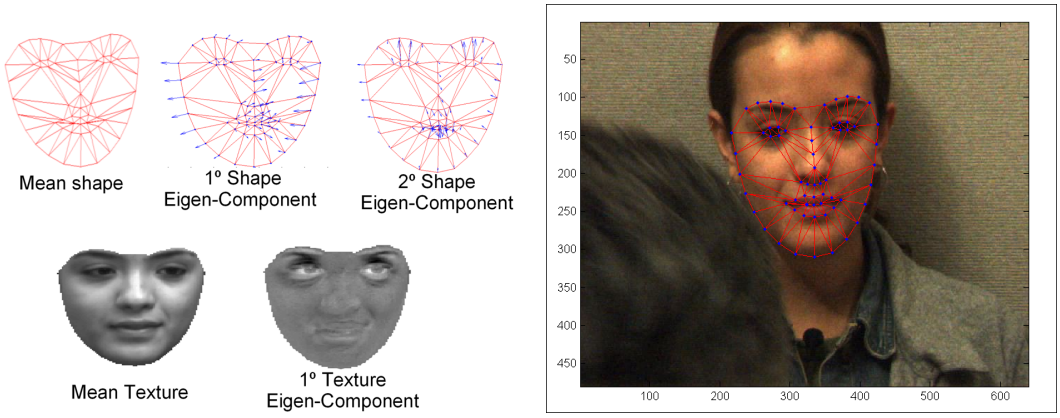

Figura 3.2: Izquierda: modelo AAM de forma y apariencia. Derecha: ejemplo de seguimiento de características faciales con AAMs. En azul, los *landmarks*. En rojo, la malla de textura.

serie de combinaciones lineales de m formas bases (o formas básicas):

$$\mathbf{O} = \mathbf{S}_0 + \sum_{i=1}^{m} l_i \mathbf{S}_i \tag{3.3}$$

donde $\mathbf{S}_i \in \Re^{2\times 66}$ es la i-ésima matriz de forma básica, siendo cada fila las variaciones en x o y de cada uno de los n puntos (a lo largo de las columnas) para este modo.

La textura (apariencia) aplicada a esta forma resultará en un combinación lineal similar:

$$\mathbf{A} = \mathbf{A}_0 + \sum_{i=1}^{m'} r_i \mathbf{A}_i \tag{3.4}$$

donde $\mathbf{A}_i \in \Re^{1\times p}$ es la i-ésima matriz de apariencia básica, siendo cada fila la intensidad en valores de grises de la variación de cada uno de los p pixels (a lo largo de las columnas) para este modo.

3.3.3.4. Ajuste del modelo

Para ajustar el AAM a una nueva imagen se minimiza una función de error del ajuste, mediante la actualización de los parámetros del modelo. Normalmente, la función de error incluye la norma L2 de la diferencia de textura de la imagen generada por el modelo y la imagen original.

$$E\left(\mathbf{p}; l_{1,\ldots,m}; r_{1,\ldots,m'}\right) = \left\| \mathbf{A}_0 + \sum_{i=1}^{m'} r_i \mathbf{A}_i - I\left(W\left[\boldsymbol{T}_s\left(\mathbf{p}\right)\left\{\mathbf{S}_0 + \sum_{i=1}^{m} l_i \mathbf{S}_i\right\}\right]\right)\right\|_2 \tag{3.5}$$

donde $I\left(W\left[\mathbf{O}\right]\right)$ son los pixels de las imágenes al ser *warpeadas* en el marco de referencia usado para la generación de la base de apariencia.

Capítulo 4

Conceptos básicos de la clasificación

4.1. Introducción

El propósito de la clasificación es el de asignar una etiqueta, en este caso "deprimido" o "no deprimido" a una muestra de datos (por ejemplo, una cara). La clasificación automática puede ser divida en dos fases: entrenamiento y validación.

La primera fase es el entrenamiento del clasificador, en la que un conjunto de muestras etiquetadas, llamado conjunto de entrenamiento, se utiliza para ajustar los parámetros internos del algoritmo de clasificación. Una vez esta fase es completada, el algoritmo con los parámetros ya ajustados puede aplicarse a nuevas muestras de datos para estimar o predecir su correspondiente etiqueta de clase (teóricamente desconocida). Este conjunto de datos es el llamado conjunto de prueba o validación.

En un problema de clasificación de dos clases, el conjunto de entrenamiento suele expresarse como:

$$\{(\boldsymbol{x}_1, y_2), \ldots, (\boldsymbol{x}_p, y_p)\} \tag{4.1}$$

donde $\boldsymbol{x}_i \in \Re^n$ es el vector de características de la muestra a clasificar y $y_i \in \{-1, 1\}$ es la correspondiente etiqueta, que indica a qué clase pertenece en función de su signo.

4.2. Diseño del clasificador

A la hora de diseñar un clasificador, podemos distinguir cuatro elecciones de diseño importantes:

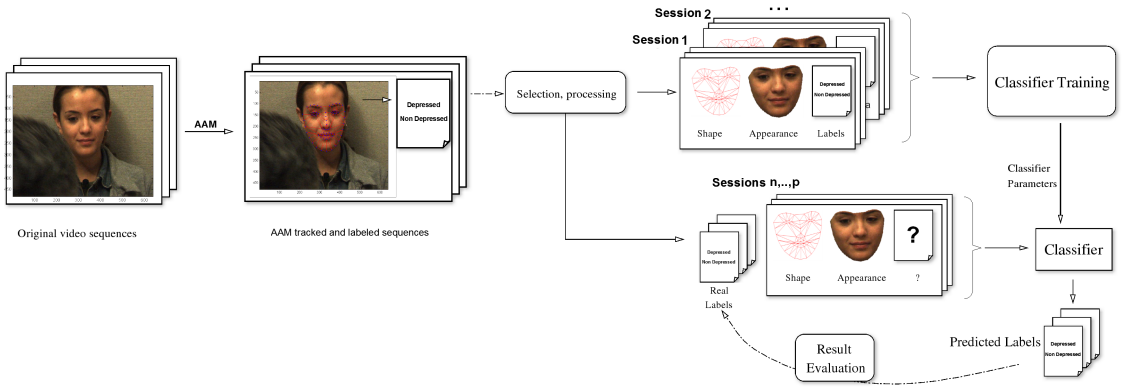

Figura 4.1: Visión general del algoritmo de clasificación.

- **Conjunto de entrenamiento.** Deben seleccionarse las muestras adecuadas para ser incluidas en el conjunto de entrenamiento. Ejemplos demasiado sutiles (no perceptibles en las características extraídas) únicamente servirían para "confundir" al clasificador en su fase de entrenamiento, resultando en un ajuste de parámetros peor que si estos ejemplos hubieran sido excluidos del conjunto. Esto nos lleva a escoger sesiones de vídeo donde los pacientes están o muy deprimidos, o muy poco deprimidos, hecho que se refleja en el criterio de intensidad recogido en la Sección 2.2.3.

- **Representación de las muestras de datos.** La muestra de datos a clasificar (conceptualmente ,una cara) debe ser representada de forma que pueda ser procesada por el clasificador. Generalmente, se utiliza un vector real multidimensional $\boldsymbol{x}_i \in \Re^n$, cuyos componentes son las características extraídas, las cuales contienen información discriminante para la clasificación.

- **Elección del clasificador.** Se han publicado resultados con casi todo los algoritmos de clasificación existentes [12, 16], los cuales dependen, además de la aplicabilidad del algoritmo, de la naturaleza de los datos a clasificar, que en este caso depende a su vez de las características extraídas de ellos y de las condiciones en que la base de datos fue generada. En particular, se han publicado muy buenos resultados empleando *Support Vectors Machines* (SVMs) en el contexto de la clasificación de secuencias de vídeo procesadas con AAMs. Los SVMs son una herramienta de clasificación muy versátil y eficiente, y será pues el algoritmo empleado en este proyecto (ver la versión completa de esta memoria para una descripción matemática detallada de los

SVMs).

- **Esquema de validación.** Como esquema de validación de la clasificación se elegirá la validación cruzada o *cross validation*. Con esta técnica, el ciclo de entrenamiento y validación se ejecuta múltiples veces en diferentes subconjuntos del conjunto total etiquetado disponible.

  El procedimiento general es el *K-fold cross validation*, en el que el conjunto total de datos se particiona en K subconjuntos. En cada ciclo de entrenamiento-validación, un subconjunto es empleado para la validación, mientras que los K-1 restantes se usan como información de entrenamiento. Este proceso se repite de forma que cada uno de las muestras forme parte del conjunto de validación una sola vez. Durante esta validación cruzada se recogen estadísticas que servirán para evaluar el rendimiento del clasificador. Además, parámetros del clasificador (como el parámetro C del SVM o el kernel gaussiano σ, cuya discusión se encuentra en la versión íntegra de esta memoria) son normalmente ajustados usando este procedimiento [2]. En el caso en el que K toma un valor igual al número total de muestras disponibles para la clasificación, en cada iteración una sola muestra es empleada para la validación, mientras que las demás se utilizan como conjunto de entrenamiento. Este método es el llamado *leave-one-out*.

4.3. Evaluación de la clasificación

4.3.1. Parámetros de evaluación

La bondad de los experimentos de clasificación se mide fundamentalmente por su exactitud, la cual se computa teniendo en cuenta la tasa de falsos y verdaderos positivos y la tasa de falsos y verdaderos negativos. En adelante, nos referiremos a ellos empleando la terminología inglesa y sus definiciones se presentan a continuación:

- **True positive rate (TPR) o tasa de positivos verdaderos.** Es el número de ejemplos positivos que son correctamente clasificados, respecto al número total de muestras positivas.

- **False negative rate (FNR) o tasa de falsos positivos.** Es el número de ejemplos positivos que son incorrectamente clasificados, respecto al número total de muestras positivas.

- **True negative rate (TNR) o tasa de negativos verdaderos.** Es el número total de ejemplos negativos que son correctamente clasificados, respecto al número total de muestras negativas.

- **False negative rate (FNR) o tasa de falsos negativos.** Es el número total de ejemplos negativos que son incorrectamente clasificados, respecto al número total de muestras negativas.

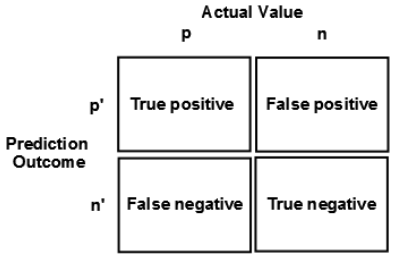

Figura 4.2: Matriz de confusión.

- **Exactitud.** En clasificación binaria es el número de *true positives* más el de *true negatives* sobre el número total de muestras.

- **Parámetro F.** Es función de las medidas de precisión y de retorno:

$$F = \frac{2 \cdot \text{precision} \cdot \text{retorno}}{\text{precision} + \text{retorno}} \tag{4.2}$$

4.3.2. Curvas ROC

Los parámetros descritos en la sección anterior dependen del valor del umbral usado para la clasificación dado por el SVM. Para aportar independencia sobre este valor a la evaluación de la clasificación, se usan las curvas ROC (*Receiver Operating Characteristic*) para representar los resultados de la clasificación. Éstas muestran los valores que toma el *False Positive Rate* (FPR) respecto al *True Positive Rate* (TPR), al realizar un barrido de todos los posibles valores del umbral del SVM.

Una medida para la evaluación de la clasificación que suele extraerse de estas curvas es el área bajo la misma, siendo la curva ROC ideal aquella de área 1.

Capítulo 5

Extracción de características

5.1. Introducción

Si la técnica de seguimiento de características faciales, en este caso basada en AAMs es lo suficientemente robusta y precisa, es concebible que el usar únicamente la información de forma sea suficiente para poder distinguir entre los pacientes deprimidos y no deprimidos. Sin embargo, la información de apariencia sería capaz de describir con mayor exactitud la morfología y las expresiones de la cara, aunque también es cierto que prácticamente cualquier cambio en la apariencia se refleja en la forma, resultando en un movimiento de los *landmarks* al realizar el ajuste del modelo AAM.

Sería posible utilizar toda la información de forma de la cara como vectores de características en la clasificación. Sin embargo, esto supondría tratar con una cantidad enorme de datos:

- Cada cara tiene 66 *landmarks*.
- Cada secuencia de vídeo tiene una media de 20.000 fotogramas.
- La base de datos Spectrum está compuesta por 149 sesiones.

Es pues necesario desarrollar un método de extracción de características que nos permita llevar a cabo una reducción de la cantidad de datos utilizados en el algoritmo de clasificación.

5.2. Reducción de los datos mediante análisis estadístico

El método de extracción de características desarrollado consta de tres niveles: el nivel fotograma, el nivel de grupo y el nivel global.

5.2.1. Nivel fotograma

Las medidas estadísticas se tomarán sobre la cantidad de variabilidad existente entre dos fotogramas consecutivos. Es decir, nuestro interés se centra en cuánto cambia una cara de un fotograma al siguiente, lo cual puede ser interpretado como una medida de velocidad.

El nivel de fotograma consiste en tomar una secuencia de fotogramas de una sesión dada y computar la diferencia entre los fotogramas consecutivos. El resultado será pues una secuencia de vectores, conteniendo cada uno de ellos la "velocidad" de los *landmarks*, en otras palabras, la diferencia de posición que cada uno de los *landmarks* experimenta entre dos fotogramas consecutivos.

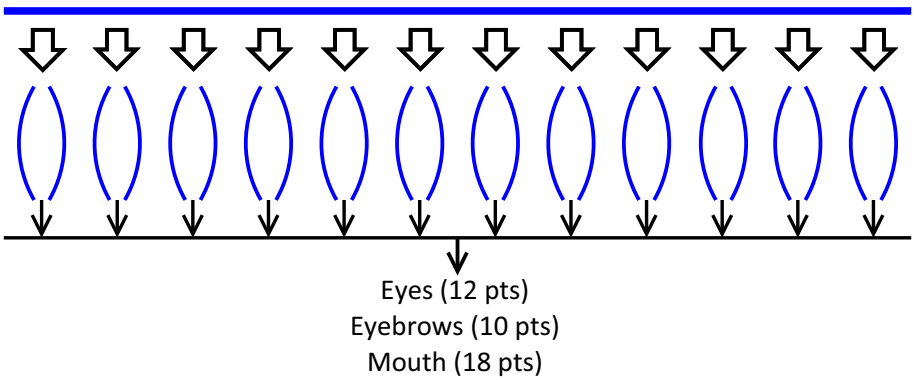

Figura 5.1: Extracción de características: nivel de fotograma.

5.2.2. Nivel de grupo

Para construir este nivel se siguen dos pasos:

En primer lugar, la secuencia de vectores resultante del nivel de fotograma se segmenta en una serie de ventanas no solapadas. El objetivo principal es el de ser capaz de capturar el movimiento del sujeto en pequeños intervalos de tiempo. La longitud

óptima de estas ventanas es desconocida a priori, pero puede ser determinada tras el análisis de los resultados de la clasificación. En particular, las simulaciones se llevarán a cabo utilizando intervalos de tres longitudes diferentes: 150 fotogramas, 300 fotogramas y 600 fotogramas, correspondientes a ventanas de 5, 10 y 20 segundos, respectivamente.

En segundo lugar, se busca encontrar un solo vector de características para cada una de las ventanas del proceso de segmentación. Para este propósito, se realizan una serie de medidas estadísticas sobre los vectores del nivel de fotograma: la media, la mediana y la desviación estándar. Concatenando todas las medidas de cada uno de los intervalos de la secuencia, conseguimos un único vector de características para cada una de ellos, lo cual constituye el resultado de este nivel de extracción.

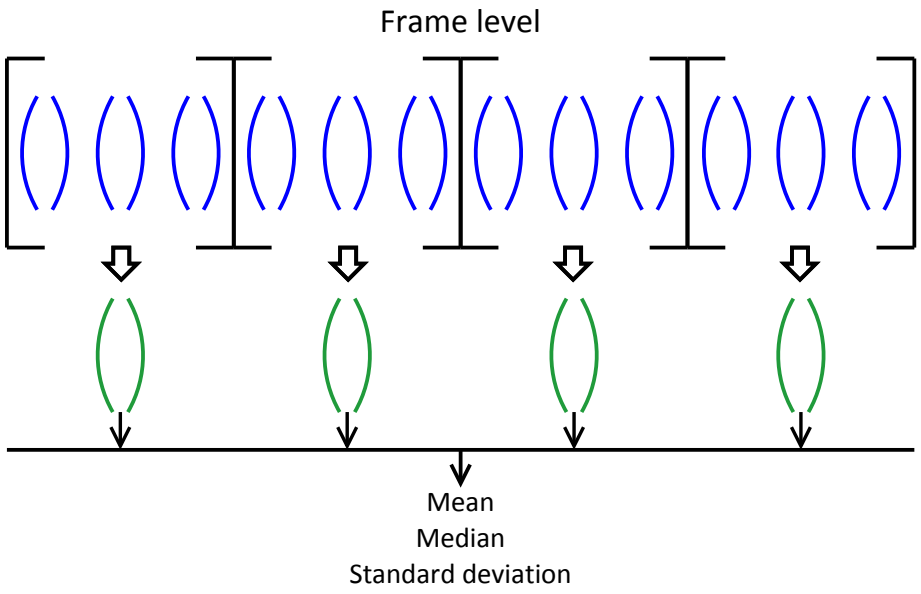

Figura 5.2: Extracción de características: nivel de grupo.

5.2.3. Nivel global

Como paso final en el proceso de extracción de características, se aplican un nuevo conjunto de medidas sobre los vectores resultantes del nivel de grupo. En este caso, estas medidas son: la media, la mediana, el máximo y el mínimo. Estas medidas se aplican a cada una de las estadísticas de los vectores del nivel de grupo, habiendo por tanto 12 posibles combinaciones entre ellas.

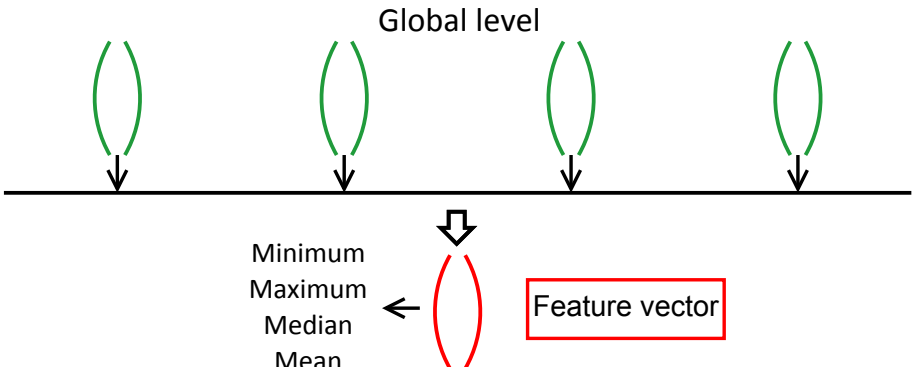

Figura 5.3: Extracción de características: nivel global.

5.3. Otras características: coeficientes de forma

El método de extracción de características presentado en la sección anterior se ha enfocado como una manera de reducir la gran cantidad de datos que supone el tener en cuenta los 66 *landmarks* que constituyen la forma completa de la cara. Sin embargo, esta técnica puede ser aplicada a cualquier tipo de datos, como por ejemplo los llamados coeficientes de forma.

En términos generales, dada una cara, los coeficientes de forma describen la variación de movimiento que ésta experimenta con respecto a una cierta forma media, dando así una idea de la cantidad de energía que contiene. Para computar los coeficientes de forma de una cara, en primer lugar será necesario construir un modelo, es decir, un conjunto de vectores que definen un cierto subespacio. Los coeficientes que se necesitan para reconstruir esta cara a partir del modelo es a lo que llamamos coeficientes de forma. En otras palabras, son los valores escalares que multiplicados por los vectores que definen el subespacio nos dan una representación de la cara dada.

El modelo de caras debe ser capaz de representar lo más precisamente posible cada una de las caras de la base de datos. La situación ideal sería pues poder tener en cuenta uno cada de los fotogramas de todas las secuencias de vídeo disponibles. Sin embargo, y como ya se ha discutido anteriormente, esto supondría manejar una cantidad excesiva de datos. Durante el proceso de seguimiento de características faciales basado en AAMs, se llevó a cabo el etiquetaje manual de los *landmarks* de aproximadamente el 5 % de los fotogramas, con el objetivo tanto de aumentar

la fiabilidad del seguimiento, como de solventar situaciones en que se producían oclusiones en la cara de los pacientes. Dado que la posición de los *landmarks* en estos fotogramas es prácticamente perfecta, y teniendo en cuenta que estos pertenecen a los mismos sujetos y sesiones que se emplean para la clasificación, esta serie de caras etiquetadas manualmente constituyen el conjunto de datos más adecuado para construir el modelo.

Una vez elegidos los datos, estos se organizan en una matriz $\mathbf{M}$ de tamaño $132 \times N$, siendo N el número total de caras manualmente etiquetadas. Cada una de las filas contiene las coordenadas de los 66 *landmarks* de una de las caras, estando las coordenadas x e y situadas una detrás de las otras. Además, se computa y se guarda la forma media de este conjunto de caras, puesto que será un elemento fundamental en el proceso de reconstrucción de los datos.

Una vez construida la matriz $\mathbf{M}$, a cada una de las caras que la forman se le es sustraída la cara media (forma media) y se le aplica la PCA a la matriz resultante mediante el método de descomposición SVD. Como resultado de esta descomposición (ver la versión íntegra de la memoria para más detalle) obtenemos una matriz $\mathbf{V}$ de tamaño 132×132, conteniendo en cada una de sus columnas los vectores que definen un subespacio que permite una representación de los datos de partida completamente incorrelada (una matriz de covarianza diagonal).

Con el objetivo de conservar únicamente aquellos vectores que representan la mayor parte de la energía de los datos, concretamente el 95 %, nos quedamos con los 10 primeros vectores de la matriz $\mathbf{V}$.

Finalmente, para hallar los coeficientes de forma de una cara dada, sólo será necesario proyectarla en el subespacio definido por la matriz $\mathbf{V}$. Es decir, siendo $\mathbf{m}$ un vector que contiene las coordenadas *x*-*y* de los *landmarks* de una cara una vez le es sustraída la cara media, sus coeficientes de forma vienen dados por:

$$\mathbf{c} = \mathbf{m} \cdot \mathbf{V} \tag{5.1}$$

Estos coeficientes pueden calcularse tanto para una cara perteneciente al conjunto empleado para construir el modelo, como para cualquier otra cara de la base de datos. Si posteriormente estos coeficientes se aplicaran con el objetivo de reconstruir la cara a la que pertenecen, la reconstrucción en el caso de los coeficientes de una cara utilizada para construir el modelo, con respecto a la reconstrucción de cualquier otra cara de la base de datos, sería más precisa. Una comparativa de los errores de reconstrucción puede encontrarse en la versión completa de la memoria.

Capítulo 6

Resultados de la clasificación

6.1. Introducción

El proceso de clasificación comprende tres puntos importantes a tener en cuenta: la implementación, los datos y el esquema de entrenamiento-validación.

- **Implementación**. La implementación del SVM se basa en la librería LibSVM [3]. Básicamente, se utilizan dos de sus funciones: *svmtrain* y *svmpredict*. La primera entrena al clasificador y construye el modelo basándose en el conjunto de entrenamiento. La segunda realiza la validación del clasificador dada una muestra y se correspondiente etiqueta, retornando pues información sobre la exactitud de la clasificación.

- **Datos para la clasificación**. El criterio de intensidad de la depresión presentado en la Sección 2.2 se aplica a todas las sesiones de la base de datos Spectrum, con el objetivo de utilizar los ejemplos de depresión más evidentes en el proceso de la clasificación. El resultado es finalmente que el número total de sesiones utilizadas en los experimentos de clasificación son: 66 sesiones "deprimidas" y 41 sesiones "no deprimidas", perteneciendo a un total de 51 sujetos.

  Atendiendo al tipo de características extraídas de estas sesiones, se realizan dos tipos de experimentos. El primer tipo utiliza como vectores de características aquellos extraídos de los *landmarks* de la cara tras aplicar el esquema de extracción de tres niveles presentado en el capítulo anterior. El segundo tipo, en vez de partir de los puntos de la cara, partirá de los coeficientes de forma de cada una de ellas, aplicándoles posteriormente el mismo procedimiento de extracción de características.

- **Esquema de entrenamiento-validación.** Se emplearán dos esquemas diferentes, ambos basados en el método de *K-fold cross validation*. La Figura 6.1 muestra el diagrama de flujo del algoritmo.

  En el primero de ellos, K toma el valor del número total de sujetos (51), pasando a ser un algoritmo *leave-one-out* modificado, al que podríamos llamar *leave-one-subject-out*. Durante este proceso, en cada iteración se utilizan todas las sesiones de un sujeto (de una a cuatro) para la validación y las restantes se dejan como conjunto de entrenamiento. Si en lugar de éste se empleara el método *leave-one-out* clásico, estaríamos introduciendo en ambos conjuntos de datos (entrenamiento y validación) sesiones del mismo sujeto, información altamente correlada.

  En el segundo esquema de entrenamiento-validación, un 25 % de los datos se emplean para la validación de la clasificación, mientras que el 75 % restante entrena al SVM, pudiendo ser considerado pues como un *4-fold*. A la hora de construir los dos conjuntos de datos se tendrá de nuevo cuidado de no incluir en ambos conjuntos sesiones del mismo sujeto. Si tenemos en cuenta que no todos los sujetos tienen el mismo número de sesiones, resulta que, estrictamente hablando, en cada iteración no habrá el mismo número de sesiones en cada conjunto de datos.

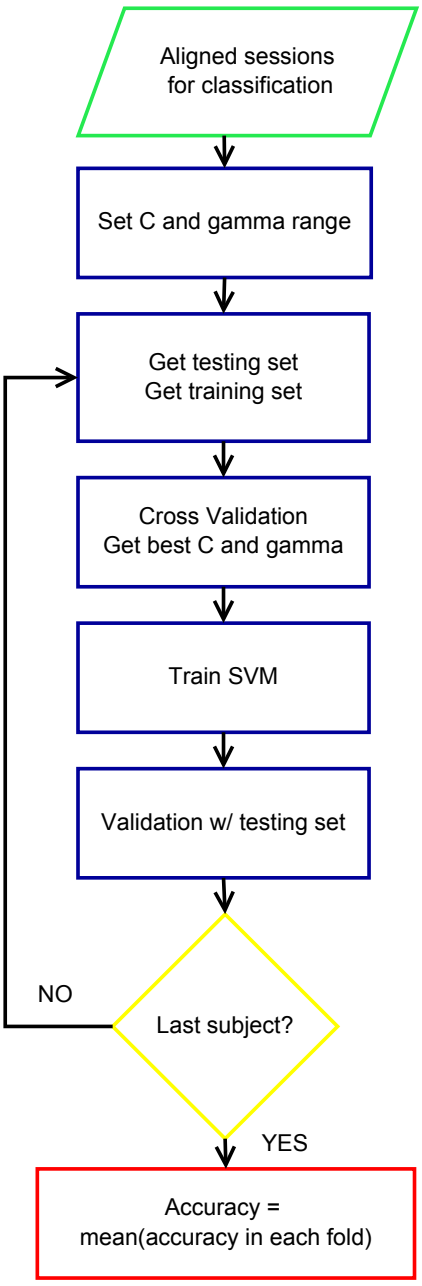

Figura 6.1: Diagrama de flujo del esquema general de entrenamiento-validación.

6.2. Experimentos y resultados

6.2.1. Visión general

Los experimentos se centrarán en partes específicas de la cara (cejas, ojos y boca) y en los coeficientes de forma. Además, se incluirán experimentos empleando todas las posibles combinaciones de estos, excepto los coeficientes de forma. Los coeficientes de forma se excluyen de estas combinaciones para evitar la presencia de información redundante en la clasificación, puesto que estos son una representación de la cara como conjunto.

Los experimentos que usan una única característica son repetidos utilizando diferentes longitudes de ventana en el nivel de grupo del proceso de extracción de características, concretamente 150, 300 y 600 fotogramas. Los experimentos que utilizan combinaciones de características sólo se realizan con un intervalo de 300 fotogramas, puesto que éste será el que ofrezca mejores resultados de clasificación, tal y como se comprobará más adelante.

Los resultados se presentan en una tabla que contiene los siguientes parámetros, los cuales evalúan la bondad de la clasificación (ver la Sección 4.3 para sus definiciones): exactitud, *True Positive Rate*, *True Negative Rate*, parámetro F y área debajo de la curva ROC.

Los resultados presentados en las siguientes secciones son únicamente una recopilación de todos los incluidos en la memoria completa y su discusión se basa exclusivamente en la exactitud de la clasificación. Para consultar las curvas ROC y las tablas de resultados completas debe consultarse la versión íntegra de esta memoria.

6.2.2. Resultados con leave-one-subject-out

Un resumen de los resultados obtenidos utilizando un esquema de entrenamiento-validación *leave-one-subject-out* se presenta en las Tablas 6.1 y 6.2.

Entre todas las longitudes de intervalo elegidas para los experimentos (150, 300 y 600 fotogramas), es la de 300 fotogramas la que ofrece mejores resultados. Este hecho destaca más en el caso de los coeficientes de forma, cuyos resultados de clasificación empleando 300 fotogramas superan en gran manera a aquellos con 150 y 600 fotogramas.

Los ojos y las cejas no parecen ser determinantes a la hora de distinguir entre los sujetos deprimidos y no deprimidos, puesto que la exactitud de la clasificación emple-

Característica \ Intervalo	150	300	600
Ojos	0.498	0.671	0.652
Cejas	0.626	0.666	0.632
Boca	0.611	0.749	0.592
Coefs. de forma	0.583	0.761	0.712

Tabla 6.1: Exactitud de la clasificación empleando características simples.

Intervalo \ Característica	Cejas Ojos	Boca Ojos	Boca Cejas	Ojos Cejas Boca
300	0.560	0.722	0.650	0.691

Tabla 6.2: Exactitud de la clasificación empleando combinación de características.

ando estas características no supera el 68 %. Sin embargo, no puede decirse lo mismo de la boca y de los coeficientes de forma. Estas características son capaces de alcanzar una exactitud del 75 % en la clasificación. Un resultado satisfactorio era esperado en el caso de los coeficientes de forma, puesto que estos son una representación de la cara completa y por tanto contienen una gran cantidad de información sobre ella; pero no en el caso de la boca. La boca por ella misma es capaz de discriminar con una exactitud razonable entre las dos clases. Aunque resulte sorprendente, si pensamos detenidamente sobre ello nos damos cuenta de que la boca puede expresar una gran variedad de sentimientos: felicidad – mediante una sonrisa –, tristeza – mediante la depresión de los *corners* de la boca –, sorpresa, etc. Además, también lleva la información asociada al habla: tanto la velocidad como la cantidad de tiempo que la persona habla.

Por su parte, los experimentos que utilizan combinaciones de características para la clasificación no ofrecen resultados satisfactorios, debido a diversos factores. Los exactitud dada por *Cejas+Ojos* es muy cercana a la aleatoriedad, lo que era de esperar si tenemos en cuenta que estamos combinando dos características que en solitario dan resultados muy pobres. En cuanto a los ejemplos restantes (*Cejas+Boca*, *Boca+Ojos* y *Boca+Ojos+Cejas*), todos ellos incluyen la boca, aunque en combinación con características menos discriminantes, dando lugar por tanto a resultados peores

que aquellos empleando únicamente la boca.

En conclusión, el mejor resultado obtenido mediante un esquema de entrenamiento-validación *leave-one-subject-out* viene dado a partir del uso de los coeficientes de forma como características en la clasificación, en conjunto con una longitud de ventana en el nivel de grupo de 300 fotogramas.

6.2.3. Resultados con 75 % entrenamiento - 25 % validación

Un resumen de los resultados obtenidos utilizando un esquema de entrenamiento-validación 75 % entrenamiento - 25 % validación se presenta en las Tablas 6.3 y 6.4.

Característica \ Intervalo	**150**	**300**	**600**
Ojos	0.540	0.619	0.584
Cejas	0.541	0.608	0.589
Boca	0.583	0.696	0.601
Coefs. de forma	0.659	0.713	0.663

Tabla 6.3: Exactitud de la clasificación empleando características simples.

Intervalo \ Características	**Cejas Ojos**	**Boca Ojos**	**Boca Cejas**	**Ojos Cejas Boca**
300	0.653	0.696	0.692	0.663

Tabla 6.4: Exactitud de la clasificación empleando combinación de características.

Como fácilmente puede comprobarse comparando las tablas de esta sección con aquellas de la Sección 6.2, los resultados de estos experimentos son en general peores. Este hecho es fácil de comprender si se tiene en cuenta que para el SVM es mucho más sencillo clasificar una, dos, o como mucho cuatro sesiones (dependiendo de las sesiones asociadas al sujeto que se deja fuera), que clasificar una cuarta parte de todas las sesiones (alrededor de 25). Además, a esto se le suma que durante la etapa de entrenamiento el SVM dispone de un menor número de sesiones para aprender, lo que hace su tarea aún más complicada. Sin embargo, en general los resultados no decaen demasiado, lo cual dota de una cierta robustez a nuestro clasificador.

Si tenemos en cuenta únicamente los resultados obtenidos con el esquema 25 % - 75 %, el comportamiento de las diferentes características empleadas para la clasificación es bastante similar a aquel del esquema *leave-one-subject-out*. Los mejores resultados vienen dados por los coeficientes de forma y la boca, y con un intervalo de 300 fotogramas de longitud, aunque en este caso dan lugar a una exactitud de aproximadamente el 70 % (5 % de pérdida de exactitud). Los resultados de los ojos y las cejas siguen siendo pobres, reforzando así la teoría de que estas características no son discriminantes para distinguir entre las clases "deprimida" y "no deprimida".

En cuanto a los resultados usando combinaciones de características, en este caso la combinación de la boca con otras características no discriminantes, como los ojos y las cejas, no parece afectar tanto como en el caso de los experimentos con el esquema *leave-one-subject-out*, puesto que los resultados no difieren mucho de aquellos obtenidos empleado únicamente la boca.

En conclusión, el mejor resultado obtenido mediante un esquema de entrenamiento-validación 25 % - 75 % viene dado a partir del uso de los coeficientes de forma como características en la clasificación, en conjunto con una longitud de ventana en el nivel de grupo de 300 fotogramas.

Capítulo 7

Conclusiones y trabajo futuro

7.1. Conclusiones

El análisis automático de las expresiones faciales se ha convertido en una práctica cada vez más posible durante los últimos diez años, principalmente debido a los avances en los campos de la visión por computador y del aprendizaje automático. Pero nadie dijo que fuera fácil. Incluso abordar un problema en el que no es siquiera necesario detectar ni reconocer expresiones faciales específicas, sino únicamente extraer información sobre el comportamiento general de la cara, resulta ser un trabajo altamente exigente.

El principal objetivo de este proyecto era construir la primera aproximación de un clasificador capaz de determinar si una persona está deprimida o no. Aunque los resultados de la clasificación no superan una exactitud del 76 %, sí parecen ser muy prometedores. Para mejorarlos, cada paso es importante: escoger las muestras adecuadas para entrenar el clasificador, el proceso de alineamiento de los datos, el procedimiento de extracción de características, la elección del clasificador y el esquema de entrenamiento-validación.

Tratar con grandes cantidades de datos, como lo son las secuencias de vídeo de la base de datos Spectrum (de 20.000 fotogramas, en media) es siempre difícil. La solución adoptada en este proyecto pasa por el empleo de una técnica de segmentación en ventanas no solapadas, resultando aquellas de 10 segundos las que ofrecen mejores resultados. Sin embargo, dado que este enventanado supone una reducción de los datos, mucha información se pierde en el camino.

A la complejidad inherente de las tareas discutidas en los párrafos anteriores, se suma la dificultad de comprender y encontrar las características adecuadas que extraer de las secuencias de vídeo. Es en este contexto en el que la estrecha colabo-

ración con expertos psicólogos fue fundamental, puesto que ayudan a señalar cuáles son las partes específicas de la cara a las que merece la pena prestarles atención.

Quizás, lo que puede extraerse del trabajo recogido en esta memoria, es que no está para nada claro dónde reside la información relevante asociada a la depresión. Si un ser humano es capaz, en la mayoría de los casos, de determinar mediante un simple análisis visual si una persona está deprimida o no, ¿por qué no dar por sentado que una máquina puede ser entrenada para llevar a cabo esta misma tarea?

7.2. Trabajo futuro

El punto más fascinante de la investigación es que, aunque los temas se centran en cuestiones muy específicas, el trabajo sobre ellas puede llegar a ser infinito. Y más aún cuando se trata con una base de datos que sigue creciendo, como es el caso de la Spectrum. La siguiente lista enumera los datos disponibles actualmente o a corto plazo:

- Seguimiento de las características faciales de las secuencias de vídeo de los pacientes.
- Seguimiento de las características faciales de las secuencias de vídeo de los terapeutas.
- Secuencias de audio de las entrevistas.
- Transcripciones de las entrevistas.
- Segmentación temporal de las entrevistas basada en vídeo.
- Seguimiento de la mirada, tanto para el paciente como para el terapeuta.

Dando por hecho que toda esta información estará disponible algún día para el investigador, la base de datos Spectrum se convertiría probablemente en la más completa del campo de la depresión. Los siguientes puntos resumen parte del trabajo futuro que se puede desarrollar sobre ellos (para más detalles, consultar la versión íntegra de esta memoria).

- Aplicación de métodos de selección de características.
- Extracción de características asociadas al habla.

- Clasificación de casos leves de depresión.
- Análisis de la interacción paciente-terapeuta.
- Empleo de los ángulos asociados a los movimientos de cabeza como características para la clasificación.
- Estudio de la relación entre el habla y la depresión.
- Empleo de Unidades de Acción (*Action Units*) [8] como características para la clasificación [18].

Resultados preliminares de algunas de estas líneas de investigación, así como los resultados de clasificación presentados en esta memoria, han sido recogidos en el artículo llamado *Detecting depression from facial actions and vocal prosody*, aceptado en la International Conference on Affective Computing and Intelligent Interaction (ACII 2009) [11].

Bibliografía

[1] M. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan. Automatic recognition of facial actions in spontaneous expressions. *Journal of Multimedia*, 2006.

[2] C.M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

[3] Chih-Chung Chang and Chih-Jen Lin. *LIBSVM: a library for support vector machines*, 2001. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[4] T. Cootes and C. Taylor. Statistical models of appearance for computer vision, 1999.

[5] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *Lecture Notes in Computer Science*, 1407, 1998.

[6] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models-their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, January 1995.

[7] Martinez J. E., Ferraz M. B., Fontana A. M., and Atra E. Psychological aspects of brazilian women with fibromyalgia. *Journal of Psychosomatic Research*, 39(167-174):338–344, 1995.

[8] P. Ekman and W. V. Friesen. *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Consulting Psychologists Press, 1978.

[9] Paul Ekman. *Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage*. W. W. Norton & Company, 2001.

[10] Rita Moretti et al. Depression and alzheimer's disease: symptom or comorbidity? *American Journal of Alzheimer's Disease and Other Dementias*, 17(6):338–344, 2002.

[11] Cohn J. F., Simon Kreuz T., Yang Y., Nguyen M. H., Tejera Padilla M., and Zhou F. Detecting depression from facial actions and vocal prosody. *Affective Computing and Intelligent Interaction*, 2009.

[12] B. Fasel and J. Luettin. Automatic facial expression analysis: A survey. 2003.

[13] S. Khandelwal. *Conquering depression*. World Health Organization, 2001.

[14] Simon Lucey, Ahmed Bilal Ashraf, and Jeffrey Cohn. Investigating spontaneous facial action recognition through aam representations of the face. In K. Kurihara, editor, *Face Recognition Book*. Pro Literatur Verlag, Mammendorf, Germany, April 2007.

[15] T. M. Mitchell. The discipline of machine learning. Technical report, ML-CMU, 2006.

[16] M. Pantic and M.S. Bartlett. *Face Recognition*, chapter Machine Analysis of Facial Expressions, pages 377–416. I-Tech Education and Publishing, Vienna, Austria, 2007.

[17] Karen Schmidt, Zara Ambadar, Jeffrey Cohn, and Lawrence Ian Reed. Movement differences between deliberate and spontaneous facial expressions: Zygomaticus major action in smiling. *Journal of Nonverbal Behavior*, 2006. (In Press).

[18] Carnegie Mellon University and Uinversity of Pittsburgh. Automated face analysis project. http://www.cs.cmu.edu/∼face/.