UNIVERSIDAD AUTÓNOMA DE MADRID

ESCUELA POLITÉCNICA SUPERIOR

PROYECTO FIN DE CARRERA

ESTIMACIÓN AUTOMÁTICA DE EDAD A PARTIR DE IMÁGENES FACIALES

Ingeniería de Telecomunicación

Antonio Rodríguez Ruiz

FEBRERO 2012

ESTIMACIÓN AUTOMÁTICA DE EDAD A PARTIR DE IMÁGENES FACIALES

AUTOR: Antonio Rodríguez Ruiz

TUTOR: Manuel A. Sánchez-Montañés Isla

Grupo de Neurocomputación Biológica (GNB)

Dpto. de Ingeniería Informática

Escuela Politécnica Superior

Universidad Autónoma de Madrid

Febrero de 2012

Agradecimientos

En primer lugar, me gustaría dar las gracias a mi tutor Manuel Sánchez-Montañés no solamente por darme la oportunidad de realizar este trabajo, sino por la ayuda incansable y la cantidad de horas que ha estado conmigo para que este proyecto pudiera salir adelante, en ocasiones incluso a costa de su propia salud. Del mismo modo, también quiero agradecer a Luis Fernando Lago su ayuda y consejos para el desarrollo del trabajo y a Richard Rojas por su infinidad de respuestas a mi infinidad de preguntas y por no limitarse simplemente a ser un compañero de laboratorio sino también a ser un amigo. Tampoco me puedo olvidar del resto del personal de la EPS, tanto docente como personales de limpieza, administración y cafetería.

Como no, agradecer a mis padres la oportunidad de haber podido realizar la carrera, así como el haberme aguantado todos estos años. También agradecer a mis familiares su apoyo e interés constante, en especial a mis abuelos Ramona y Benito y a mi tía Rafi.

No me gustaría pasarme por alto el dar las gracias a mis dos amigos de toda la vida Jaime y Cheli, que siempre han permanecido a mi lado y nunca me han fallado.

Por último y no menos importantes, es más, seguramente los más importantes, quiero agradecer todo a aquellos que he conocido durante los años de la carrera. Tengo la fortuna de haber vivido maravillosos momentos con muchísimas personas que he conocido durante mis años en la universidad y justamente por el hecho de ser tantos, no me puedo permitir el lujo de nombrar a ninguno de ellos por temor a olvidarme de alguien. Pero cualquier persona con la que haya hecho un viaje de vacaciones, o salido a tomar unas cañas, o salido a comer o cenar, o de fiesta, o hacer un picnic o una barbacoa, o me haya sufrido en partidas de videoconsola, o me haya aplastado jugando al ping-pong o al fútbol, sabe que forma parte de estos agradecimientos y que no nombrarles no significa que me olvide de ellos.

A todos ellos de nuevo, ¡GRACIAS!

A Tamara,

por ser la mejor acompañante para este viaje.

Palabras Clave

Edad, Facial, Regresión, Clasificación, Automático, Imagen, Alineamiento, Recorte, Ecualización, MAE, Tasa de acierto, LOPO, AAM, SIFT, BIF

Resumen

El objetivo de este proyecto de fin de carrera es la definición y caracterización de la implementación de un sistema que realice estimación automática de edad a partir de imágenes faciales. Este trabajo se realiza a partir de la petición de una empresa especializada en visión artificial, con vistas a ser utilizada en un contexto de seguridad.

Tras un estudio del Estado del Arte, los sistemas primeramente implementados son los que mejores resultados brindan en la literatura relacionada, que son las técnicas de modelos activos de apariencia (AAM) y características de inspiración biológica (BIF). Posteriormente se realizan y validan una serie de modificaciones que proponemos a dichas técnicas. Dentro del proceso de validación del sistema, se propone el uso de un error de referencia básico, de tal forma que lo mínimo que exigimos a un sistema es que tenga un error menor que éste. Por otra parte, se propone un sistema de estimación de edad a partir de características invariantes a escala extraídas por SIFT. Esto en sí es una novedad ya que no se ha encontrado en la literatura el uso de estas características para estimar automáticamente la edad.

Los sistemas con mejores resultados en nuestras pruebas son los basados en BIF, ya que se obtiene un error absoluto medio (MAE) de 5.5 años. Para llegar a este resultado, se ha visto la necesidad de disponer una etapa de alineamiento preciso. También se pone de manifiesto la importancia de eliminar de la imagen todos los elementos distractores posibles. Los sistemas basados en SIFT llegan a alcanzar un MAE de 6.6, y los basados en AAM un MAE de 6.5. Pero por otra parte, los sistemas basados en SIFT no necesitan alineamiento, y en principio son invariantes a cambios como rotaciones y traslaciones, por lo que esta línea de investigación se debería seguir desarrollando.

La estimación de edad también se puede enfocar como un problema de clasificación en rangos de edades. En nuestros análisis concluimos que esta manera de enfocar el problema es más robusta (el error empeora menos en conjuntos de test muy diferentes a conjuntos de entrenamiento), y es menos sensible a la calidad del alineamiento.

Key Words

Age, Facial, Regression, Classification, Automatic, Image, Alignment, Cutout, Equalization, MAE, Hit Rate, LOPO, AAM, SIFT, BIF

Abstract

The aim of this work is to define and characterize the implementation of a system that automatically estimates a person's age from his facial image. This work is motivated by the request from a computer vision company which wants to use this system in a security context.

First a study of the State of the Art was carried, identifying those published systems with better reported results. These systems use active appearance models (AAMs) and biologically inspired features (BIFs). After implementing and validating them, we propose and analyze some extensions of these techniques. We also propose a system which predicts age from scale invariant features extracted by SIFT. This is a new strategy in this context, since we have not found in the literature the use of these features to automatically estimate the age of a person. On the other hand, we propose the use of a basic reference error as an important tool in the validation phase, requiring that any implemented system has an error smaller than this reference.

Systems with better results in our tests are based on BIF, with a mean absolute error (MAE) of 5.5 years. In order to achieve this error, an accurate alignment step is needed. It is also important to eliminate all the distractors in the image. SIFT based systems achieve MAEs of 6.6, whilst AAM-based systems achieve a MAE of 6.5.Since SIFT-based systems do not require alignment and are invariant to changes such as rotations and translations, we suggest to further develop this strategy in future work .

Age estimation can also be approached as a classification problem in different age ranges. From our analysis we conclude that this way of approaching the age estimation problem is more robust and less sensitive to the alignment quality.

Índice de contenido

1 Introducción 1
1.1 Motivación 1
1.2 Objetivos 1
1.3 Organización de la memoria 2
2 Estado del arte 3
2.1 Introducción 3
2.2 Preprocesamiento de las imágenes 4
2.3 Información en la que se basan los estimadores de edad 4
2.4 Reducción de dimensionalidad 8
2.5 Algoritmos de clasificación y regresión 8
2.6 Bases de datos 9
2.7 ¿Cómo evaluar el acierto de los sistemas que estiman automáticamente la edad? 10
2.8 Resultados publicados de los sistemas más relevantes encontrados en la literatura 11
2.9 Conclusiones 12
3 Diseño 13
3.1 Preprocesamiento de las imágenes 13
3.2 Extracción de características 17
3.3 Reducción de dimensionalidad 23
3.4 Predicción de la edad 24
3.5 Evaluación de la calidad del estimador 29
3.6 Resumen de las distintos sistemas que se implementan 32
4 Pruebas y resultados 33
4.1 Estimación del valor de la edad 33
4.2 Estimación del rango de edad 44
4.3 Pruebas con otras bases de datos 45
5 Conclusiones y trabajo futuro 49
5.1 Conclusiones 49
5.2 Trabajo futuro 55
Glosario de acrónimos 57
Referencias 59
Presupuesto I
Pliego de condiciones III

Índice de figuras

Figura 1: Evolución de la cara según una persona envejece 4

Figura 2: Puntos de referencia y vector que forma la unión de dichos puntos mediante triangulación de Delaunay (la circunferencia circunscrita de cada triángulo de la red no debe contener ningún vértice de otro triángulo) para AAM 5

Figura 3: Formación del vector de envejecimiento para el método AGES 6

Figura 4: Diagrama de bloques de la implementación un sistema estimador de edad con BIF ... 7

Figura 5: Ejemplo de formación de bloques de píxeles 8

Figura 6: Distribución de los puntos característicos e identificadores de cada uno de ellos para la base de datos FG-NET 9

Figura 7: Imágenes de ejemplo de la base de datos MORPH 10

Figura 8: Alineamiento de imágenes con respecto a los ojos 14

Figura 9: Ejemplo de deformación de una imagen (derecha) al realizar alineamiento al contorno facial, ojos, nariz y comisuras labiales de una imagen con diferente pose (centro) con respecto a la original (izquierda) 14

Figura 10: Ejemplo de recorte a partir de puntos etiquetados 15

Figura 11: Ejemplo de recorte con forma geométrica fija 16

Figura 12: Ejemplo de ecualización del histograma (abajo) de una imagen (arriba) 17

Figura 13: Puntos de referencia para obtener la información de la forma (izquierda) y áreas de extracción de información de textura (derecha) 18

Figura 14: Ilustración de extracción de características mediante BIF (arriba) y ejemplo de imágenes resultantes (abajo) 19

Figura 15: Ejemplos de filtros de Gabor (arriba) e imágenes filtradas (en verde, abajo) 20

Figura 16: Ejemplo de descriptores SIFT 22

Figura 17: Jerarquía de las redes neuronales 26

Figura 18: Ejemplo de SVM y la creación del hiperplano para separar clases 26

Figura 19: Ejemplo de ajuste de un conjunto de datos a una función cuadrática 27

Figura 20: Ejemplo de funcionamiento de técnica KNN con pesos. En la figura el número de vecinos más próximos elegido es 3 y cada vecino próximo tiene indicado su valor de edad correspondiente 29

Figura 21: Ejemplo de distribución de clases en una base de datos con muestras sesgadas (izquierda) y distribución de la misma base de datos tras la eliminación del sesgo (derecha) .. 32

Figura 22: MAE calculado mediante LOPO en los sistemas implementados usando AAM basados en el mejor trabajo de la literatura [21] variando la edad de división entre grupos.... 37

Figura 23: MAE calculado para la partición 802+200 en los sistemas implementados usando AAM variando la edad de división entre grupos 37

Figura 24: Selección de imagen de referencia (arriba) y alineamiento a sus puntos de referencia (abajo) 39

Figura 25: Resultados del MAE para el sistema implementado usando SIFT, variando el número de vecinos y la forma de pesar de éstos 40

Figura 26: Resultados del MAE para el sistema implementado usando SIFT, variando el número de vecinos y teniendo en cuenta las densidades de puntos y edades..... 41

Índice de tablas

Tabla 1: Resumen de los resultados más relevantes del estado del arte 11

Tabla 2: Parámetros para los filtros Gabor de la etapa simple 20

Tabla 3: Parámetros que definen las regiones para la etapa compleja 21

Tabla 4: Resumen de los distintos estimadores de edad que se implementan..... 32

Tabla 5: Resultados para los sistemas básicos. Las unidades de los errores son años 34

Tabla 6: Resultados para los sistemas implementados usando AAM basados en el mejor trabajo al respecto de la literatura [21]..... 35

Tabla 7: Resultados para los sistemas implementados usando AAM sin usar reducción de dimensionalidad 36

Tabla 8: Resultados para los sistemas implementados usando AAM con distintos estimadores de edad y reducción de dimensionalidad que el trabajo basado 38

Tabla 9: Resultados para los sistemas con imágenes alineadas de una manera más precisa 38

Tabla 10: Resultados del MAE para el sistema implementado usando SIFT, variando el número de vecinos y la forma de pesar de éstos 40

Tabla 11: Resultados del MAE para el sistema implementado usando SIFT, variando el número de vecinos y teniendo en cuenta las densidades de puntos y edades..... 41

Tabla 12: Resultados del MAE para el sistema implementado usando BIF, variando el número de orientaciones y bandas de los filtros Gabor con imágenes recortadas mediante elipse y alineadas por los ojos..... 42

Tabla 13: Resultados del MAE para el sistema implementado usando BIF, variando el número de orientaciones y bandas de los filtros Gabor con imágenes recortadas por el contorno facial de la imagen y alineadas a varios puntos..... 43

Tabla 14: Resultados de porcentaje de acierto al clasificar entre tres clases diferentes utilizando BIF con imágenes recortadas mediante elipse y alineadas por los ojos 44

Tabla 15: Resultados de porcentaje de acierto al clasificar entre tres clases diferentes utilizando BIF con imágenes recortadas por el contorno facial de la imagen y alineadas a varios puntos 45

Tabla 16: Resultados para los sistemas implementados usando BIF con la base de datos MORPH con la configuración de filtros óptima según FG-NET 46

Tabla 17: Resultados para los sistemas implementados usando BIF para estimación del valor de la edad usando una base de datos en entrenamiento diferente a la de test 46

Tabla 18: Resultados para los sistemas implementados usando BIF para estimación del valor de la edad usando una base de datos en entrenamiento diferente a la de test eliminado imágenes de una base de datos 47

Tabla 19: Resultados para los sistemas implementados usando BIF para clasificación usando una base de datos en entrenamiento diferente a la de test 47

1 Introducción

1.1 Motivación

La clasificación automática de imágenes faciales es un área de investigación que abarca disciplinas tales como el reconocimiento de patrones o el procesado de imagen. Los objetivos que persiguen dichos sistemas son múltiples, pudiendo variar desde la identificación en una misma imagen de características tan dispares como el género, edad o emociones de un sujeto o la verificación de que dos imágenes distintas corresponden a una misma persona [1].

La estimación de la edad de una persona a partir de una imagen es un problema interesante con multitud de posibles usos, desde mejorar el reconocimiento facial, desarrollar estudios demográficos o implementar aplicaciones que van desde el control de acceso, interacción hombre-máquina, la identificación de personas y la formación y organización de bases de datos [2].

Este proyecto de fin de carrera intenta dar solución a una empresa que necesita que se implemente un sistema capaz de determinar automáticamente la edad de un sujeto a partir de una imagen de su cara. Debido a las múltiples propuestas que se han realizado hasta el momento en este campo para intentar determinar automáticamente la edad de un individuo a partir de una imagen facial, es necesario realizar primero un estudio profundo del Estado del Arte y comparación de las propuestas más relevantes, analizar las fortalezas y debilidades de cada método, estudiar si son posibles de implementar bajo un ámbito determinado y encontrar la manera de poder mejorar los resultados ofrecidos por éstos.

1.2 Objetivos

El objetivo de este proyecto es realizar un estimador automático de edad que sea lo más robusto posible en condiciones reales. Tras presentar aquellos sistemas que mejores resultados ofrecen respecto a la estimación de edad mediante imágenes faciales encontrados en la literatura, se realizará una comparación entre ellos analizando distintos aspectos, tras lo cual se decidirá llevar a cabo la implementación de los más relevantes. A partir del análisis de los resultados obtenidos, se llevará a cabo la búsqueda de mejoras en los distintos métodos, bien modificando parámetros, metodología de trabajo o bloques del sistema, teniendo en cuenta que se buscan métodos que puedan funcionar de una manera automática, es decir, sin necesidad de contar con datos introducidos de manera manual.

Tras la realización de las diferentes implementaciones, se tiene como objetivo determinar la importancia de las distintas etapas formantes de los sistemas en el funcionamiento de los mismos. Se pretende determinar el efecto que tiene en el sistema el realizar o el no realizar un tipo de procesamiento previo de las imágenes o el hecho de afrontar el problema de la estimación de la edad como un problema de clasificación o como un problema de regresión.

1.3 Organización de la memoria

La memoria consta de los siguientes capítulos:

- **Capítulo 1**: Introducción: Motivación, objetivos y organización de la memoria.
- **Capítulo 2**: Estado del arte: Repaso a los trabajos previos en el ámbito de la estimación de la edad, citación de las posibles aplicaciones, resumen de las técnicas más relevantes implementadas en la literatura y la limitación de éstas.
- **Capítulo 3**: Diseño: Explicación de las diferentes etapas y métodos dentro de éstas que han sido implementados para los sistemas desarrollados.
- **Capítulo 4**: Pruebas y resultados: Exposición de los resultados obtenidos en las diferentes implementaciones realizadas.
- **Capítulo 5**: Conclusiones y trabajo futuro: A partir de las pruebas y resultados realizados, se exponen las conclusiones de las mismas y el posible trabajo futuro en el ámbito del trabajo.

2 Estado del arte

2.1 Introducción

La estimación de la edad de una persona es un problema de mucha relevancia en determinados contextos. Ejemplos importantes de posibles aplicaciones en este ámbito son:

- Ámbito forense, para conocer la edad de un sujeto involucrado en un crimen, realizar una reconstrucción post-mortem o gracias a la progresión de las características faciales con la edad, simular el aspecto de una persona desaparecida [3] o de un fugitivo.
- Control de seguridad y vigilancia, para evitar que ciertas personas que no cumplan el requisito de tener una edad mínima concreta, puedan acceder a cierto tipo de contenidos, bienes o lugares [4].
- Biometría, como información extra a la hora de identificar a una persona [5].
- Entretenimiento, como parte de algún juego o para predecir la edad de otros sujetos [6].

El caso de que la predicción de la edad sea realizada por una máquina cobra especial importancia por la dificultad que tiene el ser humano para estimar la edad en algunas situaciones. Por ejemplo, respecto a los casos de acceso a productos limitados por la edad del cliente, se llevaron a cabo unos estudios [7] a una serie de suministradores de alcohol en la que se les mostraron una serie de fotografías de personas de 16 años en las que tenían que determinar si tenían una edad igual o mayor que 18 años y así tener derecho a adquirir dicho producto. Los resultados de este estudio fueron que el 38% de los chicos y el 56% de las chicas pasaron como adultos cuando en realidad no lo eran. Investigaciones similares [8] determinaron que se tiende a sobreestimar la edad de las personas jóvenes en aproximadamente un tercio de los casos, por lo que los suministradores de alcohol se ven obligados a pedir la identificación en la mayoría de los casos a sus clientes.

Estos estudios plantean un reto para la investigación acerca de cómo determinar con exactitud la edad de alguien. La literatura sobre este tema es diversa y dispersa, yendo desde la comprensión de transformaciones matemáticas a la relación entre la edad percibida y características tales como el atractivo y bondad. Según la literatura las características faciales asociadas con la edad están relacionadas con el desplazamiento y crecimiento de los huesos, la deformación de la piel, la pérdida de fortaleza por parte de los músculos de la cara o la aparición de arrugas, verrugas, etc. según una persona envejece [9]. Varios estudios en el campo de la estimación automática de edad se basan en estos cambios físicos y/o biológicos que ocurren con el envejecimiento facial, proporcionando las características físicas diferenciadoras de la edad, mientras otros estudian además otros factores que influyen en estimaciones de la edad tales como las características de un cierto grupo (jóvenes, adultos, ancianos…) e incluso aspectos personales de una persona, como su estado civil, lugar de residencia o situación laboral [10].

Este capítulo se estructura según las diferentes etapas que componen los sistemas estimadores de edad: el preprocesamiento de las imágenes, las distintas técnicas para extraer la información en la que se basan los estimadores de edad, las diferentes técnicas de reducción de dimensionalidad, los distintos algoritmos utilizados, las bases de datos más

comunes y las distintas maneras de enfocar el problema (ya sea como regresión o como clasificación).

2.2 Preprocesamiento de las imágenes

Todos los trabajos encontrados en la literatura en primer lugar realizan un preprocesamiento de la imagen. Las distintas técnicas de esta etapa van desde la ecualización de la imagen, la normalización de todas las imágenes (por ejemplo, para que todas tengan un mismo tamaño), el alineamiento de determinadas partes de la cara hasta la eliminación del fondo de la imagen tras localizar qué es cara en la imagen (ver por ejemplo [11] [12] [13] [14]).

2.3 Información en la que se basan los estimadores de edad

El problema de la estimación de la edad es un problema complejo ya que el envejecimiento es algo incontrolable y propio de cada persona. Todos los sistemas para estimar la edad cuentan con dos etapas comunes que son la extracción de características apropiadas para estimar edad y las técnicas de estimación de la misma. Este apartado aborda las distintas técnicas de extracción de características que son más relevantes en este contexto.

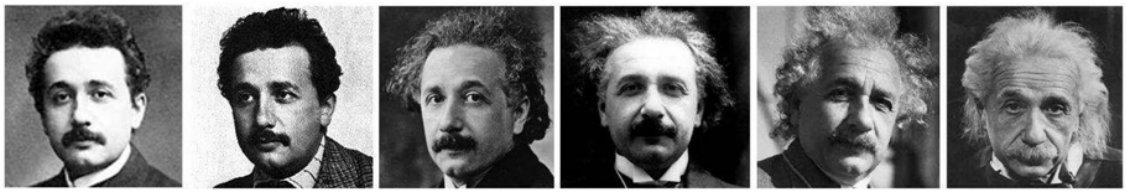

Figura 1: Evolución de la cara según una persona envejece

Los primeros trabajos que intentaban estimar la edad a partir de imágenes faciales [15] [16] se basaron en el estudio de la geometría a partir de la teoría del desarrollo cráneo-facial [17]. Dichos trabajos consistían en clasificar las imágenes en tres grupos (bebés, jóvenes adultos y adultos mayores), obteniendo 100% de acierto. La estimación de la edad basada en modelos antropométricos solamente puede diferenciar la forma de la cabeza hasta el periodo adulto, ya que a partir de una cierta edad adulta, la forma del cráneo no varía demasiado, por lo que ésta es una técnica que puede resultar útil para edades jóvenes pero no para adultos. Además cuenta con la limitación de poder trabajar solamente con imágenes faciales frontales y necesitar imágenes de alta resolución.

Debido a las limitaciones mencionadas, a partir de dichos trabajos surgieron nuevas líneas de investigación en el campo, con el desarrollo de distintas técnicas para la estimación de la edad basado en la exploración de características intrínsecas, tales como arrugas, y/u otras características extrínsecas, tales como el género, situación económica, nivel de vida, etc. A continuación se explican cualitativamente métodos relevantes para extraer la información relevante para estimar la edad, detallando los resultados más adelante.

2.3.1 Modelos activos de apariencia (AAM)

La técnica de los modelos activos de apariencia (*Active Appearance Models, AAM*) [18] construye un modelo general a partir de un conjunto de imágenes. Dicho modelo intenta explicar la variabilidad de la forma y textura en dicho conjunto. De esta forma, cada imagen se interpreta como una variación del modelo general. En AAM las imágenes son representadas

por un vector de datos en el que figuran las coordenadas de los puntos de referencia (por ejemplo, ojos, nariz, labios, etc.) y/o información local en esos puntos, como es la textura.

Los modelos activos de apariencia han sido usados en diferentes problemas de reconocimiento automático, como detección de expresiones faciales y estimación de la edad. Dado un conjunto de imágenes faciales y los correspondientes puntos de referencia, se genera un modelo estático de la variación de forma utilizando análisis de componentes principales (*Principal Component Analysis, PCA*). De esta manera, se construye un modelo estático de variación de textura, siendo este modelo lo único necesario para extraer el vector de características, a partir de la ecuación del conjunto de imágenes de entrenamiento:

$$X_i = \bar{X} + Px_i$$

donde X_i representa la forma de la imagen de entrenamiento, $\bar{X}$ es el valor medio de las imágenes, P es la matriz de autovectores generada y x_i es el vector de características, pudiendo despejarse en función del resto de elementos de la ecuación.

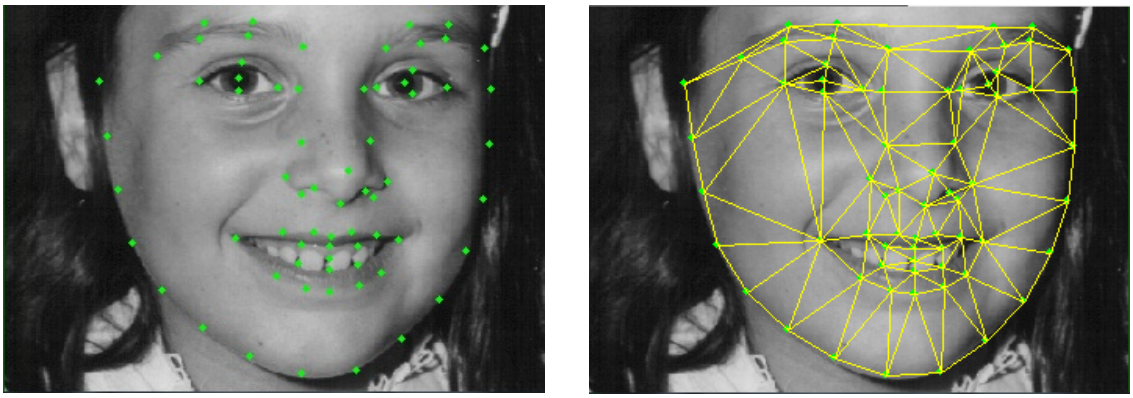

Figura 2: Puntos de referencia y vector que forma la unión de dichos puntos mediante triangulación de Delaunay (la circunferencia circunscrita de cada triángulo de la red no debe contener ningún vértice de otro triángulo) para AAM

La técnica de los modelos activos de apariencia ha sido utilizada en varias implementaciones de sistemas destinados a determinar la edad. Una de las primeras aproximaciones [19] se llevó a cabo sobre una base de datos que contenía 400 imágenes de individuos cuyas edades oscilaban entre 0 y 35 años. A partir de los parámetros obtenidos del AAM, se construía una función cuadrática para predecir la edad mediante regresión. Esta implementación, al ser una de las primeras que usaban esta técnica para estimar la edad, sugería posibles líneas de futuro a seguir para estos sistemas, tales como refinar la estimación con bloques intermedios más complejos o trabajar con bases de datos más heterogéneas en cuanto a edades y sujetos. Sistemas posteriores [14] usaban bases de datos más amplías y heterogéneas, tanto en el número de sujetos como en la variedad de edades de éstos, simulando así unas condiciones más realistas. Se proponían diferentes algoritmos, como el EM (*Expectation-Maximization*) para afrontar el problema de la regresión. Otros sistemas [12], tras tratar las características de las imágenes con diferentes algoritmos, aplicaban regresión mediante SVR mejorando levemente resultados de anteriores sistemas. Según la literatura, el sistema que mejor estima la edad a partir de características AAM [20] [21], realizaba una reducción de dimensionalidad tal que cada imagen con su correspondiente edad se representaba mediante un vector de una sola dimensión gracias a una técnica denominada

regresión espectral. Tras ello se procedía a dos bloques, uno en el que se clasificaban las imágenes entre jóvenes y adultos y uno último en el que se realizaba regresión tras la información del primero.

Las características extraídas mediante los modelos activos de apariencia, por lo general funcionan bien para estimar edad. Sin embargo, dichos sistemas necesitan contar con un etiquetamiento de una serie de puntos de la cara, algo que debe realizar un sistema capaz de localizar esa serie de puntos y que de esta manera liga el correcto funcionamiento del sistema a la correcta localización de estos puntos. Existen bases de datos con una serie de puntos de referencia etiquetados de manera manual que pueden utilizarse para formar modelos activos de apariencia.

2.3.2 Aging Pattern Subspace (AGES)

En el método AGES [13] la idea básica consiste en modelar el patrón de envejecimiento a partir de una secuencia de imágenes individuales de un mismo sujeto ordenadas en el tiempo con la intención de construir un subespacio representativo para cada edad. La información que maneja este método es la proporcionada por el método AAM. Sin embargo, se diferencia de éste en que trabaja con secuencias de imágenes de una misma persona. Mediante el algoritmo iterativo EM, puede sintetizarse las edades ausentes dentro de una secuencia de un mismo sujeto. El valor de las edades ausentes dentro del conjunto de imágenes de entrenamiento para un sujeto se obtiene a partir de la media del resto de imágenes que tienen esa misma edad de otros sujetos.

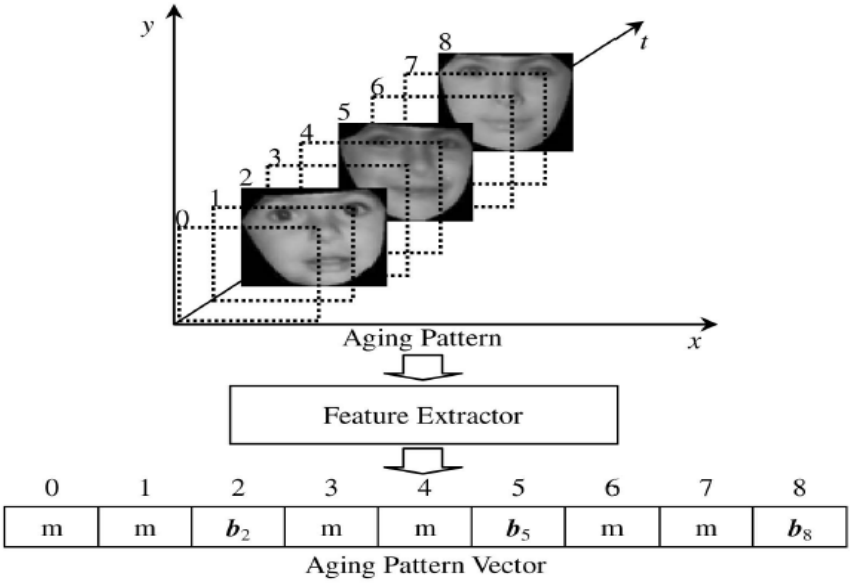

Figura 3: Formación del vector de envejecimiento para el método AGES

La técnica cuenta con dos etapas: aprendizaje de los patrones de envejecimiento y estimación de la edad. En la primera etapa, la técnica adoptada para construir un subespacio que capture la variación principal en el conjunto de datos de entrenamiento es PCA. Para realizar la proyección a dicho subespacio se procede del siguiente modo:

$$y = W^T(x - \mu)$$

donde μ es el valor medio de x, que es el vector que contiene las características de cada sujeto, es decir, el vector de envejecimiento, y W^T es la transpuesta de W, que se compone de

los autovectores de la matriz de covarianza de x. El vector de envejecimiento está mayormente incompleto, por lo que a partir de las características de envejecimiento, se utiliza el algoritmo EM para aprender un subespacio representativo.

Respecto a utilizar este método en situaciones realistas, cuenta con el inconveniente de tener que localizar los puntos de referencia para realizar el cálculo de AAM, ligándose de nuevo el funcionamiento de la técnica a la eficacia en la localización de puntos. Además esta técnica supone que todas las personas tienen las mismas características faciales para una determinada edad, lo cual no es cierto ya que el proceso de envejecimiento es algo individual y dependiente de varios factores externos.

2.3.3 Características de inspiración biológica (BIF)

La idea básica de las características de inspiración biológica (*Biologically Inspired Features, BIF*) [22] se deriva del modelo de respuesta a estímulos visuales de los primates en el reconocimiento visual de objetos. Las imágenes son tratadas por dos bloques llamados *capa simple* y *capa compleja*, que extraen las características que definen un objeto.

En el uso encontrado de BIF dentro del contexto de estimación de edades [11] la primera etapa, llamada capa *simple (S)*, consiste en una serie de filtros de Gabor para filtrar la imagen. En la segunda etapa, llamada capa *compleja (C)*, se realiza la operación máximo o desviación estándar sobre los resultados del bloque anterior. Tras este último bloque, cada imagen queda definida por su vector de características BIF.

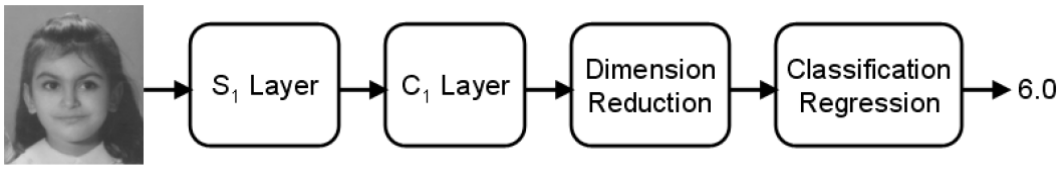

Figura 4: Diagrama de bloques de la implementación un sistema estimador de edad con BIF

Las imágenes en los sistemas que usan BIF suelen tener un procesamiento previo mediante una serie de recortes para realizar una selección de la parte de la imagen de la cual vaya a obtenerse la información, ya que puede ocurrir que se extraiga información que actúe como ruido de todo aquello de la imagen que no sea cara, como cuadros, paisaje, muebles, etc. También las imágenes son escaladas a un mismo tamaño para buscar alineamiento de partes de la cara.

Se debe tener en cuenta que debido a que las etapas que componen BIF cuentan con varios parámetros a fijar (por ejemplo el número de filtros u orientaciones de Gabor a utilizar en la etapa *simple,* o el número de dimensiones finales que se usan tras realizar PCA), los resultados finales estarán muy ligados a la selección de estos parámetros. Generalmente, según la literatura [23] [11] la estimación de edad usando las características de inspiración biológica es de las técnicas que obtienen mejores resultados.

2.3.4 Patch-kernel

Otra estrategia diferente consiste en ver una cara como un conjunto de trocitos ("*patches*") diferentes [24]. Lo importante no es ahora la relación entre diferentes *patches*, sino cómo se distribuyen estadísticamente. El primer paso en esta estrategia consiste en obtener un modelo estadístico general de cómo se distribuyen dichos *patches* para cualquier cara de la base de datos. Normalmente se usa para esto una mezcla de gaussianas. Una vez obtenido este modelo general, se reajusta para cada cara particular, con lo que el vector de características de cada cara representa la desviación respecto al modelo general. Finalmente se construye un estimador basado en *kernel-regression* usando estas características.

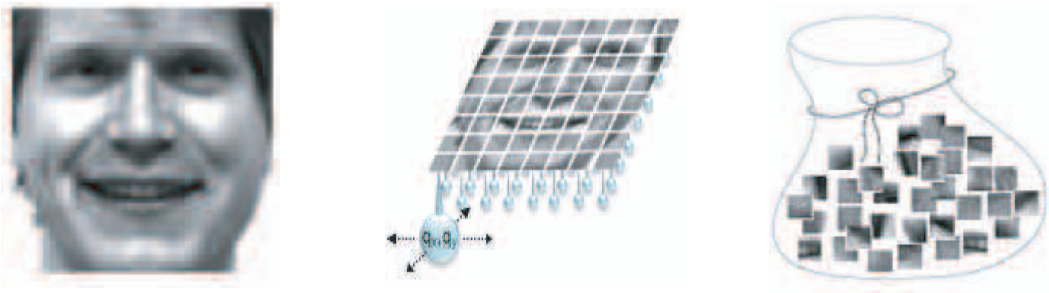

Figura 5: Ejemplo de formación de bloques de píxeles

Una de las grandes ventajas de esta técnica es que al trabajar directamente con trozos extraídos de la imagen, es invariante frente a rotaciones, traslaciones o cambios de escalas de la misma.

2.4 Reducción de dimensionalidad

Las características extraídas de una cara suelen ser vectores de alta dimensionalidad, por lo que en la mayoría de ocasiones es necesario aplicar una reducción de dimensionalidad para así eliminar información poco relevante de los vectores de características que puede llegar a actuar como distractor y así hacer más eficiente el funcionamiento de los sistemas [11] [13] [25] [21].

En la mayoría de los sistemas, la reducción de dimensionalidad se realiza mediante análisis de componentes principales (Principal Component Analysis, PCA) [26]. Intuitivamente esta técnica sirve para hallar las causas de la variabilidad de un conjunto de datos y ordenarlas por importancia. Técnicamente se busca la proyección que capture la mayor parte de la varianza de los datos. Otra técnica muy usada para reducir la dimensionalidad basada en proyecciones es el análisis discriminante lineal de Fisher [27].

Otra técnica [21] para reducir la dimensionalidad es la regresión espectral (*Spectral Regression, SR*) [28]. Esta técnica afronta el problema de reducir la dimensionalidad para problemas de aprendizaje en el marco de la regresión, lo que evita la descomposición en matrices de autovectores densas. A partir de las características discriminantes de los datos se puede descubrir una estructura discriminante intrínseca de éstos.

2.5 Algoritmos de clasificación y regresión

El problema de estimar la edad se puede tratar como un problema de regresión (predecir la edad concreta de un individuo) o desde el punto de vista de la clasificación

(estimar la edad de un sujeto dentro de un posible rango de edades o considerar cada edad como una clase).

Los algoritmos de clasificación y regresión más comunes usados en la literatura son las máquinas de vectores de soporte (*Support Vector Machines, SVM*) y regresión por vectores de soporte (*Support Vector Regression, SVR*) [29] [30], que son un conjunto de algoritmos de aprendizaje. Otros sistemas [24] utilizan otras técnicas para estimar la edad con la información del vector de características, como el denominado *Kernel Regression* [31] o para realizar operaciones lineales sencillas [16] para determinar la edad. Otro algoritmo habitual para clasificar son las redes neuronales [12] que son un paradigma de aprendizaje y procesamiento automático inspirado en la forma en que funciona el sistema nervioso de los animales.

2.6 Bases de datos

La calidad de la colección de imágenes faciales es crucial en el diseño de sistemas para la estimación de la edad. Lo deseable es que las bases de datos sean heterogéneas, es decir, que cuenten con variedad de imágenes de diferentes personas con distintas edades, en diferentes condiciones (iluminación, pose, calidad de imagen…), etc. Confeccionar una base de datos así es costoso, especialmente por el hecho de contar con varias imágenes de una misma persona a lo largo de su vida. Seguidamente se detallan las bases de datos más comúnmente utilizadas en la literatura para evaluar los sistemas de predicción automática de edad.

2.6.1 FG-NET

Una de las bases de datos más utilizadas es FG-NET [32]. Ésta cuenta con 1002 imágenes de 82 sujetos con edades entre los 0 y 68 años en distintas poses, condiciones de iluminación o expresión. Cada sujeto, cuenta con varias imágenes de él mismo con diferentes edades. Cada imagen está etiquetada con 68 puntos faciales (*landmarks*).

Figura 6: Distribución de los puntos característicos e identificadores de cada uno de ellos para la base de datos FG-NET

Sin embargo, esta base de datos cuenta con una carencia a la hora de distribuir las edades de sus muestras, ya que existe un gran sesgo entre edades, habiendo gran número de imágenes de jóvenes pero pocas de adultos o ancianos [21].

2.6.2 MORPH

La base de datos MORPH [33] es de carácter público y fue realizada por la Universidad de Carolina del Norte. Las imágenes están distribuidas en dos álbumes, y cuentan con la información de la edad, género y raza. El rango de edades con el que cuenta MORPH es de 15 a 68 años.

Esta base de datos está equilibrada respecto a la distribución de la edad. Sin embargo, no cuenta con muestras de sujetos con edades menores a 15 años.

Figura 7: Imágenes de ejemplo de la base de datos MORPH

2.6.3 Yamaha Aging Database (YGA)

La base de datos Yamaha Aging Database (YGA), es una base de datos comercial que cuenta con 8000 imágenes de alta resolución de 1600 sujetos asiáticos [27], siendo la mitad de ellos varones y la otra mitad mujeres. El rango de edades que abarca es desde 0 a 93 años y los sujetos están bajo distintas condiciones de iluminación, expresión o maquillaje.

Generalmente ofrece buenos resultados en los sistemas que la utilizan para la estimación de la edad debido a la carencia de imágenes de personas que no sean de raza asiática oriental, por lo que es una base de datos muy homogénea, cuando lo interesante para evaluar la robustez de los sistemas de estimación de edad es que las bases de datos sean lo más heterogéneas posible.

2.7 ¿Cómo evaluar el acierto de los sistemas que estiman automáticamente la edad?

Debido a las características del problema de la estimación de la edad, éste puede evaluarse de dos maneras: como un problema de clasificación, en el que cada edad es una clase, o como un problema de regresión.

Mientras que los sistemas que afrontan el problema como clasificación de la edad dentro de un rango de edades se pueden evaluar simplemente mediante la tasa de casos bien clasificados, la manera más común de evaluar el funcionamiento en los sistemas que llevan a cabo una estimación del valor de la edad es mediante el cálculo del error medio absoluto (*Mean Absolute Error, MAE*). El MAE permite hacerse una idea del funcionamiento general del sistema y es definido como la media del error absoluto entre las etiquetas estimadas (edad o clase) y las verdaderas etiquetas:

$$MAE = \frac{\sum_{i=1}^{N_i} |\hat{z}_i - z_i|}{N_t}$$

donde $\hat{z}_i$ es la etiqueta (edad) estimada para la muestra i, z_i es la etiqueta (edad) verdadera y N_t es el número total de muestras.

Para los sistemas que estiman la edad, el valor de MAE se expresa en años, por lo que con dicho valor se puede hacer una idea de cómo de bueno es un sistema, siendo mejores aquellos sistemas que obtengan menor MAE. El resultado ideal sería que el MAE fuese nulo, sin embargo esto no es posible ya que los sistemas se modelan con los datos de varias personas, siendo el proceso de envejecimiento individual para cada una y dependiente de múltiples factores externos como el clima, lugar de residencia u horas de exposición solar [25].

2.8 Resultados publicados de los sistemas más relevantes encontrados en la literatura

En apartados anteriores se habló sobre los distintos bloques que se pueden encontrar en los sistemas estimadores de la edad más relevantes. En la siguiente tabla, por cada sistema se hace referencia al MAE obtenido si se realiza estimación del valor de la edad o la tasa de acierto si se realiza clasificación de la edad en un rango, las bases de datos utilizadas, la información de la imagen en la que se basan, la técnica de reducción de dimensionalidad (si es aplicada) y el algoritmo de estimación utilizado.

Información usada	Reducción de dimensionalidad	Algoritmo	Base de datos	MAE (regre.)	% Acierto (clasif.)	Referencia
AAM	-	Función cuadrática	No conocida	5.1	-	[19]
AAM	-	SVR	FG-NET	5.3	-	[14]
AAM	-	SVR	YGA	7.0	-	[14]
AAM	PCA	SVR	FG-NET	5.1	-	[12]
AAM	SR	SVM+SVR	FG-NET	4.3	-	[21]
AGES	PCA	Propio	FG-NET	6.7	-	[13]
AGES	PCA	Propio	MORPH	8.8	-	[34]
AGES	PCA	Propio	FG-NET	-	40.92 (14 grupos)	[25]
BIF	PCA	SVR	FG-NET	4.8	-	[11]
BIF	PCA	SVR	YGA	3.9	-	[11]
BIF	Fisher	SVM	YGA	2.6	-	[27]
Geometría de la cabeza	-	Propio	No conocida	-	100 (3 grupos)	[16]
Patch-kernel	-	Kernel regression	FG-NET	5.0	-	[24]
Patch-kernel	-	Kernel regression	YGA	4.9	-	[24]

Tabla 1: Resumen de los resultados más relevantes del estado del arte

Respecto a los sistemas que trabajan con AAM, en su primera aproximación [19] los resultados fueron buenos debido a la base de datos en unas condiciones poco realistas al ser las imágenes de alta resolución y de un entorno controlado. Los posteriores sistemas [14] [12] [21] tuvieron en cuenta este aspecto y realizaron simulaciones más realistas con el uso de

imágenes en distintas condiciones tanto de resolución como de poses de los sujetos, por lo que los resultados de éstos son más relevantes en la literatura.

Referente a los sistemas que usan AGES [13] [25], éstos obtienen MAE peores que otros sistemas. Sin embargo utilizan la información no solamente para hacer estimación del valor de la edad sino también para clasificar dentro de un rango, algo en lo que no se centran la mayoría de los sistemas.

En los sistemas basados en BIF [11] [27], los resultados obtenidos son generalmente buenos, usando varios clasificadores diferentes como SVM o SVR, mientras que las técnicas que patch-kernel o geometría de la cara [16] [24] también ofrecen buenos resultados por lo general, pero no son muy utilizados en la literatura.

2.9 Conclusiones

Basándose en lo relatado en este capítulo, se puede concluir que el buen funcionamiento de los sistemas estimadores de edad está fuertemente ligado a la correcta extracción de características determinantes de la edad a partir de la imagen, tales como la forma de cabeza o la localización de arrugas.

Existen distintas técnicas que obtienen las características intrínsecas de la edad, aunque el hecho de poder aplicar cada técnica puede depender también de otra serie de sistemas que sean capaces de localizar una serie de puntos de referencia para realizar su implementación, por ejemplo AAM que tiene la necesidad de localizar una serie de puntos de referencia faciales para poder llevarse a cabo.

Este trabajo se afronta desde el punto de vista de llevar a cabo una implementación real de determinación de la edad de una persona, por lo que a partir de lo estudiado en la literatura, se decide descartar en primer lugar la implementación de sistemas basados en geometría de la cabeza (primeros trabajos en estimación de edad) debido a que estos sistemas solamente pueden predecir la edad de alguien hasta llegar a una edad adulta y solamente trabaja con imágenes frontales. Los otros dos sistemas que se descartan son el basado en AGES y en trozos de la imagen debido a que los resultados obtenidos en la literatura son peores que los que se obtienen con las otras dos técnicas en las que se deciden basar los diseños que se realizarán, que son las técnicas de modelos activos de apariencia con clasificador SVM y SVR [21] y la técnica de extracción de BIF [11]. Tras las implementaciones, si los resultados obtenidos son satisfactorios se podrá comprobar cómo funcionan dichos sistemas en otros ámbitos, como la clasificación o con otras bases de datos, de cara a entender la robustez que dichos sistemas presentan y posibles líneas de trabajo futuro respecto a ellos.

3 Diseño

En el capítulo anterior se abordaban los sistemas más relevantes encontrados en la literatura, analizándolos usando una misma estructura de bloques. Esto permitía ver más claramente las diferencias fundamentales entre dichos sistemas.

En este capítulo se explican en detalle las implementaciones realizadas basadas en dichos sistemas y otras propuestas nuevas que se realizan. La herramienta que se usa en la implementación de cada bloque es *Matlab*®.

A continuación se detallan las diferentes alternativas que se han implementado y validado en cada una de las diferentes etapas que componen el sistema: preprocesamiento de la imagen, extracción de características, reducción de dimensionalidad, estimadores de edad y la forma de evaluar la calidad del estimador.

3.1 Preprocesamiento de las imágenes

La etapa de preprocesamiento es fundamental en sistemas de análisis automático de imágenes [35], sin embargo éste no es un paso trivial. Por ejemplo es importante evitar que durante el tratamiento de la imagen se conserve/añada información no necesaria o se elimine información relevante en el funcionamiento del sistema. A continuación se detallan las diferentes formas de procesamiento previo que se han utilizado.

3.1.1 Alineamiento

El alineamiento consiste en hacer que todas las imágenes tengan la misma posición con respecto a una imagen de referencia en uno o varios lugares de la cara, tales como los ojos, la nariz, el contorno de la cara, etc. Es deseable que la imagen de referencia disponga de unos puntos de referencia bien diferenciados para que el alineamiento sea más sencillo de realizar. Estos puntos de referencia pueden ser seleccionados de manera manual, escogiendo el usuario los puntos a alinear sobre cada imagen, o contando con imágenes que previamente dispongan de una serie de puntos etiquetados, o de manera automática a partir de sistemas que sean capaces de localizar ciertos puntos relevantes en la cara.

Existen distintas alineaciones que pueden llevarse a cabo aplicando distintas transformaciones a las imágenes, bien sea solamente girándolas, trasladándolas y/o escalándolas, o bien sea deformando ligeramente la imagen.

3.1.1.1 Transformaciones rígidas

Llamamos transformaciones rígidas a aquellas que mantienen la forma original de la imagen y por tanto no la deforman. Estas transformaciones son los cambios de escala de una imagen, las traslaciones y/o las rotaciones.

Se puede comprobar claramente en el ejemplo de la figura 8 cómo las imágenes han sido giradas y escaladas para conseguir el alineamiento con respecto a los ojos. Uno de los aspectos negativos que se comprueban con estas transformaciones y que es común a todas cuando se realiza alineamiento son la aparición de regiones sin imagen (trozos negros de las imágenes alineadas) que al ser independiente el giro para cada imagen, puede empeorar el

funcionamiento del sistema, y la posible pérdida de información al escalar las imágenes ya que se está alterando la resolución original de éstas.

Figura 8: Alineamiento de imágenes con respecto a los ojos

3.1.1.2 Transformaciones no rígidas

Llamamos transformaciones no rígidas a aquellas que pueden llegar a deformar la imagen original para realizar alineamiento. Estas deformaciones se pueden realizar cuando se busca alinear con precisión a más de dos puntos.

Las imágenes que sufren más deformación a la hora de realizar esta serie de transformaciones no rígidas son aquellas que pretenden realizar la alineación a una imagen con distinta pose a la suya, por ejemplo, cuando se intenta alinear una imagen de una persona con la cara girada a un lado a una serie de puntos determinados por la pose de una persona que está con la cara orientada hacia cualquier dirección diferente.

Figura 9: Ejemplo de deformación de una imagen (derecha) al realizar alineamiento al contorno facial, ojos, nariz y comisuras labiales de una imagen con diferente pose (centro) con respecto a la original (izquierda)

Esta técnica de alineamiento no rígido puede obtener buenas alineaciones si las imágenes tienen poses similares pero ofrece serias deformaciones en los casos donde esto no ocurra, como puede observarse en la figura 9, pudiendo afectar gravemente al correcto funcionamiento del sistema.

3.1.1.3 Limitaciones en el alineamiento

Como se mencionó anteriormente, para realizar alineaciones es necesario contar con unos puntos de referencia a los cuales alinear. El alineamiento depende totalmente de la detección precisa de los puntos, por lo que cuantos más puntos se deseen alinear, más fuentes de error va a haber en dicho proceso. Alinear con transformaciones no rígidas permite obtener alineamientos mucho más precisos, pero se necesitan usar más puntos de referencia. En la práctica es deseable trabajar con técnicas capaces de localizar automáticamente puntos de referencia de la cara que sean fáciles de detectar (por ejemplo los ojos).Por otra parte muchas veces es preferible hacer alineamientos más sencillos (por ejemplo transformaciones rígidas) ya que necesitan menos puntos.

3.1.2 Recorte

Las imágenes con las que trabajan los sistemas que estiman edad, suele contener además de la cara elementos como el fondo de la imagen u otras partes del cuerpo del individuo. Estos elementos pueden constituir información no relevante a la imagen y que ésta haga que el funcionamiento de los sistemas estimadores de edad empeore. Por lo tanto, realizar un recorte de la imagen para eliminar la mayor parte de estos elementos distractores puede resultar de gran beneficio para el sistema.

Si el sistema dispone o puede extraer los puntos del contorno de la cara, se pueden usar estos para recortar la región de la imagen donde se ubica la cara, aunque puede perderse información relevante de la imagen para estimar la edad si el recorte no es realizado en la totalidad de la cara, como ocurre con el recorte observado en la siguiente figura donde no se incluye información referente al pelo o la frente.

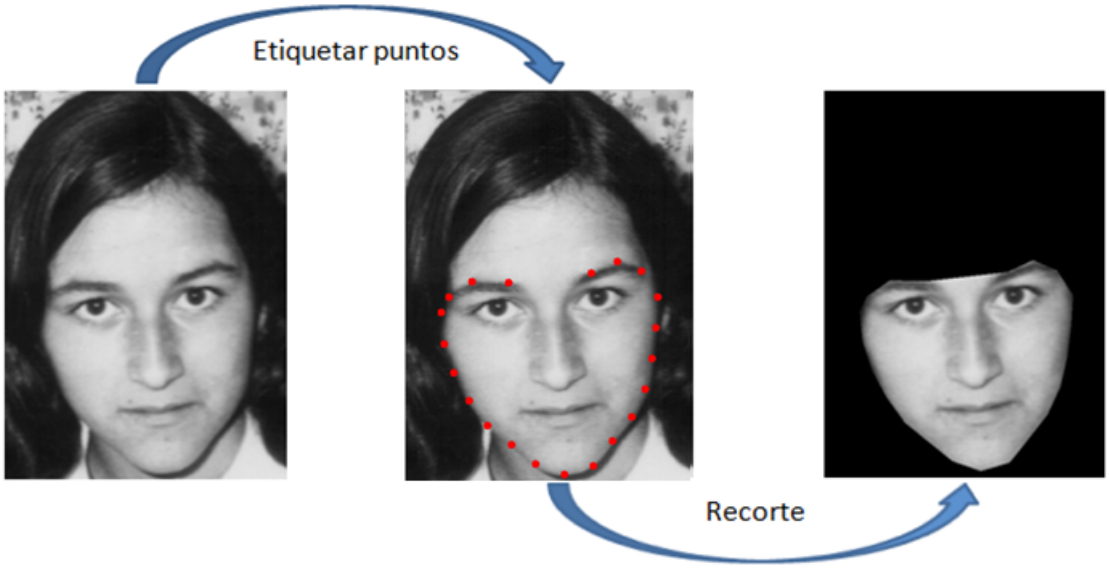

Figura 10: Ejemplo de recorte a partir de puntos etiquetados

Otro tipo de recorte más sencillo que el anterior consiste en realizar un recorte con una determinada forma geométrica fija después de localizar una parte relevante de la cara, por ejemplo realizar un recorte con forma de elipse a partir de la localización de los ojos o la nariz. Este tipo de recorte es más sencillo debido a que no necesita localizar tantos puntos. Sin embargo el uso de este tipo de recorte tiene el riesgo de que se recorta "a ciegas" ya que se aplica siempre el mismo recorte a cada imagen y puede que se elimine información necesaria para la estimación de la edad. Además este recorte puede que no elimine la suficiente información no necesaria de la imagen, dejando partes distractoras sin recortar para los sistemas.

Figura 11: Ejemplo de recorte con forma geométrica fija

3.1.3 Ecualización

Dentro del conjunto de imágenes que se utilizan para entrenar el sistema existen diferencias no sólo de pose o el número de elementos distractores entre diferentes imágenes, sino también en las condiciones de iluminación. Ésta es otra circunstancia que puede influir negativamente en el funcionamiento de sistema, ya que pueden existir detalles de una imagen que debido a la iluminación no se tengan en cuenta en el mismo grado de importancia que ese mismo detalle en otras imágenes con distinta iluminación. Para evitar este efecto, se puede ecualizar el histograma de cada imagen para intentar que existan el mismo número de píxeles por cada nivel de gris del histograma de una imagen monocroma.

El problema que conlleva realizar la ecualización de una imagen es que se ecualiza toda la imagen, es decir, también se tiene en cuenta elementos distractores como el fondo o la ropa, por lo que puede que la ecualización de la parte de la imagen en la que aparece la cara no sea tan eficiente para los sistemas de estimación de edad como podrían ser. Para ecualizar solamente la información relevante a la cara se puede realizar recortes como los explicados en el apartado anterior y posteriormente ecualizar solamente la información relevante de la imagen.

En la siguiente figura se puede observar un ejemplo claro de que tras realizar la ecualización de una imagen, el histograma de la misma pasa de tener una cantidad de píxeles en una única región de la escala de grises a tenerlos en toda la gama de la escala de grises.

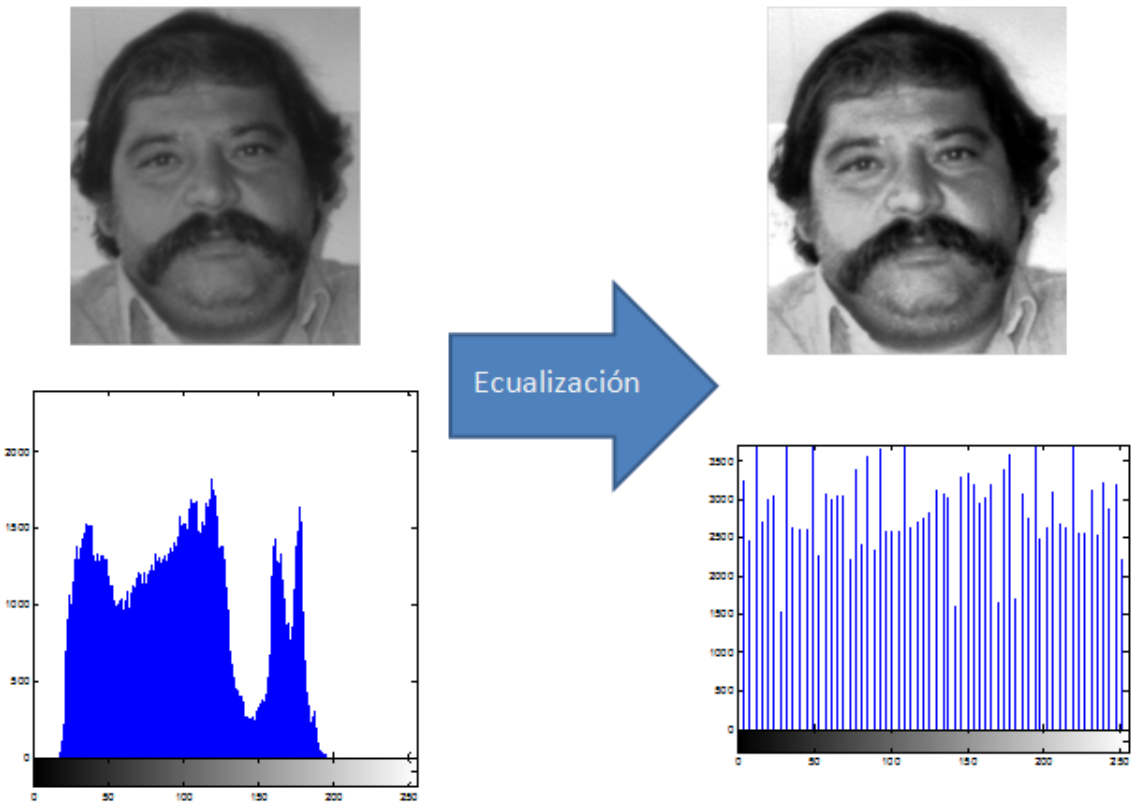

Figura 12: Ejemplo de ecualización del histograma (abajo) de una imagen (arriba)

3.2 Extracción de características

Según lo analizado en el capítulo *Estado del Arte*, la etapa crítica en los sistemas estimadores de la edad es la extracción de características complejas de las imágenes. El comportamiento de esta etapa está fuertemente ligado al procesamiento previo de las imágenes, por lo que la mejor opción es extraer estas características después de haber eliminado información irrelevante en la imagen.

En este apartado se explican tres técnicas diferentes para la extracción de características: dos de ellas son las que mejores resultados ofrecen en la literatura, que son los modelos activos de apariencia (AAM) y las características de inspiración biológica (BIF), y una tercera técnica como propuesta de extracción de características propias de la edad que no ha sido encontrada en la literatura para la estimación de la edad, que es el método *Scale-Invariant Features Transform* (SIFT).

3.2.1 Modelos activos de apariencia (AAM)

Los modelos activos de apariencia [18] son una de las técnicas más utilizadas para extraer las características de una imagen facial. Se basa en capturar la variabilidad de la forma y la textura de un conjunto de datos. Este método se basa en los modelos activos de forma (ASM) [36], donde se utiliza el análisis de componentes principales (PCA) para explicar los datos (caras en los sistemas que estima edad) en base a un número reducido de componentes.

Para crear este modelo estadístico se necesita un conjunto de imágenes con una serie de puntos localizados y etiquetados Estos puntos describen la forma del objeto. El modelo para capturar la información de la forma se puede construir como:

$$x = \bar{x} + \Phi_s b_s$$

donde las nuevas formas x son sintetizadas al deformar la forma media $\bar{x}$ usando combinaciones lineales de los autovectores de la matriz de covarianza Φ_s, siendo b_s el vector de características de cada imagen. De manera equivalente se define el modelo para capturar la información de la textura:

$$g = \bar{g} + \Phi_g b_g$$

La forma y la textura de cada imagen queda definida por b_s y b_g, siendo la manera de combinar ambas informaciones:

$$b = \begin{pmatrix} W_s b_s \\ b_g \end{pmatrix} = \begin{pmatrix} W_s \Phi_s^T (x - \bar{x}) \\ \Phi_g^T (g - \bar{g}) \end{pmatrix} = \Phi_c c;\ \Phi_c = \begin{pmatrix} \Phi_{cs} \\ \Phi_{cg} \end{pmatrix}$$

donde c es el denominado vector de apariencia que contiene la información de textura y de forma y W_s es una matriz de peso, siendo ajustable sus parámetros.

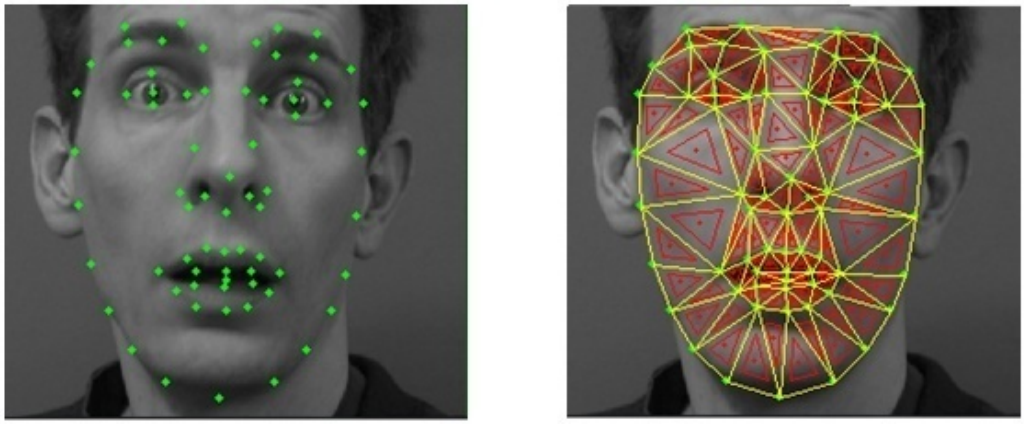

Figura 13: Puntos de referencia para obtener la información de la forma (izquierda) y áreas de extracción de información de textura (derecha)

La precisión de este método depende del número de vectores propios que se conservan tras la reducción de dimensionalidad, siendo este valor determinado por la varianza explicada seleccionada tras la reducción de dimensionalidad mediante análisis de componentes principales.

3.2.2 Características de inspiración biológica (BIF)

Las características de inspiración biológica (*Biologically Inspired Features, BIF*) [22] se inspiran en la respuesta a estímulos visuales de las células de la corteza visual de primates. Las imágenes son tratadas por dos bloques llamados *simple* y *complejo* que extraen las características que definen un objeto. Estas características son extraídas a partir de un banco de filtros de Gabor y operaciones matemáticas relativamente sencillas sobre las imágenes filtradas. De estas operaciones se obtiene un vector de características.

A continuación, se muestra una figura que ilustra el funcionamiento de la técnica de manera esquemática. En este ejemplo la etapa *simple tiene* ocho tamaños distintos de filtros de Gabor y cuatro orientaciones (representados por cada color diferente) y se observa cómo la etapa *compleja* trabaja con la información de cada banda (dos filtros de Gabor consecutivos y de diferente tamaño) definida en la etapa anterior.

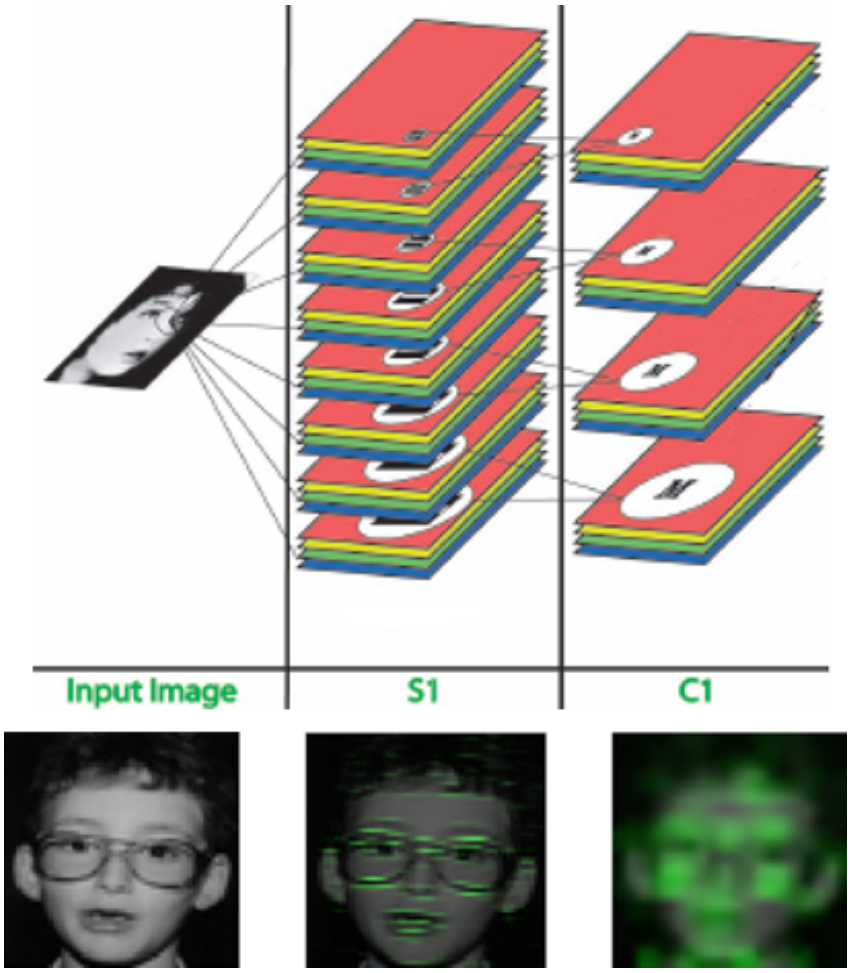

Figura 14: Ilustración de extracción de características mediante BIF (arriba) y ejemplo de imágenes resultantes (abajo)

3.2.2.1 Etapa simple

La finalidad de esta primera etapa es crear un modelo del comportamiento de las células simples de la corteza visual de mamíferos. La manera en la que esto se consigue es tratando cada imagen con un banco de filtros de Gabor, donde estos filtros se definen como:

$$G(x,y) = \exp\left(-\frac{X^2 + \gamma^2 Y^2}{2\sigma^2}\right)\cos\left(\frac{2\pi}{\lambda}X\right)$$

donde $X = \cos\theta + y\sin\theta$ y $Y = -x\sin\theta + y\cos\theta$ son las rotaciones de los filtros de Gabor con un ángulo θ entre 0 y π, γ es el cociente de aspecto espacial que especifica la elipticidad del filtro (se suele fijar a un valor de 0.3), σ es el ancho efectivo y λ es la longitud de onda.

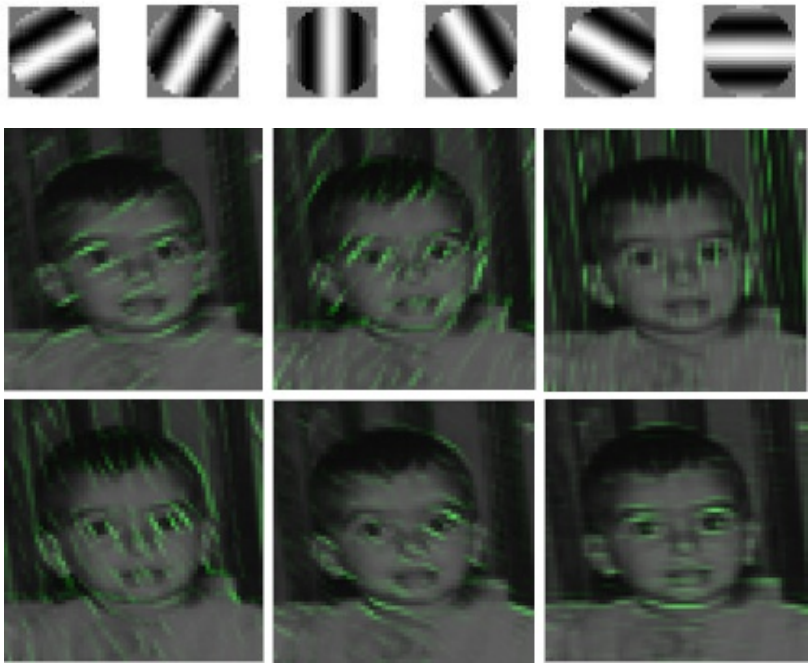

Figura 15: Ejemplos de filtros de Gabor (arriba) e imágenes filtradas (en verde, abajo)

En este método, el banco de filtros utilizado varía tanto en tamaño como en orientación. Por cada conjunto de dos filtros consecutivos en tamaño, se define una banda. A la hora de aplicar los filtros de Gabor a las imágenes, un parámetro que queda a elección del usuario es el número de bandas y de orientaciones de los filtros a utilizar. Esto es debido a que el número de orientaciones y bandas óptimo depende del conjunto de datos con el que se trabaje. Seguidamente se adjunta una tabla con los valores de los filtros para la etapa *simple*:

	Etapa *simple*		
	Tamaño filtro (píxel)	σ	λ
Banda 1	5x5	2.0	2.5
	7x7	2.8	3.5
Banda 2	9x9	3.6	4.6
	11x11	4.5	5.6
Banda 3	13x13	5.4	6.8
	15x15	6.3	7.9
Banda 4	17x17	7.3	9.1
	19x19	8.2	10.3
Banda 5	21x21	9.2	11.5
	23x23	10.2	12.7
Banda 6	25x25	11.3	14.1
	27x27	12.3	15.4
Banda 7	29x29	13.4	16.8
	31x31	14.6	18.2
Banda 8	33x33	15.8	19.7
	35x35	17.0	21.2

Tabla 2: Parámetros para los filtros Gabor de la etapa *simple*

3.2.2.2 Etapa compleja

En esta etapa, los resultados obtenidos de la etapa *simple* se ven sometidos a una serie de operaciones que tienen como resultado un único vector. Por cada banda y orientación, la operación que se procede a realizar por regiones es la siguiente: cada región se somete al cálculo de su desviación estándar, $std = \sqrt{\frac{1}{N_s x N_s}\sum_{i=1}^{N_s x N_s}(F_i - \bar{F})^2}$; que revela la variabilidad de las salidas de las células simples dentro de la vecindad N_s x N_s, donde F_i es el valor máximo de los dos resultados de la etapa *simple* obtenidos por los dos filtros de esa banda para el píxel i: F_i = max(x_i^j, x_i^{j+1}); donde x_i^j y x_i^{j+1} son los valores filtrados de la etapa anterior con escalas j y j + 1, $\bar{F}$ es el valor medio de los valores filtrados de la vecindad N_s x N_s.

Las regiones utilizadas para el cálculo de la desviación estándar son variadas, cambiando el tamaño de cada región y el solapamiento que existe entre ellas. Al igual que para la etapa anterior, en esta etapa *compleja* se definen una serie de bandas que se corresponden con las definidas en dicha etapa anterior y que posteriormente permiten cierta flexibilidad a la hora de decidir qué información utilizar para la estimación de la edad. Las regiones para realizar el cálculo de desviación estándar se definen como se adjunta en la siguiente tabla:

	Etapa *compleja*	
	Tamaño filtro (píxel)	**Solapamiento (píxel)**
Banda 1	6x6	3
Banda 2	8x8	4
Banda 3	10x10	5
Banda 4	12x12	6
Banda 5	14x14	7
Banda 6	16x16	8
Banda 7	18x18	9
Banda 8	20x20	10

Tabla 3: Parámetros que definen las regiones para la etapa *compleja*

3.2.3 Transformada de características invariantes a escala (SIFT)

El algoritmo *Scale-invariant feature transform* (SIFT) [37] se utiliza para encontrar y describir puntos característicos en imágenes. Ha sido usado en diferentes campos como el reconocimiento de objetos, video tracking, modelado 3D o reconocimiento de gestos [38], pero no se han encontrado trabajos donde se use en estimadores de edad. Este método toma una imagen y la transforma en un amplio conjunto de vectores de características locales. Cada uno de estos vectores es invariante a escala, rotación o traslación de la imagen. Estos vectores se obtienen a partir de una serie de pasos que se detallan en los siguientes apartados.

Paso 1: Detección de máximos y mínimos de la escala

En esta etapa se identifica la posición y la escala de los puntos candidatos a ser características invariantes. El espacio de la escala se define por $L(x, y, \sigma)$, y se obtiene tras convolucionar la función Gaussiana $G(x, y, \sigma)$ con la imagen $I(x, y)$. Para que la detección de los puntos de interés sea rápida y eficaz, el algoritmo utiliza la diferencia de Gaussianas convolucionada con la imagen. Dicha diferencia puede ser calculada a partir de las diferencias

entre dos escalas vecinas separadas por un factor constante k. La función de detección se define como:

$$D(x, y, \sigma) = \big(G(x, y, k\sigma) - G(x, y, \sigma)\big) * I(x, y) = L(x, y, k\sigma) - L(x, y, \sigma)$$

donde L es una imagen suavizada con un filtro de Gauss.

Paso 2: Localización de puntos de interés

En esta etapa se pretende eliminar los puntos localizados que no sean relevantes. A cada punto se le aplica un modelo para determinar su localización y escala. Para detectar los máximos y mínimos locales de $D(x, y, \sigma)$ cada punto se compara con sus ocho píxeles vecinos en la imagen actual y con los nueve píxeles vecinos de las imágenes superior e inferior. De este modo se seleccionan únicamente los puntos más estables.

Paso 3: Asignación de la magnitud del gradiente y de la orientación

Esta etapa tiene como objetivo asignar una orientación y una magnitud del gradiente a cada punto de interés localizado. Para cada imagen $L(x, y)$ con determinada escala, se calcula la magnitud del gradiente $m(x, y)$ y la orientación $\Theta(x, y)$ mediante diferencias entre píxeles:

$$m(x, y) = \sqrt{\big(L(x + 1, y) - L(x - 1, y)\big)^2 + \big(L(x, y + 1) - L(x, y - 1)\big)^2}$$

$$\Theta(x, y) = \tan^{-1} \frac{L(x, y + 1) - L(x, y - 1)}{L(x + 1, y) - L(x - 1, y)}$$

Paso 4: Descriptor del punto de interés

Para conseguir el descriptor del punto, en primer lugar se obtienen las magnitudes y orientaciones de los vecinos del punto de interés, utilizando la imagen suavizada más cercana a la escala calculada en el segundo paso. Una vez obtenidos, para asegurar la invariancia a la orientación, las coordenadas del descriptor y las orientaciones del gradiente se rotan de forma relativa a la orientación extraída antes.

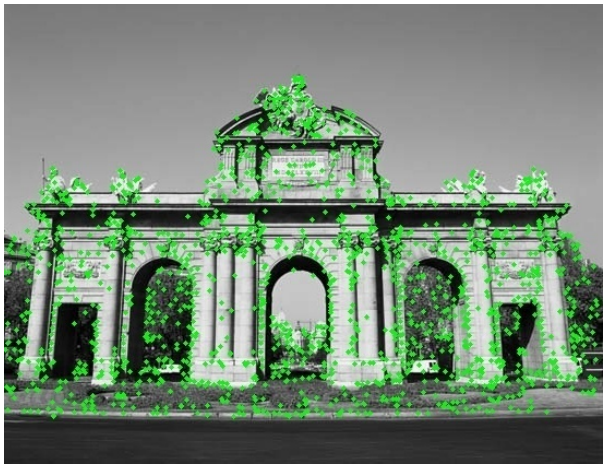

Figura 16: Ejemplo de descriptores SIFT

3.3 Reducción de dimensionalidad

Una vez que se extraen las características, las imágenes suelen estar representadas por vectores de alta dimensionalidad. Estos vectores de alta dimensionalidad pueden contener información poco relevante y/o altamente redundante, además de ocasionar que los sistemas que los utilicen sean más lentos en su funcionamiento, por lo que la mayoría de las ocasiones es interesante aplicar una reducción de dimensionalidad que mantenga la información más relevante de los vectores y así mejore en líneas generales el funcionamiento del sistema al poder ser más eficiente.

En este apartado se detallan los tres métodos diferentes implementados para realizar reducción de dimensionalidad.

3.3.1 Regresión Espectral (SR)

La regresión espectral (*Spectral Regression, SR*) [28] tiene como objetivo principal el lograr una reducción de dimensionalidad de un conjunto de vectores a una única dimensión. Se basa fundamentalmente en combinar técnicas de regresión y análisis espectral de grafos. Es una técnica que puede ser supervisada o no, lo que da una idea de su flexibilidad.

Esta técnica construye primero un grafo de afinidad a partir de una serie de muestras etiquetadas y no etiquetadas para encontrar una estructura intrínseca discriminante y obtener una serie de respuestas a un conjunto de datos dado. A partir de estas respuestas, se aplica una regresión ordinaria para así poder crear una función de aprendizaje.

Dado un conjunto de entrenamiento con N muestras etiquetadas $x_1, x_2, \dots, x_N$ y un conjunto de test con $(M - N)$ muestras sin etiquetar $x_{N+1}, x_{N+2}, \dots, x_M$ con $x_i \in \Re^n$ y x_i perteneciente a una de c clases, siendo n_k el número de muestras etiquetadas en la clase k $(\sum_{k=1}^{c} n_k = N)$, el algoritmo para realizar SR se resume como sigue:

<u>Paso 1: Construcción del grafo de adyacencia G</u>

El grafo G es construido para modelar la estructura local y la información de las etiquetas a través de tres pasos: (a) conectar los nodos i a los nodos j si x_i es de los p vecinos más próximos de x_j y viceversa, (b) conectar los nodos i a los nodos j si x_i es de la misma clase que x_j y (c) eliminar las uniones de los nodos i y j si x_i y x_j no son de la misma clase.

<u>Paso 2: Construcción de la matriz de pesos W</u>

W es una matriz de dispersión simétrica de MxM que se construye como sigue: (a) $W_{ij} = 0$ si no hay nodos entre i y j, (b) $W_{ij} = 1/n_k$ si x_i y x_j son de la misma clase, los demás $W_{ij} = \delta\, s(i,j)$, donde δ $(0 < \delta \leq 1)$ es un peso dado y $s(i,j)$ es una función de evaluación de semejanza entre x_i y x_j interpretada mediante la función del calor.

<u>Paso 3: Generación de las respuestas</u>

Del problema $Wy = \lambda Dy$ se obtienen los c autovectores $y_0, y_1, \dots, y_{c-1}$, donde D es la matriz diagonal de MxM donde el valor de la diagonal (i,i) es la suma de la columna o fila i de W.

Paso 4: Mínimos cuadrados con regularización

$v_k = arg_v \min (\sum_i^M (v^T x_i - v)^2 + \alpha \|v\|^2)$ donde y_i^k es el elemento i de y_k. Se calcula los $c-1$ autovectores $v_1, v_2, \dots, v_{c-1} \in \Re^n$. v_k $(k \in [1, c-1])$ es la solución del problema de regularización de mínimos cuadrados.

Paso 5: SR-embedding

Siendo $V = [v_1, v_2, \dots, v_{c-1}]$ una matriz de transformación de n x $(c-1)$, los componentes x_i de dimensión nx1 son introducidos en el subespacio para extraer los vectores de menor dimensionalidad z_j

$$x \rightarrow z = V^T x$$

3.3.2 Análisis de componentes principales (PCA)

El análisis de componentes principales (*Principal Component Analysis, PCA*) [26] es una técnica estándar para reducir la dimensionalidad de un conjunto de datos. Intuitivamente la técnica sirve para hallar las causas de la variabilidad de un conjunto de datos y ordenarlas por importancia. El funcionamiento técnico de PCA se basa en buscar la proyección según la cual los datos queden mejor representados en términos de mínimos cuadrados. Los nuevos componentes principales o factores serán una combinación lineal de las variables originales, y además serán linealmente independientes entre sí.

PCA construye una transformación lineal que elige un nuevo sistema de coordenadas en el cual la dirección de mayor varianza del conjunto original de datos coincide con el primer eje de coordenadas, la segunda mayor varianza coincide con el segundo eje, y así sucesivamente. Cuando se usa PCA como técnica de reducción de dimensionalidad, se desechan los ejes que tienen asociados varianzas menores. Una forma estándar de elegir cuántos ejes son desechados es fijar el tanto por ciento de varianza original que se desea capturar y quedarse con el conjunto más pequeño de ejes que la explican. Para ello se van seleccionando los ejes de mayor a menor varianza hasta que la suma de sus varianzas asociadas llegue a la varianza total que se quiere capturar.

Sean n vectores de m dimensiones, PCA pretende encontrar un número de factores $p < m$ que explican aproximadamente el valor de las m variables de cada sujeto. Se define F_j como el conjunto de muestras. A partir de los datos, puede construirse la matriz de correlación muestral $\mathcal{R} = [r_{ij}] \in M_{mxn}$ donde $r_{ij} = \frac{cov(F_i, F_j)}{\sqrt{var(F_i)}\sqrt{var(F_j)}}$; y puesto que la matriz de correlación es simétrica, es diagonalizable y se verifica que $\sum_{i=1}^m \lambda_i = 1$, siendo λ_i sus valores propios, por lo que gracias a esta propiedad, los factores principales más relevantes identificados matemáticamente se representan por la base de autovectores de la matriz $\mathcal{R}$.

3.4 Predicción de la edad

En los sistemas implementados se parte la base de datos en conjunto de entrenamiento y conjunto de test. En un primer lugar, el sistema se ajusta ("entrena") usando los datos del conjunto de entrenamiento, tanto sus características como el valor de la variable a predecir ("aprendizaje supervisado"). En segundo lugar, el sistema ya entrenado se usa para

predecir la edad en los ejemplos de test, obteniendo de esta forma una estimación del error que tendrá el sistema en ejemplos nuevos.

En este apartado se explican los diferentes algoritmos para los estimadores de edad utilizados en las implementaciones. Se hace una división entre aquellos que se utilizan para estimar el valor concreto de la edad y aquellos que se usan para estimar la edad dentro de un rango.

3.4.1 Estimación del valor concreto de la edad

La estimación del valor concreto de la edad, puede enfocarse como un problema de clasificación, (cada edad es una clase, con lo que hay tantas clases como edades posibles) o como un problema de regresión, en la que los resultados son valores enteros o reales. En este punto lo normal es preguntarse cuál es el mejor método para afrontar dicho problema. Para responder a esta cuestión, se han hecho estudios del comportamiento de sistemas estimadores de la edad con un clasificador representativo como máquinas de vectores de soporte (*Support Vector Machine, SVM*) y con un regresor representativo como la regresión por vectores de soporte (*Support Vector Regression, SVR*) [12]. La conclusión general que se obtiene es que ninguno de los dos presenta unos resultados generalmente mejores que el otro, sino que su comportamiento varía con los datos con los que se trabaja, siendo mejor uno donde el otro no lo es y viceversa. En otros estudios [39] se sugiere que los mejores resultados se obtienen de la fusión de ambos métodos, tras lo cual se puede conseguir un híbrido que se aproveche de las fortalezas de cada uno.

El objetivo de este apartado es explicar en detalle los distintos estimadores que se han considerado en las implementaciones.

3.4.1.1 Por clasificación

En primer lugar, se explican una serie de técnicas que son utilizadas cuando el problema de la estimación del valor concreto de la edad se afronta como un problema de clasificación, es decir, cada edad es una clase, con lo que hay tantas clases como edades posibles.

3.4.1.1.1 *Distancia a medias*

Con este método se pretende clasificar las muestras que se encuentran en test en función de la distancia que tengan al valor medio de las muestras medias que haya en el conjunto de entrenamiento.

Por cada clase de las muestras que se encuentran en entrenamiento, se calcula su media obteniendo así un vector por cada clase que resulta ser el valor medio de cada clase, por lo que así se tienen únicamente tantos vectores como clases haya. Para asignar una clase a una nueva muestra, se calcula su distancia (generalmente la distancia euclídea) a cada vector medio y aquel que resulte más cercano a la nueva muestra será el que defina la clase del vector que se está testeando.

3.4.1.1.2 *Redes neuronales (ANN)*

Las redes neuronales (*Artificial Neurals Networks, ANNs*) son un tipo de algoritmo de aprendizaje automático que permite construir modelos de clasificación/regresión usando como modelos estructuras inspiradas en el cerebro. En el tipo de red neuronal más conocido, que es el que se utiliza en este trabajo (red *feedforward* entrenada mediante *backpropagation*) [40], la red neuronal consta de una capa de entrada, una o varias capas ocultas y otra capa de salida. Dichas capas están formadas por unidades sencillas llamadas neuronas, que realizan una operación matemática simple. En este tipo de redes neuronales, los atributos del patrón a predecir se introducen en la capa de entrada (con tantas neuronas como atributos haya), y la predicción de la red está en el/los valor/es de las neuronas de la capa de salida. La red está organizada para transmitir señales desde las neuronas de entrada a las de salida a través de las ocultas.

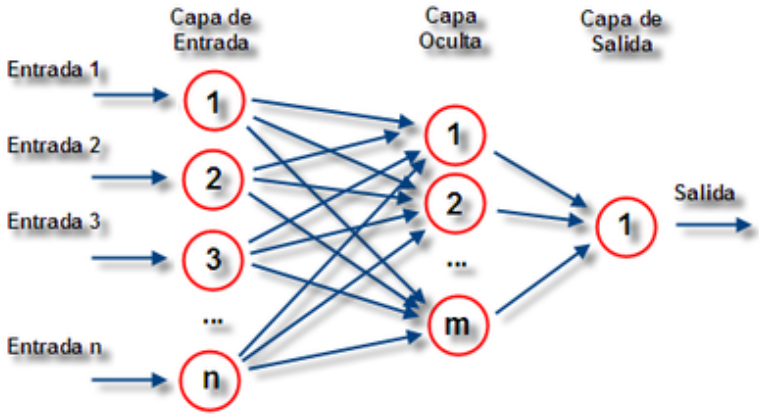

Figura 17: Jerarquía de las redes neuronales

La salida de la red neuronal depende de los valores de las diferentes conexiones entre las neuronas. De esta manera, en la fase de entrenamiento se ajustan los valores de tal forma que la salida de la red para dichos ejemplos sea cercana a la deseada. La técnica estándar para realizar esta búsqueda en redes *feedforward* es la conocida como *backpropagation*.

3.4.1.1.3 *Máquinas de vectores de soporte (SVM)*

Intuitivamente, SVM [30] es una técnica que traslada el espacio de atributos de partida a un espacio de mayor dimensionalidad, donde se intentan separar las clases por medio de hiperplanos, intentando que el espacio que las separa sea lo más amplio posible.

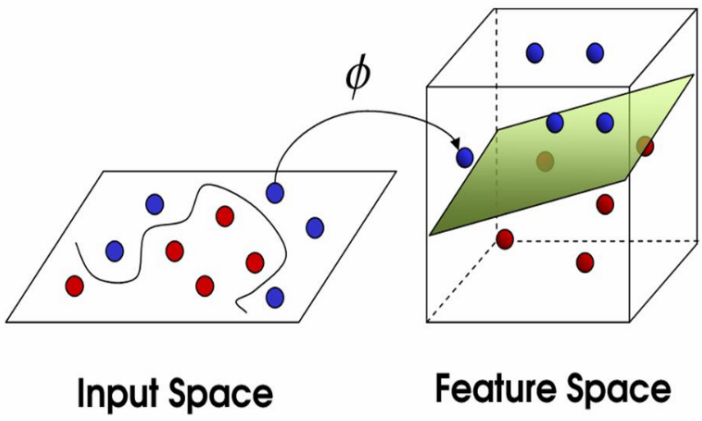

Figura 18: Ejemplo de SVM y la creación del hiperplano para separar clases

Desde un punto de vista matemático, en el caso de un clasificador binario, el problema es el siguiente: Dado N puntos de entrenamiento (x_1, y_1), (x_2, y_2),..., (x_N, y_N) con $x_i \in \Re^n$ y además $y_i \in \{-1,1\}, i = 1, \dots, N$ y suponiendo que estos puntos son linealmente independientes, se tiene que encontrar el conjunto de N_s vectores de soporte s_i $(N_s \leq N)$, los coeficientes de pesos α_i, la constante b y la superficie de decisión lineal tales que la distancia de los vectores de soporte es máxima en $wx + b = 0$; donde $w = \sum_{i=1}^{N_s} \alpha_i y_i s_i$.

3.4.1.2 Por regresión

En segundo lugar, se explican las técnicas usadas cuando el problema de estimación del valor concreto de la edad se afronta como un problema de regresión debido a que las distintas edades son una serie de valores consecutivos.

3.4.1.2.1 *Ajuste lineal por mínimos cuadrados (MvRegress, MVR)*

El ajuste lineal por mínimos cuadrados es una técnica en la que se modela la relación entre pares $\{x, y\}$ mediante una función lineal $\hat{y} = wx + b$. Los parámetros libres de dicha función lineal (w, b) se escogen de tal forma que la diferencia cuadrática media entre y e $\hat{y}$ sea mínima.

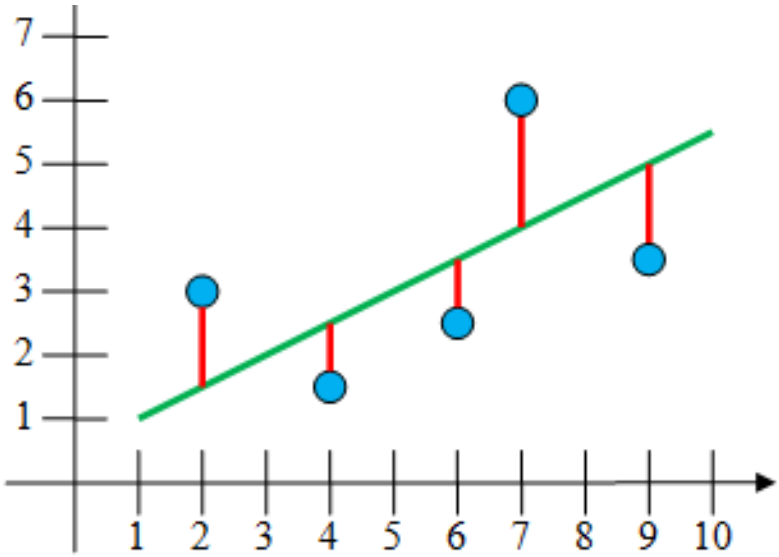

Figura 19: Ejemplo de ajuste de un conjunto de datos a una función cuadrática

3.4.1.2.2 *Función classify*

Esta es una función incluida en la herramienta de desarrollo *Matlab*®. Classify ajusta una densidad normal multivariable a cada posible valor de la variable a estimar. En los sistemas de edad, hay tantas clases como diferentes valores de edad en entrenamiento. A cada ejemplo de test se le asigna la clase (edad) cuya función de densidad sea más cercana.

3.4.1.2.3 *Regresión con vectores de soporte (SVR)*

La idea básica de la regresión por vectores de soporte (*Support Vector Regression, SVR*) consiste en realizar un mapeo de los datos de entrenamiento $x \in X$, a un espacio de mayor dimensión a través de un mapeo no lineal, donde podemos realizar una regresión lineal. Dado N puntos de entrenamiento (x_1, y_1), (x_2, y_2),..., (x_N, y_N) con $x_i \in \Re^n$ y además $y_i \in \{-1,1\}, i = 1, \dots, N$; se tiene que construir un hiperplano y obtener valores de w y b. El

hiperplano w es seleccionado con una pequeña normalización, mientras simultáneamente se minimiza la suma de las distancias de estos puntos de entrenamiento al hiperplano, medido por la función de pérdida de intensidad de ε [29].

$$|y_i - (wx_i + \mathrm{b})|_\varepsilon = \begin{cases} 0;\ si\ |y_i - (wx_i + \mathrm{b})| \leq \varepsilon \\ |y_i - (wx_i + \mathrm{b})| - \varepsilon;\ en\ cualquier\ otro\ caso \end{cases}$$

donde el valor de ε es seleccionado por el usuario, y la compensación para encontrar un hiperplano con una buena regresión es controlado a partir del parámetro C, teniendo así:

$$min_{w,b,\xi,\xi^*}\ \frac{1}{2}\ \|w\|^2 + C \sum_{i=1}^{N_s} (\xi_i + \xi_i{}^*)$$

$$y_i - (wx_i + \mathrm{b}) \leq \varepsilon + \xi_\mathrm{i}\ con\ \xi_\mathrm{i} \geq 0$$

$$-y_i + (wx_i + \mathrm{b}) \leq \varepsilon + \xi_\mathrm{i}{}^*\ con\ \xi_\mathrm{i}{}^* \geq 0$$

3.4.1.2.4 *K-Vecinos más próximos (KNN) con pesos*

El método de clasificación de k-vecinos próximos (*K-Nearest Neighbors, KNN*), [41] es un algoritmo de aprendizaje supervisado en el que dado unos vectores de entrenamiento en un espacio multidimensional, para un nuevo vector se calcula la distancia entre los vectores de entrenamiento y aquél y se seleccionan los k más cercanos. La medida de distancia más utilizada en este método es la distancia euclídea:

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^{p} (x_{i_r} - x_{j_r})^2}$$

donde los valores de los atributos del i-esimo ejemplo (donde $1 \leq i \leq n$) se representan por el vector p-dimensional $x_i = (x_{i_1}, x_{i_2}, \dots, x_{i_p}) \in X$

Para la estimación de edad, se propone una técnica nueva a partir del método de K-Vecinos próximos que consiste en pesar los k vecinos más próximos de diferentes maneras. Esta técnica se utiliza sobre la información dada por puntos característicos localizados por SIFT. A partir de la localización de los k vecinos más próximos de un punto característico, se realiza el cálculo de la media, mediana o moda de entre las clases de los k vecinos más próximos para asignar un valor de edad a ese punto característico. Para el caso de utilizar la moda, en caso de no haber ningún valor que sea la moda dentro de un conjunto de valores, se elige el de menor valor. Cuando todos los puntos tienen asignado un valor, se realiza la media ponderada de todos ellos y se asigna el valor obtenido como valor de la edad de la imagen. La siguiente figura ilustra un ejemplo de cómo funciona esta técnica en el caso de que se localizaran tres puntos característicos en una imagen de test.

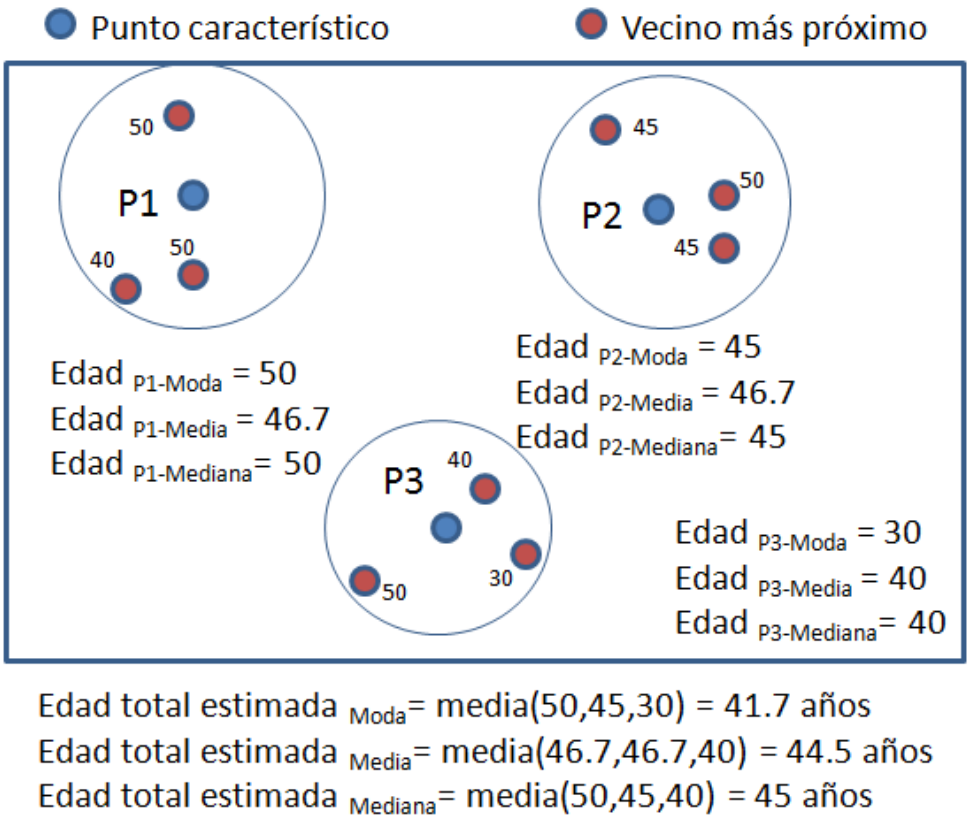

Figura 20: Ejemplo de funcionamiento de técnica KNN con pesos. En la figura el número de vecinos más próximos elegido es 3 y cada vecino próximo tiene indicado su valor de edad correspondiente

3.4.2 Estimación del rango en el que está la edad

Existe otra manera de afrontar el problema de estimación de la edad que es estimando el rango de edades al cual pertenece una edad. Este es un problema de clasificación y como tal, los distintos algoritmos utilizados para ello son los explicados en el apartado de estimación del valor de la edad referentes a estimar este valor mediante clasificación. Estos algoritmos son la distancia a medias, las redes neuronales y SVM.

3.5 Evaluación de la calidad del estimador

Existen diferentes aspectos a tener en cuenta para evaluar el funcionamiento de un estimador de edad. La forma en la que se afronta el problema (clasificación frente a regresión), las distintas formas de realizar los conjuntos de entrenamiento y de test o las bases de datos utilizadas y su composición están estrechamente ligadas para determinar la calidad de los sistemas implementados. En este apartado se detallan estos tres diferentes campos y su posible influencia en el comportamiento del estimador.

3.5.1 MAE frente a Error de clasificación

Como se vio en el capítulo *Estado del Arte* en el apartado sobre cómo evaluar los sistemas estimadores de edad, el error medio absoluto (MAE) es la manera estándar de expresar los resultados para los sistemas que realizan estimación del valor de la edad y el error de clasificación se usa para los que realicen clasificación en rangos de edades.

3.5.1.1 Estimación del valor concreto de la edad

En los sistemas que determinan el valor de la edad lo deseable es tener un valor de MAE lo más cercano a cero como sea posible. Sin embargo, se sabe que esto no es posible ya que el proceso de envejecimiento es algo individual de cada persona e incontrolable.

El valor de MAE que obtienen los sistemas, además de estar ligado a la calidad de técnica usada para estimar la edad, también está muy relacionado con la base de datos utilizada en la implementación. Si ésta es poco heterogénea en la distribución de las edades, es decir, las edades están muy concentradas en torno a un valor, al entrenar y testear con esa base de datos, lo normal es que se obtengan unos valores de MAE pequeños al haber utilizado también un margen de edades pequeño, por lo que no necesariamente un MAE de bajo valor indica que el sistema funciona satisfactoriamente.

En este trabajo se propone comparar el MAE obtenido por el sistema con lo que se denomina como *error base.* Dicho error base es el error obtenido por un sistema que asigna la misma edad $\hat{y} = c$ a todos los patrones (es decir, no hace uso de sus características). Desde el punto de vista Bayesiano, se puede demostrar que la c que minimiza el MAE en este contexto es la mediana de los y. Por tanto, siempre se comparará el MAE de los sistemas implementados con este error básico, ya que lo mínimo que se podrá esperar de un sistema es que tenga un error menor que esta base.

3.5.1.2 Clasificación dentro de un rango de edades

Al igual que en el caso de la estimación del valor de la edad, para la clasificación se busca obtener un 100% de acierto, pero se sabe que este hecho imposible por las características de la estimación de la edad y su individualidad.

De manera análoga al caso anterior, una buena manera de estimar si un resultado en clasificación es satisfactorio, es comprobando que el porcentaje de este valor sea mayor que el porcentaje de ocurrencias de la clase mayoritaria del conjunto de entrenamiento.

3.5.1.3 ¿Cuándo es aceptable un estimador?

En líneas generales, la evaluación de la calidad de los estimadores está muy ligada a cómo estén distribuidas las edades o clases en el conjunto de entrenamiento, siendo necesario comparar el MAE y el acierto en clasificación con la mediana de las edades de entrenamiento o porcentaje de la clase mayoritaria respectivamente.

3.5.2 Formación del conjunto de entrenamiento y de test

Existen diferentes maneras de seleccionar las imágenes que forman parte del conjunto de entrenamiento y de test. De las distintas posibles formaciones de éstos puede derivarse distintos resultados de MAE o acierto de clasificación, ya que pueden darse los casos en que los datos de test se asemejen mucho a los de entrenamiento y por tanto deriven en unos buenos resultados o puede ocurrir el caso contrario.

Para evitar el efecto que tienen estos casos, existen diferentes técnicas para la formación de los conjuntos de entrenamiento y de test que se detallan a continuación.

3.5.2.1 Particiones aleatorias

Una primera idea que se tiene para formar los conjuntos de entrenamiento y test es simplemente realizar una selección aleatoria de imágenes para un conjunto y otro, teniendo en cuenta que si la base de datos cuenta con varias imágenes de una misma persona se debe evitar que haya imágenes de un mismo sujeto en ambos conjuntos. A elección del diseñador del sistema queda el número de muestras en entrenamiento y en test, si bien es interesante que en entrenamiento haya al menos las mismas muestras que en test y que el conjunto de éstas sea lo más heterogéneo posible. Con el fin de no quedarse con una única prueba cuyos resultados que puede no representar el funcionamiento general del sistema, se puede aplicar este método con distintas particiones aleatorias y así finalmente tener un MAE o acierto en clasificación en media que sirve para evaluar el sistema.

3.5.2.2 Leave One Image Out (LOIO)

Esta técnica consiste en utilizar todas las imágenes de la base de datos para formar el conjunto de entrenamiento excepto una, que formaría parte del conjunto de test. Es recomendable realizar varias particiones para evitar tener los resultados de una única partición que no representa necesariamente el funcionamiento general del sistema. El problema de la utilización de esta técnica es que al trabajar con bases de datos que cuentan con varias imágenes de un mismo sujeto, se dará el caso en el que haya imágenes de este mismo sujeto en el conjunto de entrenamiento y de test a la vez. Para evitar este caso existe la técnica LOPO.

3.5.2.3 Leave One Person Out (LOPO)

Esta técnica se aplica cuando se tienen varias imágenes de un mismo sujeto. Para evitar que imágenes de una misma persona se encuentren en entrenamiento y en test, se seleccionan las muestras de todos los sujetos menos de uno en el conjunto de entrenamiento, quedando las muestras del sujeto fuera de entrenamiento para test, y así conseguir que la comparación de características de una persona se realice con otras personas y no con ella misma. De nuevo para evitar realizar una única prueba, se puede seguir esta técnica para todos los sujeto de la base de datos derivando así un MAE o acierto en clasificación en media.

3.5.3 Bases de datos

Las bases de datos sirven para realizar las pruebas en los desarrollos de diferentes sistemas y para entrenarlos. Una cualidad que se desea de éstas, sea cual sea el ámbito de desarrollo, es que sean lo más heterogéneas posibles para así poder abarcar el mayor espectro posible de la materia a tratar. Sin embargo, no solamente basta con que la información de la base de datos sea heterogénea, sino que también interesa que no exista un sesgo entre clases, ya que si esto ocurre, a pesar de que la base de datos sea rica en variedad de muestras de cada clase, si una o varias clases tienen muchas más muestras que las demás, el sistema tenderá más a predecir futuras muestras como pertenecientes a estas clases que a las demás.

Para evitar los problemas que pueden ocurrir por el sesgo en las bases de datos, en ocasiones se procede a eliminar imágenes del conjunto con mayor número de muestras con la intención de que el número de muestras en cada clase sea similar.

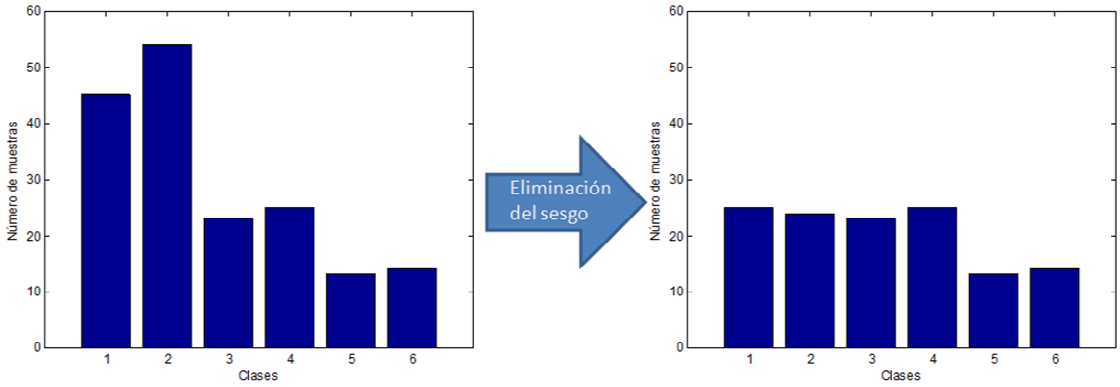

Figura 21: Ejemplo de distribución de clases en una base de datos con muestras sesgadas (izquierda) y distribución de la misma base de datos tras la eliminación del sesgo (derecha)

En la figura 21 se puede observar el ejemplo de un conjunto de datos de seis clases con claro sesgo hacia las clases 1 y 2, que puede derivar en que el sistema tienda a estimar esas dos clases por encima de la demás. Para evitar ese efecto, se eliminan muestras de estas clases así hacer que el sistema no tienda a predecir siempre estas mismas clases.

3.6 Resumen de las distintos sistemas que se implementan

La siguiente tabla ilustra los diferentes sistemas que se llevan a cabo a partir de las implementaciones realizadas en las distintas etapas explicadas a lo largo de este capítulo. Se implementan pues tanto sistemas propuestos en la literatura como propuestas nuevas estudiadas en el PFC.

Sistema	Características	Reducción de dimensionalidad	Algoritmo estimador de edad
[21]	AAM	SR	SVR
PFC	AAM	PCA	SVR
PFC	AAM	PCA	ANN
PFC	AAM	PCA	MVR
PFC	SIFT	No	KNN con pesos
[11]	BIF	PCA	SVR
PFC	BIF	PCA	MVR
PFC	BIF	PCA	Función *Classify*
PFC	BIF	PCA	SVM

Tabla 4: Resumen de los distintos estimadores de edad que se implementan

Las condiciones en las que se realizan las implementaciones y los resultados obtenidos para las mismas, se detallan en el siguiente capítulo *Pruebas y resultados*.

4 Pruebas y resultados

En el desarrollo de este capítulo se van a presentar los diferentes resultados experimentales obtenidos para los sistemas implementados. Estos sistemas son implementados utilizando las distintas técnicas de cada una de las etapas de los sistemas estimadores de edad que fueron explicadas en el capítulo anterior.

Los resultados que primero se presentan tienen como objetivo principal resolver el problema de la estimación del valor de la edad. Tras este primer análisis, se realizan una serie de pruebas desde el punto de vista de clasificación de la edad de los sujetos dentro de un rango de edades, siendo éste un problema menos exigente que la determinación del valor de la edad pero útil en otros ámbitos y para comprobar el correcto funcionamiento de la implementación.

4.1 Estimación del valor de la edad

En este apartado se afronta el problema de la estimación de la edad como un problema de regresión, es decir, como un problema en el que se intenta determinar la edad exacta de los sujetos.

En un primer lugar se presentan aquellos resultados de los sistemas implementados que no realizan una extracción de características propias de la edad por cada imagen, es decir, que trabajan directamente con información de píxeles. En segundo lugar los resultados que se presentan son los obtenidos en los sistemas implementados que sí realizan extracción de características de la edad y que por tanto, son más sofisticados. Los resultados obtenidos en los sistemas sin extracción de características sirven como referencia para comparar posteriormente con los resultados de los sistemas que sí extraen características.

Por último, se debe tener en cuenta que para este apartado, todos los resultados expuestos se obtienen tras la utilización de la base de datos FG-NET, ya que ésta es una base de datos muy utilizada en la literatura y con mucha información de las imágenes que facilita la implementación de sistemas estimadores de la edad. El uso de esta base de datos permite comprobar en un primer lugar si el sistema implementado tiene un funcionamiento correcto y a partir de este momento, si es interesante, probar otras bases de datos para confirmar el correcto funcionamiento del sistema.

4.1.1 Sistemas básicos

Este apartado tiene como objetivo presentar los resultados de sistemas que toman una imagen facial, reducen su dimensionalidad usando PCA y aplican un algoritmo de *machine learning* estándar para estimar la edad, es decir, no se realiza una extracción previa de características faciales complejas, con lo que sirven de referencia para implementaciones más sofisticadas. En la siguiente tabla se presentan los resultados obtenidos para estos sistemas además del error base de cada conjunto de entrenamiento. La reducción de dimensionalidad aplicada es PCA mientras que las imágenes están ecualizadas y alineadas a los ojos. También se detallan las técnicas seguidas para cada etapa de los sistemas estimadores de edad (exceptuando la extracción de características) y la manera en la que se forman los conjuntos de entrenamiento y de test en la base de datos utilizada.

Datos que utiliza el sistema	Conjuntos de entrenamiento y test	Estimador de edad	Error base	MAE
Imágenes recortadas por elipse	(Entr.): 802 primeras imágenes (Test): 200 últimas imágenes	Red neuronal	10.2	7.7
		MVR	10.2	**6.0**
	(Entr., Test): 100 particiones aleatorias	Red neuronal	9.8	7.2
		MVR	9.8	7.4
	(Entr., Test): LOPO	Red neuronal	9.7	7.1
		MVR	9.7	7.4

Tabla 5: Resultados para los sistemas básicos. Las unidades de los errores son años

En esta primera implementación, hay tres pruebas diferentes. La primera utiliza las 802 primeras imágenes de la base de datos para entrenar y el resto para hacer el test, siguiendo así la sugerencia de algunos trabajos de la literatura relacionada [21]. Con el afán de no utilizar una única partición que puede no reflejar el comportamiento real del sistema, la creación de 100 particiones aleatorias para entrenamiento y test (sin muestras del mismo sujeto en ambas) dan una mejor idea del funcionamiento de la implementación. Nótese que en estas 100 particiones, se mantiene el mismo número de imágenes en entrenamiento y en test que en la primera prueba. Una tercera prueba para esta base de datos se realiza mediante el método LOPO, ya que este método es el más utilizado en la literatura para evaluar sistemas [12] [27] [14] [11], además de funcionar mejor para bases de datos pequeñas y garantizar que no se tienen muestras del mismo sujeto en entrenamiento y test.

De los resultados obtenidos, se observa que para la partición única de 802 imágenes en entrenamiento y el resto en test, se obtiene un buen resultado en comparación con el error base para el estimador MvRegress, sin embargo, no se puede derivar el buen funcionamiento de esta partición como buen funcionamiento del sistema, ya que se puede observar que al realizar las 100 particiones y LOPO el resultado empeora.

Las dos primeras pruebas en realidad son las primeras que se realizan a modo intuitivo para comprobar el funcionamiento del sistema, primero con una partición aislada y después con varias particiones aleatorias. Sin embargo, el resto de pruebas que se llevan a cabo son mediante LOPO por las ventajas de este método, anteriormente nombradas. Además, en todas las pruebas el error será representado en años.

4.1.2 Sistemas con extracción de características complejas

En este apartado los resultados que se exponen son para aquellos sistemas que extraen características faciales complejas mediante las diferentes técnicas explicadas en el capítulo *Diseño*. En dicho capítulo se explicaron tres técnicas diferentes para extraer características, por lo que este apartado cuenta con tres subapartados, uno por cada técnica utilizada.

4.1.2.1 Extracción de características mediante AAM

Muchos de los mejores resultados publicados en la literatura con respecto al problema de determinar la edad, según se expuso en el capítulo *Estado del Arte,* se obtienen con el uso de los modelos activos de apariencia. Es por eso que se elige esta técnica para extraer características de la edad para las diferentes implementaciones que a continuación se comentan.

La primera implementación de esta técnica se basa principalmente en el mejor resultado encontrado en la literatura [21] que utiliza AAM, en donde no se realiza un preprocesamiento de la imagen previo, y hay una etapa de clasificación previa a la determinación del valor de la edad. En ella se divide el conjunto de muestras entre sujetos menores de 21 años y mayores de esta edad a partir de un SVM. Los resultados obtenidos para las distintas pruebas se pueden observar en la siguiente tabla.

Datos que utiliza el sistema	Reducción de dimensionalidad	Estimador de edad	Conjuntos de entrenamiento y test	Error base	MAE
AAM	SR	SVR	(Entr.): 802 primeras imágenes (Test): 200 últimas imágenes	10.2	**4.5**
			(Entr., Test): 100 particiones aleatorias	9.8	6.8±1.0
			(Entr., Test): LOPO	9.7	6.9

Tabla 6: Resultados para los sistemas implementados usando AAM basados en el mejor trabajo al respecto de la literatura [21]

Las tres distintas pruebas que se llevan a cabo son 1) el uso de las 802 primeras imágenes para formar el conjunto de entrenamiento y el resto como conjunto de test, tal y como se realiza en el trabajo en el que se basan las pruebas [21]; 2) el uso de 100 particiones aleatorias de entrenamiento y test; 3) el uso de particiones mediante LOPO para obtener una idea del comportamiento general del sistema. Se realiza LOPO en lugar de LOIO debido a que con este método es posible que haya imágenes de una misma persona en ambos conjuntos. Además se prioriza la evaluación mediante LOPO en vez de particiones aleatorias ya que con LOPO se evita que haya muestras de un mismo sujeto en los dos conjuntos formados y este método ayuda a una mejor evaluación de los sistemas cuando éstos usan bases de datos pequeñas.

Se puede observar que el mejor resultado obtenido es para la partición única formada por las 802 primeras imágenes en entrenamiento y el resto para test, mientras que cuando se realizan más particiones los resultados empeoran. De hecho, el resultado obtenido usando la partición 802+200 (MAE=4.5) está claramente fuera de la estadística obtenida con 100 particiones aleatorias (MAE=6.9±1.0), lo que indica que los resultados en la partición 802+200

son muy poco representativos de la calidad del sistema. Estas pruebas también han sido realizadas usando la información facilitada por AAM respecto a la forma aparte de las pruebas que usan forma y textura. Estas pruebas han concluido unos resultados idénticos que los obtenidos al usar el vector de apariencia (forma y textura), lo que da una idea de que en este sistema la información relevante se obtiene de la forma con la que trabaja el modelo activo de apariencia.

De cara a entender el papel de la reducción de dimensionalidad, se realizan pruebas sobre el funcionamiento del sistema implementado sin realizar reducción de dimensionalidad mediante regresión espectral, de la misma manera que se hizo en otros trabajos de la literatura [20], que es la parte más novedosa del anterior sistema implementado. Seguidamente se muestran los resultados obtenidos.

Datos que utiliza el sistema	Reducción de dimensionalidad	Estimador de edad	Conjuntos de entrenamiento y test	Error base	MAE
AAM	No	SVR	(Entr.): 802 primeras imágenes (Test): 200 últimas imágenes	10.2	**6.3**
			(Entr., Test): LOPO	9.7	9.2

Tabla 7: Resultados para los sistemas implementados usando AAM sin usar reducción de dimensionalidad

De nuevo se observa que el mejor resultado obtenido es para la partición formada por las 802 primeras imágenes en entrenamiento y el resto para test, mientras que los resultados obtenidos para LOPO son bastante peores. Estos resultados sirven para comprobar que realmente no es correcto evaluar este sistema con la primera partición, ya que generalmente ofrece mejores resultados que las particiones formadas de manera aleatoria o LOPO. Es por este motivo por el que carece de interés seguir realizando pruebas con esta partición para las próximas implementaciones. Otra observación de los resultados es que éstos son peores si no se realiza reducción de dimensionalidad, por lo que en las próximas implementaciones que usan AAM se lleva a cabo este paso.

De cara a estudiar el papel del clasificador de sujetos entre mayores y menores de una cierta edad, se realiza una prueba conservando la reducción de dimensionalidad mediante SR. El parámetro que varía es la edad que separa los sujetos en dos conjuntos. En el sistema original se hacía esta separación con menores de 21 años en un conjunto y el resto de sujetos en otro. Si no se realiza esta separación previa a la determinación del valor de la edad del sujeto, el valor de MAE que se obtiene evaluando el sistema con el método LOPO es de 6.8 años, es decir, un valor aproximado al obtenido con la separación previa de sujetos en dos grupos en una edad de 21 años, que parece indicar que esta separación previa no es necesaria. A continuación se muestran una gráfica de los resultados obtenidos al variar dicha edad de separación entre sujetos de cara a determinar si realmente el bloque que realiza una división entre los sujetos es realmente necesario para mejorar el funcionamiento del sistema. Las

pruebas se realizan para el conjunto formado con LOPO, por lo que ya es conocido que el error base es de 9.7 años.

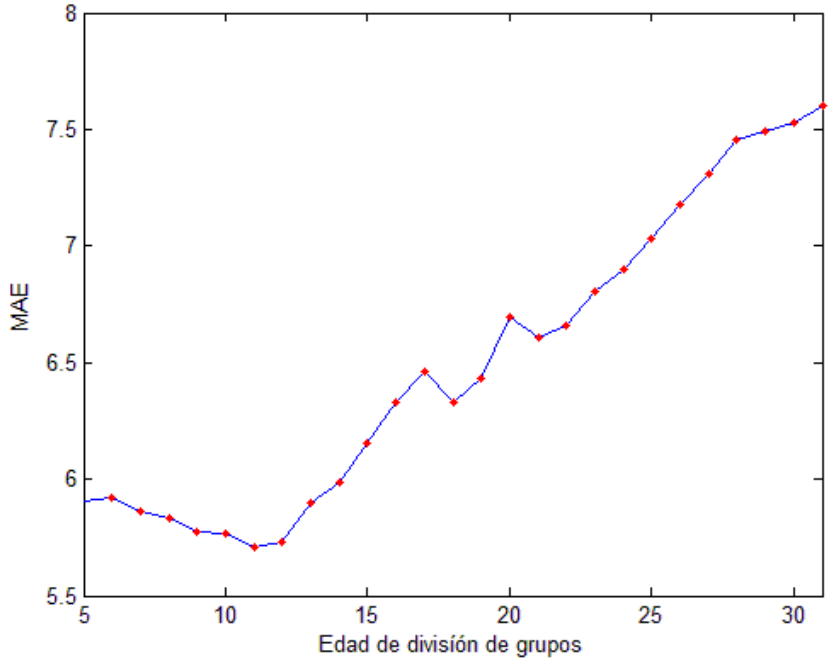

Figura 22: MAE calculado mediante LOPO en los sistemas implementados usando AAM basados en el mejor trabajo de la literatura [21] variando la edad de división entre grupos

A raíz de estos resultados se observa que la variación de la edad que divide los sujetos en dos conjuntos influye en el MAE obtenido al final, pudiendo llegar a valer un valor aproximado de 5.7 años para una edad de separación entre adultos de 11 años. De cara a evaluar si la elección de valor de 21 años es un valor ajustado a una única partición, seguidamente se observa la gráfica para la partición 802+200 variando la edad de división.

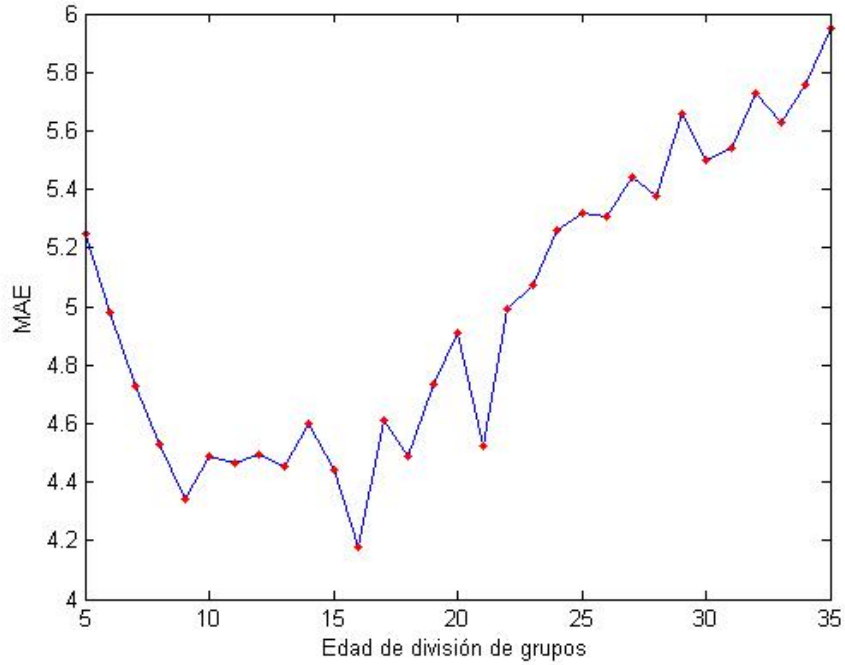

Figura 23: MAE calculado para la partición 802+200 en los sistemas implementados usando AAM variando la edad de división entre grupos

Como se observa en la gráfica anterior, existe un mínimo local para el valor de edad de 21 años para la partición 802+200. Este mínimo local no aparecía cuando se hizo el mismo cálculo de la variación del MAE mediante LOPO, por lo que la elección de este valor para hacer división de grupos entre las muestras se ajusta a su buen funcionamiento para una partición en concreto, pero se ha comprobado que no es el valor adecuado en el funcionamiento del sistema con diferentes particiones.

En otras implementaciones llevadas a cabo a partir de las características extraídas mediante los modelos activos de apariencia, tras los resultados anteriores se decide no realizar división entre dos grupos entre el conjunto de muestras ya que se observa que esta división está ligada al conjunto de muestras utilizado, variando únicamente el método de reducción de dimensionalidad utilizado y los estimadores de edad. Los resultados obtenidos para estas implementaciones son los siguientes.

Datos que utiliza el sistema	Reducción de dimensionalidad	Estimador de edad	Conjuntos de entrenamiento y test	Error base	MAE
AAM	PCA	Red neuronal	(Entr., Test): LOPO	9.7	6.5
		MVR			6.8
		SVR			7.2

Tabla 8: Resultados para los sistemas implementados usando AAM con distintos estimadores de edad y reducción de dimensionalidad que el trabajo basado

A raíz de los resultados se observa que los distintos estimadores ofrecen resultados similares que no varían mucho entre sí, mientras que el uso de PCA también ofrece buenos resultados pero peores para el estimador SVR que usando la reducción de dimensionalidad SR.

Los resultados de AAM usando solamente la forma (coordenadas espaciales) o el vector de apariencia (forma y textura) ya se comentaron que dan exactamente los mismos resultados en este sistema, con lo que la información relevante de los modelos activos de apariencia en este problema viene dada básicamente por la forma. Sabiendo que la información relevante en este sistema se extrae de la forma, se realiza una prueba sin división entre dos grupos, en la que se recortan las imágenes por el contorno facial, se ecualizan y alinean varios puntos de las imágenes en bruto a una imagen referencia, estos puntos son el contorno facial, las pupilas, la punta de la nariz y las comisuras labiales.

Datos que utiliza el sistema	Reducción de dimensionalidad	Estimador de edad	Conjuntos de entrenamiento y test	Error base	MAE
Imágenes alineadas a varios puntos	PCA	Red neuronal	(Entr., Test): LOPO	9.7	5.8
		MVR			6.1

Tabla 9: Resultados para los sistemas con imágenes alineadas de una manera más precisa

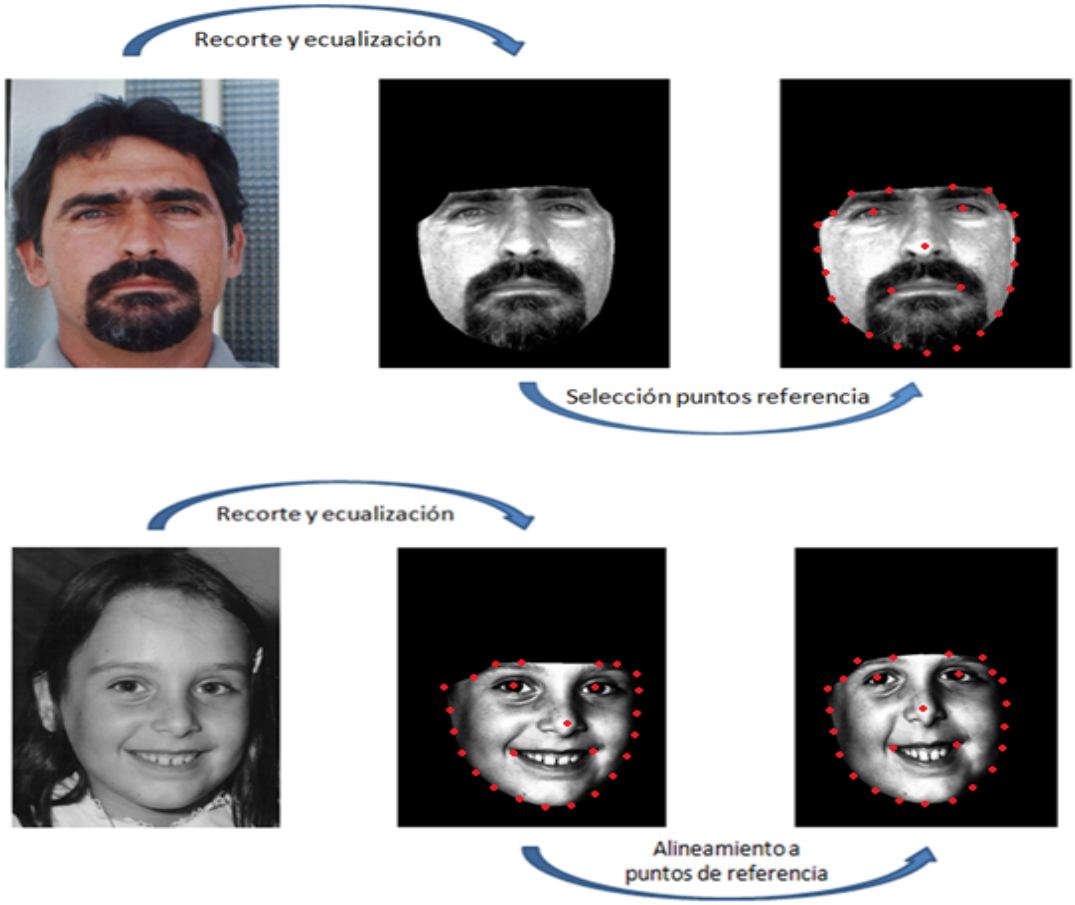

Figura 24: Selección de imagen de referencia (arriba) y alineamiento a sus puntos de referencia (abajo)

Se puede comprobar que los resultados tras este paso son ligeramente mejores al caso en el que se usaba AAM sin divisor de grupos y muy parecidos a los mejores resultados obtenidos al variar el clasificador previo, por lo que se deduce que realmente estos sistemas obtienen su información más relevante de la forma.

4.1.2.2 Extracción de características mediante SIFT

Una técnica muy utilizada en diversos campos que extrae características de las imágenes es SIFT, sin embargo, no existen estudios que la utilicen en el ámbito de la estimación de la edad. Debido a que esta técnica reporta muy buenos resultados para otro ámbito de clasificación como es el de género [42] y que las características SIFT son invariantes a traslaciones, rotaciones y escala, el uso de esta técnica es interesante para determinar si puede ser útil su utilización en este ámbito o no.

Las implementaciones que usan esta técnica son realizadas con los conjuntos de entrenamiento y test formados mediante LOPO, ya que es una manera más óptima de evaluar el comportamiento de los sistemas implementados que simplemente utilizando las primeras 802 imágenes en entrenamiento. La implementación cuenta como estimador de edad con la técnica KNN pesados, explicada en el capítulo *Diseño*. Los siguientes resultados son obtenidos en las distintas simulaciones variando el número de vecinos y la forma de pesar éstos.

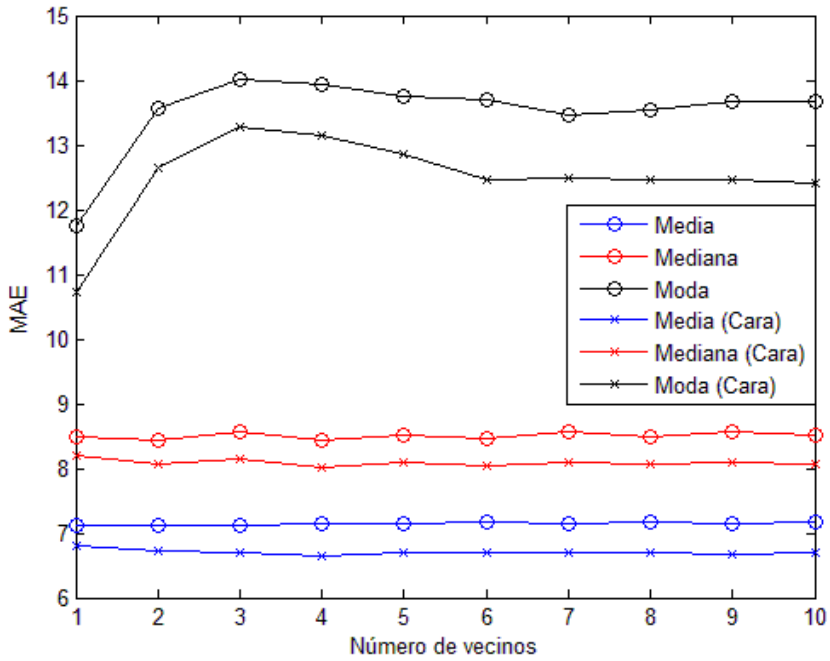

Figura 25: Resultados del MAE para el sistema implementado usando SIFT, variando el número de vecinos y la forma de pesar de éstos

	Forma de pesar los vecinos					
Número de vecinos	**Media**	**Media (cara)**	**Mediana**	**Mediana (cara)**	**Moda**	**Moda (cara)**
1	7.1	6.8	8.5	8.2	11.8	10.7
2	7.1	6.7	8.4	8.1	13.6	12.6
3	7.1	6.7	8.6	8.2	14.0	13.3
4	7.2	6.6	8.4	8.0	13.9	13.1
5	7.1	6.7	8.5	8.1	13.7	12.9
6	7.2	6.7	8.5	8.0	13.7	12.5
7	7.2	6.7	8.6	8.1	13.5	12.5
8	7.2	6.7	8.5	8.1	13.5	12.4
9	7.1	6.7	8.6	8.1	13.7	12.7
10	7.2	6.7	8.5	8.1	13.7	12.4

Tabla 10: Resultados del MAE para el sistema implementado usando SIFT, variando el número de vecinos y la forma de pesar de éstos

Se muestran los resultados correspondientes de realizar SIFT a toda la imagen y los correspondientes a realizar un recorte por el contorno facial, gracias a los puntos etiquetados que dispone la base de datos FG-NET. Se observa que los resultados que se obtienen cuando el sistema trabaja con imágenes recortadas son mejores que los correspondientes a la imagen completa, lo cual es lógico ya que al realizar el recorte se elimina información no relevante y que puede actuar como ruido. Además también se puede comprobar que el método que mejor funciona es pesar los vecinos próximos mediante la media, es decir, la nueva clase predicha es la media de las clases de sus k vecinos próximos, siendo el mejor MAE obtenido de 6.6 años aproximadamente.

Por las características de SIFT, es posible que una imagen tenga más puntos característicos que otras, lo cual puede conllevar a que las imágenes con más puntos sean las que comúnmente tengan más vecinos próximos en los test. Una posible solución a esto es

pesar cada vecino próximo por el número de puntos característicos que tenga esa imagen, es decir, si una imagen tiene muchos puntos característicos, a la hora de determinar la nueva clase de un punto de test, los puntos de la imagen con muchos puntos tendrán menos valor que los puntos de una imagen con pocos. Del mismo modo, puede que una clase tenga más muestras que el resto de clases, siendo más previsible predecir esa clase, por ello se puede proceder del mismo modo que con el número de puntos para evitar este efecto. Se muestran los resultados obtenidos teniendo en cuenta estos aspectos y una combinación de ambos.

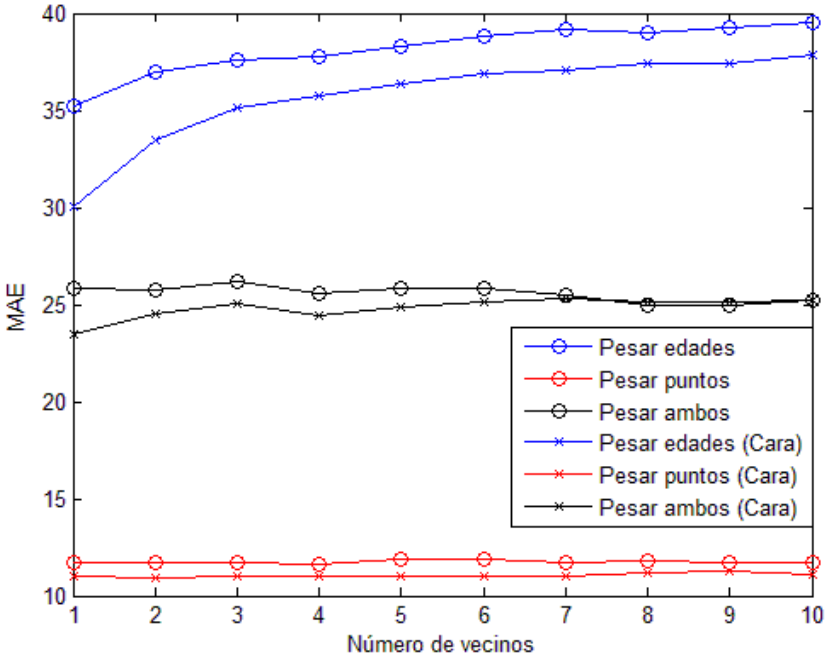

Figura 26: Resultados del MAE para el sistema implementado usando SIFT, variando el número de vecinos y teniendo en cuenta las densidades de puntos y edades

	Forma de pesar los puntos					
Número de vecinos	**Pesar edades**	**Pesar edades (cara)**	**Pesar puntos**	**Pesar puntos (cara)**	**Pesar ambos**	**Pesar ambos (cara)**
1	35.2	30.0	11.7	11.0	25.8	23.4
2	36.9	33.5	11.7	11.0	25.7	24.5
3	37.6	35.1	11.7	11.0	26.2	25.5
4	37.8	35.7	11.6	11.1	25.6	24.4
5	38.3	36.3	11.9	11.0	25.8	24.8
6	38.9	36.9	11.9	11.1	25.9	25.7
7	39.1	37.0	11.7	11.0	25.5	25.9
8	39.0	37.4	11.8	11.2	25.0	25.1
9	39.3	37.4	11.7	11.2	24.9	25.1
10	39.5	37.8	11.7	11.1	25.2	25.2

Tabla 11: Resultados del MAE para el sistema implementado usando SIFT, variando el número de vecinos y teniendo en cuenta las densidades de puntos y edades

Como se comprueba, de nuevo los resultados obtenidos con recorte de la cara son mejores que para los de la imagen completa, sin embargo, en esta ocasión los resultados son bastante peores que los obtenidos con los KNN pesados.

4.1.2.3 Extracción de características mediante BIF

La técnica para extraer características faciales complejas que mejores resultados ofrece según la literatura para la estimación de edad es la de las características de inspiración biológica. La implementación realizada se basa en el primer trabajo de la literatura [11] que utilizó BIF y que reportaba los mejores resultados en estimación de edad. En este apartado los resultados se obtienen mediante la técnica de LOPO, conociendo así el error base de 9.7 años, Además todas las implementaciones tienen en común la ecualización del histograma de cada imagen y que la reducción de dimensionalidad se realiza mediante PCA.

4.1.2.3.1 *Con alineamiento y recorte de imagen básicos*

Los resultados que se presentan son obtenidos tras realizar un recorte de la imagen en forma de elipse a partir de la localización de los ojos y una alineación con respecto a éstos. Debido a que BIF usa filtros de Gabor, se presentan los resultados variando el número de bandas utilizadas (de bandas menores a mayores) y el número de orientaciones de los filtros.

BIF de imágenes recortadas por elipse y alineadas a los ojos					
Nº de bandas = 2	**Número de orientaciones**				
Estimadores de edad	**4**	**6**	**8**	**10**	**12**
SVR	6.4	6.2	6.2	6.2	6.2
MVR	6.3	6.2	6.3	6.3	6.3
Función Classify	6.7	6.4	6.4	6.5	6.3
Nº de bandas =4	**Número de orientaciones**				
Estimadores de edad	**4**	**6**	**8**	**10**	**12**
SVR	6.4	6.2	6.2	6.2	6.2
MVR	6.4	6.3	6.3	6.2	6.2
Función Classify	6.4	6.3	6.2	6.3	6.2
Nº de bandas = 6	**Número de orientaciones**				
Estimadores de edad	**4**	**6**	**8**	**10**	**12**
SVR	6.3	6.2	6.1	6.1	6.1
MVR	6.3	6.1	6.2	6.1	6.1
Función Classify	6.2	6.1	**6.0**	**5.9**	**6.0**
Nº de bandas = 8	**Número de orientaciones**				
Estimadores de edad	**4**	**6**	**8**	**10**	**12**
SVR	6.2	6.1	6.1	6.1	6.1
MVR	6.3	6.1	6.2	6.1	6.1
Función Classify	6.2	6.2	**6.0**	6.1	6.2

Tabla 12: Resultados del MAE para el sistema implementado usando BIF, variando el número de orientaciones y bandas de los filtros Gabor con imágenes recortadas mediante elipse y alineadas por los ojos

Para esta primera implementación se tienen los mejores resultados generalmente cuando mayor es el número de orientaciones y bandas usadas. Se reportan unos resultados similares a los obtenidos anteriormente con AAM, pero sin la necesidad de localizar varios puntos previamente, sino con un simple recorte tras localizar solamente dos puntos.

4.1.2.3.2 *Con alineamiento basado en múltiples puntos y recorte basado en el contorno de la cara*

Una segunda implementación del sistema varía la forma de realizar el recorte y el alineamiento de imágenes, realizando un recorte del contorno facial gracias a puntos del contorno de la cara cuyas coordenadas se dan en la base de datos FG-NET y un alineamiento a distintos puntos de la cara (ojos, punta de la nariz, comisuras labiales y contorno facial). Para realizar este alineamiento se busca una imagen de referencia que tenga estos puntos lo suficientemente diferenciados y la cara con orientación frontal para un mejor alineamiento. Los resultados para esta implementación son los siguientes.

BIF de imágenes recortadas contorno facial y alineadas a varios puntos						
	Nº de bandas = 2	**Número de orientaciones**				
	Estimadores de edad	**4**	**6**	**8**	**10**	**12**
	SVR	6.3	6.2	6.2	6.2	6.2
	MVR	6.0	5.9	5.8	5.8	5.9
	Función Classify	6.0	5.7	5.6	5.7	5.6
	Nº de bandas =4	**Número de orientaciones**				
	Estimadores de edad	**4**	**6**	**8**	**10**	**12**
	SVR	6.2	6.1	6.1	6.1	6.1
	MVR	5.9	5.6	5.6	5.6	5.6
	Función Classify	5.9	5.7	5.6	5.6	5.6
	Nº de bandas = 6	**Número de orientaciones**				
	Estimadores de edad	**4**	**6**	**8**	**10**	**12**
	SVR	6.2	6.0	6.0	6.0	6.0
	MVR	5.8	**5.5**	**5.5**	**5.5**	**5.5**
	Función Classify	5.9	5.6	**5.5**	5.6	**5.5**
	Nº de bandas = 8	**Número de orientaciones**				
	Estimadores de edad	**4**	**6**	**8**	**10**	**12**
	SVR	6.1	6.0	6.0	6.0	6.0
	MVR	5.8	**5.5**	**5.5**	**5.5**	**5.5**
	Función Classify	5.8	5.6	5.6	**5.5**	**5.5**

Tabla 13: Resultados del MAE para el sistema implementado usando BIF, variando el número de orientaciones y bandas de los filtros Gabor con imágenes recortadas por el contorno facial de la imagen y alineadas a varios puntos

Los resultados en esta implementación que usa las imágenes alineadas a varios puntos y recortadas mejoran los de la implementación anterior que solamente usaba imágenes recortadas con elipse y alineadas a los ojos. Es lógico que esto ocurra pero es más costoso de realizar ya que se necesita localizar una cantidad importante de puntos previamente. Los mejores resultados se obtienen de nuevo a mayor número de bandas y orientaciones.

4.2 Estimación del rango de edad

Tras lo visto en el apartado anterior, se puede comprobar que de los métodos implementados, el que mejores resultados aporta son aquellos que realizan una extracción de características de inspiración biológicas (BIF). Es por ello que es interesante ahondar en el funcionamiento de estos sistemas y probarlos para otro tipo de problemas, como puede ser clasificar la edad de una persona dentro de un rango.

Para probar el funcionamiento de este extractor de características como clasificador, se prueba un caso práctico, que es la clasificación de imágenes como niños (de 0 a 8 años), adolescentes (de 9 a 17 años) y adultos (mayores de 18 años). Con la base de datos que se trabaja, se tienen unos porcentajes de muestras de cada rango creado de 34.53%, 29.34% y 36.13% respectivamente, siendo los adultos la clase mayoritaria para todos los conjuntos de entrenamiento, por lo que el porcentaje de acierto base es 36.13%. Los resultados mostrados para la estimación del rango de edad son bajo las mismas condiciones que los mostrados para la determinación de la edad exacta mediante el uso de BIF.

BIF de imágenes recortadas por elipse y alineadas a los ojos						
	Nº de bandas = 2	**Número de orientaciones**				
	Estimadores de edad	**4**	**6**	**8**	**10**	**12**
	SVM	68.2%	68.6%	69.8%	69.5%	69.8%
	Función Classify	70.8%	70.3%	71.2%	71.2%	71.4%
	Nº de bandas =4	**Número de orientaciones**				
	Estimadores de edad	**4**	**6**	**8**	**10**	**12**
	SVM	69.7%	71.0%	70.4%	70.1%	70.4%
	Función Classify	69.5%	72.7%	73.0%	72.7%	**73.1%**
	Nº de bandas = 6	**Número de orientaciones**				
	Estimadores de edad	**4**	**6**	**8**	**10**	**12**
	SVM	68.8%	70.4%	69.9%	69.4%	70.2%
	Función Classify	70.9%	71.7%	**73.1%**	72.9%	73.0%
	Nº de bandas = 8	**Número de orientaciones**				
	Estimadores de edad	**4**	**6**	**8**	**10**	**12**
	SVM	69.9%	70.2%	69.7%	69.9%	69.2%
	Función Classify	71.3%	72.5%	**73.2%**	**73.4%**	**73.1%**

Tabla 14: Resultados de porcentaje de acierto al clasificar entre tres clases diferentes utilizando BIF con imágenes recortadas mediante elipse y alineadas por los ojos

Para esta prueba, prácticamente se doblan los resultados del acierto base, por lo que realmente el sistema está siendo capaz de clasificar imágenes dentro de un rango de una manera más o menos buena. Como ocurría para estimación del valor de la edad, los mejores resultados se obtienen cuantas más bandas y orientaciones se usen.

BIF de imágenes recortadas contorno facial y alineadas a varios puntos						
	Nº de bandas = 2	**Número de orientaciones**				
	Estimadores de edad	**4**	**6**	**8**	**10**	**12**
	SVM	69.1%	68.7%	69.8%	69.7%	69.8%
	Función Classify	69.9%	70.3%	70.8%	70.7%	70.0%
	Nº de bandas =4	**Número de orientaciones**				
	Estimadores de edad	**4**	**6**	**8**	**10**	**12**
	SVM	69.5%	71.0%	70.3%	69.8%	69.6%
	Función Classify	68.9%	72.7%	72.9%	72.2%	72.6%
	Nº de bandas = 6	**Número de orientaciones**				
	Estimadores de edad	**4**	**6**	**8**	**10**	**12**
	SVM	69.6%	70.5%	70.2%	70.3%	70.9%
	Función Classify	68.9%	71.7%	72.9%	73.1%	**73.9%**
	Nº de bandas = 8	**Número de orientaciones**				
	Estimadores de edad	**4**	**6**	**8**	**10**	**12**
	SVM	70.5%	70.2%	69.4%	69.7%	70.7%
	Función Classify	69.3%	72.6%	**73.2%**	**73.2%**	**73.5%**

Tabla 15: Resultados de porcentaje de acierto al clasificar entre tres clases diferentes utilizando BIF con imágenes recortadas por el contorno facial de la imagen y alineadas a varios puntos

En esta implementación no parece haber una gran mejora con respecto al caso de uso de elipse para recortar y alineamiento por los ojos, aunque se mantienen unos resultados satisfactorios. Como viene ocurriendo en los casos anteriores, los mejores resultados se obtienen a más bandas y orientaciones.

4.3 Pruebas con otras bases de datos

Para evaluar un sistema a fondo, no basta con realizar pruebas a partir de una única base de datos, sino que es conveniente analizar dicho sistema con otras bases de datos para comprobar el funcionamiento del sistema en diferentes condiciones.

Ya que los mejores resultados que han sido obtenidos fueron a partir del uso de las características de inspiración biológica, las pruebas que se realizan con otras bases de datos se realizarán con esta técnica. En este caso, la base de datos utilizada es MORPH, que cuenta con una serie de imágenes de sujetos, la mayoría de ellas imágenes frontales, no existiendo un sesgo de edad entre las muestras, es decir, el reparto de edades en esta base de datos es más o menos equilibrada.

En la siguiente tabla se presentan los resultados obtenidos utilizando esta base de datos para estimar el valor de la edad de un sujeto. Las condiciones en las que se realiza este análisis son tras realizar 20 particiones aleatorias, ya que realizar LOPO en una base de datos tan extensa como la de estas pruebas es de una alta carga computacional. Además estas particiones se forman de tal manera que no haya muestras del mismo sujeto en entrenamiento y en test, siendo el 75% de las imágenes parte del conjunto de entrenamiento y el resto de test. Se simula en las siguientes pruebas bajo los valores de número de bandas y número de orientaciones de los filtros de Gabor que mejores resultados reportaron para la base de datos FG-NET (8 bandas y 8 orientaciones), además de realizar un procesamiento previo de las imágenes alineándolas a los ojos, ecualizándolas y realizando el recorte en elipse de la cara y siendo la reducción de dimensionalidad aplicada PCA.

Datos que utiliza el sistema	Conjuntos de entrenamiento y test	Estimador de edad	Error base	MAE
Imágenes recortadas por elipse	(Entr.): MORPH (75%) (Test): MORPH (25%)	SVR	13.1	6.9
		MVR		6.4
		Función Classify		7.5

Tabla 16: Resultados para los sistemas implementados usando BIF con la base de datos MORPH con la configuración de filtros óptima según FG-NET

En este caso, se obtienen resultados ligeramente peores respecto al valor de MAE que los obtenidos para FG-NET, sin embargo, el error base que se tiene en estas pruebas es mayor, por lo que resulta lógico que estos valores de MAE puedan resultar superiores.

Con la intención de explorar a fondo el funcionamiento del sistema, se implementa un sistema en el que una de las dos bases de datos actúa como conjunto de entrenamiento y la otra como conjunto de test. Dichas implementaciones se hacen tras realizar un procesamiento previo de las imágenes alineándolas a los ojos, ecualizándolas y realizando el recorte en elipse de la cara y siendo la reducción de dimensionalidad aplicada PCA. La primera prueba llevada a cabo consiste de nuevo en estimar el valor de la edad, siendo los resultados los siguientes.

Datos que utiliza el sistema	Conjuntos de entrenamiento y test	Estimador de edad	Error base	MAE
Imágenes recortadas por elipse	(Entr.): FG-NET (Test): MORPH	SVR	21.2	17.8
		MVR		13.1
		Función Classify		17.5
	(Entr.): MORPH (Test): FG-NET	SVR	24.6	14.3
		MVR		12.2
		Función Classify		11.7

Tabla 17: Resultados para los sistemas implementados usando BIF para estimación del valor de la edad usando una base de datos en entrenamiento diferente a la de test

Como en la anterior prueba, se tienen unos MAE altos pero se debe considerar que el valor del error base es mucho mayor para esta clase de pruebas, ya que se mezclan bases de

datos totalmente diferentes. Además para la base de datos MORPH no hay muestras de sujetos menores de 15 años, por lo que se realiza una nueva prueba eliminando de la base de datos FG-NET los sujetos menores de esta edad, obteniendo los siguientes resultados.

Datos que utiliza el sistema	Conjuntos de entrenamiento y test	Estimador de edad	Error base	MAE
Imágenes recortadas por elipse	(Entr.): FG-NET (edades ≥15años) (Test): MORPH	SVR	16.7	13.3
		MVR		11.1
		Función Classify		14.4
	(Entr.): MORPH (Test): FG-NET (edades ≥15años)	SVR	13.7	8.7
		MVR		7.8
		Función Classify		9.2

Tabla 18: Resultados para los sistemas implementados usando BIF para estimación del valor de la edad usando una base de datos en entrenamiento diferente a la de test eliminado imágenes de una base de datos

Al eliminar de la base de datos FG-NET los sujetos con edades no reflejadas en MORPH, los resultados mejoran como era de esperar. Debe tenerse en cuenta que en este caso, cuando se entrena con FG-NET es normal que los resultados sean peores que cuando se entrena con MORPH, ya que hay unas 450 imágenes en entrenamiento y unas 4500 imágenes en test.

Siguiendo con el estudio, se realiza una prueba con respecto a clasificar la edad dentro de un rango de edades. Para que esta prueba sea más fácil de interpretar, se afronta el problema de clasificación de imágenes mayores o menores de 18 años, eliminando una serie de imágenes de cada base de datos, de tal manera que el porcentaje de sujetos de cada una de las dos clases en cada base de datos sea del 50% para cada conjunto. De nuevo los valores del número de bandas y orientaciones usado es el del mejor caso reportado anterior, la reducción de dimensionalidad aplicada es PCA y se realiza un procesamiento previo de las imágenes alineándolas a los ojos, ecualizándolas y realizando el recorte en elipse de la cara

Datos que utiliza el sistema	Estimador de edad	Conjuntos de entrenamiento y test	Porcentaje de acierto base	Porcentaje de acierto
Imágenes recortadas por elipse	SVM	(Entr.): FG-NET (Test): MORPH	50%	65.4%
	Función Classify			62.2%
	SVM	(Entr.): MORPH (Test): FG-NET		57.3%
	Función Classify			56.2%

Tabla 19: Resultados para los sistemas implementados usando BIF para clasificación usando una base de datos en entrenamiento diferente a la de test

Los resultados superan la tasa de acierto base aunque no por mucho. La diferencia de características entre ambas bases de datos es la causante principal de estos resultados.

5 Conclusiones y trabajo futuro

En este último capítulo se exponen las conclusiones del trabajo llevado a cabo en este proyecto de fin de carrera. La mayor parte de estas conclusiones se extraen a partir de los resultados presentados en el capítulo anterior. El objetivo principal de este capítulo es, en base a los resultados mostrados, contestar a la pregunta: ¿qué sistema concreto se debe implementar si se quiere estimar automáticamente la edad en base a los rasgos faciales? De esta forma se soluciona el problema inicial de este proyecto, que es definir, implementar y caracterizar una implementación concreta de detección automática de edad que nos ha pedido una empresa especializada en visión artificial, con vistas a ser utilizada en un contexto de seguridad.

En un segundo apartado se exponen las posibles líneas de trabajo futuro con respecto a los sistemas que se han tratado en este trabajo, añadiendo nuevas etapas a las implementaciones, variando el funcionamiento de las mismas o introduciendo técnicas más sofisticadas y su posible desarrollo.

5.1 Conclusiones

Durante el transcurso de este proyecto de fin de carrera se han llevado a cabo diferentes implementaciones de sistemas estimadores de edad. Cada sistema utiliza diferente tipo de información y la estructura de las diferentes etapas que lo forman varía.

5.1.1 Sistemas básicos

En un primer lugar se realizaron una serie de pruebas implementando sistemas que no realizaban extracción de características complejas para determinar la edad, sino que simplemente trabajaban con las componentes principales de la información suministrada por la imagen a nivel de píxeles. La intención de estas pruebas no era más que servir de comparador para las implementaciones más sofisticadas realizadas con extracción de características complejas. Una conclusión que se deriva de las pruebas es que es importante comparar los resultados obtenidos con el error base, ya que no podemos hablar de MAE en términos absolutos porque un bajo valor de MAE no significa necesariamente que el sistema sea bueno si el error base es bajo y lo mismo en valores con MAE altos que no necesariamente significan un sistema no satisfactorio. De hecho existen artículos en la literatura [13] cuyos resultados obtenidos están por encima del error base y por tanto no son tan buenos como parecen. Lo mismo se aplica a las tasas de acierto cuando se habla de estimación de la edad dentro de un rango.

Una de las primeras conclusiones que se podía empezar a intuir con estas implementaciones era la necesidad caracterizar el error del sistema con más de una partición, ya que para la partición 802+200 (primeras 802 imágenes de la base de datos para entrenamiento y el resto para test) el MAE era de 6.0 años para el estimador MVR, mientras que con otras particiones este valor sube bastante (por ejemplo, si se hace una estadística de los MAE en test con 100 particiones aleatorias de entrenamiento y test, el MAE promedio obtenido son 7.4 años con el mismo estimador). Cuando no hay muchos patrones disponibles, una técnica estándar en *machine learning* para estimar el error de generalización es el *Leave One Person Out* (LOPO) [14] [11]. Su utilización inmediata en problemas de clasificación de

imágenes es lo que se conoce *Leave One Image Out* (LOIO). Sin embargo, en problemas donde hay sujetos con diferentes imágenes (en nuestro caso, imágenes obtenidas a diferentes edades) es más riguroso utilizar LOPO, en el que cada conjunto de test está formado por todas las imágenes de una única persona. En nuestro caso el MAE de test obtenido mediante LOPO es 7.1 años para el estimador por red neuronal y 7.4 años para MVR, por debajo en ambos casos del error base.

5.1.2 Sistemas con extracción de características complejas

Dentro de las implementaciones con técnicas de extracción de características complejas la primera que se probó fue una estrategia basada en modelos activos de apariencia. Basándose en el trabajo que mejores resultados mostraba en la literatura usando esta técnica [21] se realizó una primera implementación. En dicho trabajo se proponía realizar una partición con las primeras 802 imágenes de la base de datos FG-NET para entrenar y el resto para realizar test. Los resultados obtenidos en ese trabajo fueron replicados (MAE=4.5años) para esa partición pero se comprobó que si se utilizaban otras particiones de entrenamiento y test, tanto usando particiones aleatorias como LOPO, estos resultados empeoraban bastante, con un MAE de 6.8 años. Después de analizar los resultados se concluye que esto se debe principalmente a que con esta partición se tiene en test solamente diez sujetos de más de 40 años, tres por encima de 50 años y solamente uno por encima de 60 años, por lo que en test la mayoría de muestras son de edades jóvenes. Esto unido al sesgo existente en la base de datos FG-NET produce que se obtengan unos buenos resultados para esta partición en concreto. De estos resultados se concluye algo que ya se vino anticipando durante el desarrollo de la memoria, que es la necesidad de evaluar los sistemas con más de una partición, ya que es probable que se dé la circunstancia de que una única partición reporte unos resultados mucho mejores o peores que los que el sistema realmente puede conseguir.

Se ha comprobado por tanto que dicho trabajo, que hemos replicado por ser el que mejor MAE tenía en la literatura utilizando AAM, no es en realidad tan bueno sino que está sobreajustando a la partición concreta con la que trabajan. De hecho, ninguna de las particiones aleatorias que se han probado en esta implementación tiene un MAE menor que el de la partición 802+200. Sin embargo este trabajo disponía de algunas características interesantes por las cuales destacaba, como eran el uso de la regresión espectral con la que reducía cada vector que definía a un sujeto con su edad a una sola dimensión. La necesidad de usar este método para el funcionamiento del sistema se comprueba al realizar estas mismas pruebas sin reducción de dimensionalidad (tabla 7), donde los resultados empeoran hasta obtener un MAE cercano al error base, por lo que el funcionamiento del sistema no es satisfactorio. Esto es debido a que al realizar reducción de dimensionalidad de cada AAM se elimina información no relevante que puede actuar como ruido dentro del sistema. Otra novedad en este sistema es el clasificador intermedio entre menores y mayores de 21 años. Se puede ver que esta edad de corte no es la que mejores resultados da con la base de datos utilizada en estas pruebas, como puede comprobarse en la figura 22. Sin embargo no se puede elegir la edad de división en función de los distintos resultados obtenidos en test, ya que de esta manera se estaría sobreajustando un parámetro en función de las muestras con las que se trabaja. Es por ello que con la intención de evaluar el comportamiento de dicho clasificador se elimina del sistema, obteniendo un valor de 0.1 años menor en MAE que usando este

clasificador previo con la edad de 21 años. De esto se concluye que dicho clasificador intermedio en realidad no está aportando nada relevante al sistema.

En otra prueba, se trabaja solamente con el vector de forma de AAM en vez de con las características suministradas por vector y forma de manera conjunta, obteniéndose en este caso los mismos resultados que en pruebas anteriores, de donde se puede concluir que la información relevante sobre la edad para este sistema del vector AAM no está en la textura sino en la forma. La prueba realizada para comprobar esto fue alinear todas las caras a unos puntos de referencia para que estuvieran alineadas, y trabajar a nivel de píxel con ellas con estimadores sencillos, obteniendo en estos casos unos resultados en MAE de 5.8 años con el mejor estimador, resultados mejores que los de la implementación con clasificador para dividir entre dos grupos y con SVR. Solamente este valor es superado cuando el parámetro del clasificador previo se ajusta al conjunto de muestras, por lo que se deduce que la información de las coordenadas espaciales es muy relevante para el funcionamiento del sistema.

En la tabla 8 se puede comprobar el funcionamiento del sistema con otros estimadores y reducción de dimensionalidad, donde los resultados mejoran a los estimadores y reducción de dimensionalidad sugeridos, mientras que el cambio de reducción de dimensionalidad de SR a PCA empeora los resultados para el estimador SVR que sugiere el trabajo. Cerrando el tema de los sistemas con modelos activos de apariencia, se concluye que los resultados no son tan satisfactorios como se ilustra en el trabajo debido a que se realiza un sobreajuste de parámetros y componentes del sistema a una sola partición, mientras que para las demás particiones se tienen peores resultados que los que se pueden conseguir con sistemas básicos, como se ilustra en la tabla 8.

En otros sistemas de la literatura [39] [12] que también usaban AAM, se obtenían buenos resultados tras una serie de pruebas LOPO, siendo la diferencia principal con los sistemas implementados el uso de estimadores de edad más sofisticados, por lo que puede que el comportamiento mejore con estimadores más sofisticados, pero todo esto a costa de añadir un grado de complejidad al sistema. Esto añadido a la necesidad de tener que localizar una serie de puntos para las imágenes de entrenamiento previamente para realizar AAM, y a la necesidad en la práctica de disponer de un método automático que haga esto de una manera robusta para imágenes no vistas anteriormente lleva a descartar la opción de implementar un sistema basado en AAM.

Debido al interés en llevar a cabo sistemas automáticos robustos, se buscaron técnicas que pudieran trabajar directamente con la imagen sin necesidad de tener que localizar una gran cantidad de puntos. Una de las técnicas utilizadas en las implementaciones que no necesita estos puntos es SIFT, que además no necesita que las imágenes sean alineadas, es invariante a cambios de escala, traslación y rotación y ofrecía buenos resultados en otros problemas de clasificación diferentes a edad [42]. Se han realizado implementaciones bajo dos circunstancias distintas que son trabajar con la imagen completa o con la selección previa de la cara de la imagen gracias a la localización de los ojos. Dichos recortes pueden realizarse tras una localización previa de los ojos con técnicas relativamente sencillas y realizar el recorte con una forma geométrica fija. Los resultados obtenidos para la imagen recortada son mejores que para la imagen completa, lo que resulta lógico ya que solamente se trabaja con la información

deseada y nos ayuda a concluir la necesidad de realizar un recorte previo de la imagen con el afán de trabajar solamente con la cara.

En las implementaciones realizadas se propone una técnica que usa KNN pero pesa de manera diferente las contribuciones de cada vecino. Dicha técnica consiste en realizar una media ponderada, la moda o la mediana de las edades de los vecinos próximos. La primera prueba llevada a cabo con esta nueva técnica (figura 25, tabla 10) ofreció los mejores valores cuando la forma de pesar era la clase media de entre un conjunto de vecinos próximos, seguido de la mediana y finalmente la moda. De estos resultados se puede concluir que cuando se trabaja con varias clases como en el problema de la edad en la que cada edad es una clase, se reportan mejores resultados haciendo la media de las clases vecinas. Sin embargo, el mejor resultado reportado en MAE es 6.6 años, inferior a lo obtenido en la implementación anterior, mientras que las otras técnicas dan resultados peores, cercanos al error base usando la mediana o incluso por encima de éste usando la moda, aunque se debe tener en cuenta que este valor de MAE se obtiene sin la necesidad de ningún proceso manual previo. Analizando las características de la técnica SIFT, se concluye que una imagen puede tener más puntos característicos que otra y por tanto tener más vecinos próximos que otras imágenes. El mismo efecto ocurre cuando existe una clase con muchas más muestras que las demás. Para evitar este efecto se hizo una implementación que penalizaba los puntos de una imagen cuantos más puntos tuviera esa misma imagen, utilizando esta misma técnica también para las edades. Los resultados son bastante malos para esta implementación, en el caso de penalizar la edad debido a que cuando un vecino próximo pertenece a una edad poco común ésta pesa demasiado en el sistema y se tiende a predecir esta edad, mientras que en los sistemas que penalizan por puntos los resultados están por encima del error base. Se concluye que esta forma de pesar los puntos o clases no es óptima a la vista de los malos resultados

En vista de los resultados, y a pesar de las ventajas que aporta el uso de la técnica SIFT por no necesitar localizar puntos (en todo caso localizar puntos sencillos para realizar un recorte con forma geométrica) y ser invariante a escala, rotación y traslación, se descarta la implementación de un sistema que use esta técnica, que a pesar de superar los resultados de los sistemas básicos que no extraen características de la imagen, no ofrece unos resultados suficientemente satisfactorios como sí se obtienen en cambio por ejemplo con el género [42]. Es por ello que se sugiere que esta línea de investigación debería de seguir siendo desarrollada debido a su buen funcionamiento en otros ámbitos.

Siguiendo con la intención de realizar sistemas que no necesiten la localización de varios puntos, se lleva a cabo el desarrollo de sistemas que usen la técnica para extraer características de inspiración biológica BIF basado en el primer trabajo que utilizó esta técnica [11]. Esta técnica es la que mejores resultados reportaba en la literatura relacionada, como se vio en el capítulo *Estado del Arte*. La primera prueba realizada trabajaba con imágenes con recorte previo de la cara. Debido a que BIF usa un banco de filtros Gabor, el valor óptimo de número de orientaciones y bandas varía con la base de datos. Las pruebas realizadas permiten concluir que generalmente los mejores resultados se encuentran cuantas más bandas y orientaciones se usen, coincidiendo con lo expuesto en otros trabajos de la literatura [11] [23], debido a que usando más bandas y orientaciones se tienen más vectores con más información referente a la edad. Los resultados (tabla 12) no varían mucho con los distintos estimadores

utilizados, siendo el valor de MAE entre 5.9 y 6.1 años para los casos con más bandas y orientaciones usadas (mejores que los obtenidos con las implementaciones con SIFT). De cara a estudiar el mejor funcionamiento posible de este sistema, se realizó una prueba con un alineamiento previo de las caras a varios puntos, al igual que en el estudio del sistema con AAM, obteniendo una mejora en el valor del MAE hasta los 5.5 años. Estos son los mejores resultados obtenidos de las distintas implementaciones realizadas, sin embargo, son obtenidos utilizando localización a varios puntos, lo que no es deseado por introducir un factor de complejidad mayor al sistema. El sistema en el que la implementación está basada [11] obtenía unos valores de 4.8 años en MAE con la base de datos utilizada en esta implementación pero realizando un alineamiento no determinado y a varios puntos de la cara. La manera de acercarse a este valor es realizando el mismo alineamiento, pero al no ser conocido, la implementación realizada consigue acercarse solamente a 0.7 años de este valor. En realidad los resultados obtenidos con las imágenes recortadas simplemente por elipse, a pesar de tener un año más en su valor de MAE, tienen como ventaja frente al sistema de la literatura su sencillez al no necesitar muchos puntos para alinear. De todo esto se puede concluir que cuando el alineamiento se realiza a varios puntos de la cara los resultados de MAE obtenidos mejoran. En la situación de que se dispongan de sistemas capaces de localizar varios puntos de la cara el alineamiento será mejor y por tanto los resultados mejorarán, pero se debe tener en cuenta que si el sistema para localizar puntos de la cara no funciona correctamente, el alineamiento no será correcto y por tanto los resultados empeorarán.

Debido a los resultados que se obtienen con BIF para la estimación del valor de la edad, además de ser un sistema que no necesita localizar muchos puntos, con el añadido de que computacionalmente es menos complejo que AAM y SIFT, se elige realizar el sistema que desea la empresa usando esta técnica. En vista a esta decisión, se realizan más pruebas basadas en esta técnica relacionadas con la estimación de la edad, como es la clasificación de la edad dentro de un rango. Estas pruebas se realizaron a partir del hipotético problema de clasificar las imágenes entre niños, adolescentes y adultos, estando estas tres clases aproximadamente equilibradas en el reparto de muestras, teniendo la clase mayoritaria un 36.13% de las muestras totales. De nuevo los mejores resultados se obtienen a mayor número de bandas y orientaciones, obteniendo una tasa de acierto del 73% aproximadamente tanto para el caso de trabajar con imágenes recortadas como con imágenes alineadas a varios puntos. Que el clasificador obtenga resultados parecidos con la imagen recortada que con la alineada a varios puntos permite concluir que el alineamiento a varios puntos para este sistema es importante a la hora de estimar el valor de la edad pero no tanto para la clasificación de la edad dentro de un rango de edades, lo cual es una ventaja en el uso de este sistema como clasificador. Se considera que un 73% es un valor bueno para el problema planteado, ya que dobla la tasa de acierto base, por lo que se comprueba el buen carácter clasificador del sistema implementado.

Ya que no existe un número óptimo de orientaciones y bandas a utilizar pero se ha comprobado que generalmente los mejores resultados se obtienen cuantas más bandas y orientaciones se usan, se eligen los valores de 8 bandas y 8 orientaciones (10 o 12 orientaciones reportan resultados similares y añaden un coste computacional mayor) para la implementación del sistema. Bajo estos valores se llevan a cabo más pruebas con otras bases de datos. En este caso la base de datos utilizada es MORPH, debido a su gran cantidad de

imágenes y a que no sufre un sesgo entre edades como sufría FG-NET. En este caso, los resultados en MAE obtenidos para el mejor estimador son 6.4 años para las imágenes recortadas, lo que dista solamente en 0.5 años del MAE obtenido bajo las mismas condiciones con la otra base de datos, pero teniendo en cuenta que el error base es mayor, se considera un buen resultado. Estos resultados son mejores que los reportados en la literatura para esta base de datos para estimación de la edad [34], debido a que esta base de datos no había sido utilizada con técnicas tan sofisticadas o con técnicas sin necesidad de localizar puntos. Cabe destacar que en este caso uno de los mejores estimadores de edad para la anterior base de datos, la función *Classify* de Matlab®**,** obtiene los peores resultados de los estimadores utilizados anteriormente, por lo que se puede concluir que no existe un estimador mejor que otro, aunque a tenor de los resultados, el estimador de edad elegido sería el MVR que es el que mejores resultados ofrece en el cómputo global de los resultados.

5.1.2.1 Pruebas cruzando bases de datos

Las últimas pruebas consistieron en cruzar las bases de datos FG-NET y MORPH de tal forma que el conjunto de entrenamiento estuviera formado solamente por imágenes de MORPH y el conjunto de test estuviera formado solamente por imágenes de FG-NET y viceversa. Los resultados no son tan buenos como cuando ambos conjuntos se obtienen de una única base de datos (sin embargo el valor de MAE comparado con el error base es bueno). Esto es debido a existen grandes diferencias entre ambas bases de datos, dándose el caso de que una no tiene sujetos menores de 15 años (MORPH) mientras que la otra (FG-NET) tiene la mayoría de sus sujetos (57.3%) bajo esta edad. Es por ello que se realizó una prueba eliminando las muestras menores de 15 años en FG-NET (tabla 15), mejorando los resultados de las implementaciones ya que en uno de los dos conjuntos de entrenamiento o test no hay carencia de muestras de una clase que son mayoría en el otro conjunto. En estas pruebas de nuevo los mejores resultados son con el estimador MVR, que no hace más que confirmar la buena elección de este estimador para el sistema a implementar para la empresa. La conclusión que se obtiene de estas simulaciones es que es importante entrenar a los sistemas con bases de datos que tengan una distribución parecida a test. En caso de que la distribución en test sea completamente desconocida, recomendamos utilizar un conjunto de entrenamiento una distribución de edades lo más homogénea posible.

Las pruebas llevadas a cabo cruzando bases de datos dan una idea general sobre el funcionamiento del sistema en los casos en los que las imágenes de entrenamiento difirieran mucho de las de test, lo cual puede ocurrir en aplicaciones prácticas. Los resultados obtenidos en estas pruebas eran unos valores de MAE de unos 8 a 11 años para el estimador MVR, que fue el elegido para la implementación. Estas pruebas se realizaron con bases de datos que diferían en la distribución de clases y poses de los sujetos (en MORPH las imágenes son en su mayoría frontales y en FG-NET hay diferentes poses). Es por ello que la recomendación para que los resultados que se obtengan en la práctica se acerquen a los conseguidos en las primeras implementaciones (MAE aproximado de 6 años) es realizar un entrenamiento con sujetos en diferentes poses: tanto frontal, de lado, inclinado, etc.; además de evitar que existan edades con muchas más muestras que otras, porque de esta manera el sistema puede estar sesgado hacia estas edades mayoritarias.

Las recomendaciones para formar el conjunto de entrenamiento del sistema pueden variar según las condiciones en las que vaya a utilizarse el estimador. Si se conoce que va a ser utilizado bajo imágenes de la misma postura en todos los casos (siendo por ejemplo una situación posible su utilización siempre bajo imágenes frontales), entonces es lógico que se entrene con imágenes en esa misma pose. En estos casos incluso sería posible que el sistema detectase previamente si las imágenes con las que trabajan cumplen unos requisitos de pose para estimar su edad con la detección previa de una serie de puntos clave de la cara, como pueden ser los ojos, nariz y boca, y así determinar si la pose de la cara en la imagen entra dentro del conjunto de imágenes que pueden ser estimadas por el sistema. Si se desconoce el tipo de imágenes que se van a someter al sistema, entonces lo más recomendable es entrenar al sistema con imágenes en múltiples poses, con varias fotos de cada sujeto, con sujetos de distintas edades y distintas razas y evitando que una clase sea mayoritaria con respecto a otras. Esto se puede hacer creando una base de datos o a partir de bases de datos existentes, como pueden ser las utilizadas en este trabajo, pudiendo eliminar imágenes de esas bases de datos para tener en cuenta los aspectos mencionados anteriormente de no conocer qué tipo de imágenes de van a someter al sistema.

5.2 Trabajo futuro

El trabajo realizado en este proyecto de fin de carrera tiene diversas líneas de continuación posibles que pueden permitir la mejora de los sistemas estimadores de edad llevados a cabo. Un posible trabajo de futuro sería la implementación de un reconocedor de género previo. Algunos trabajos [14] [24] [12] realizan una división manual en la base de datos con la que trabajan entre hombres y mujeres, aunque esta división podría realizarse de manera automática como ya se ha hecho en algunos trabajos [42]. Tras esta división, implementan dos sistemas estimadores de edad, uno por cada género, mejorando los resultados obtenidos con respecto a cuándo se usan todos los sujetos en un mismo sistema. La realización de un sistema reconocedor de género puede entonces mejorar el funcionamiento de los sistemas implementados en este trabajo, aunque también liga el correcto funcionamiento del sistema al correcto funcionamiento del reconocedor de género.

Respecto al sistema implementado con SIFT, se han comentado las distintas fortalezas del uso de esta técnica, pero a pesar de estas virtudes los resultados obtenidos no fueron satisfactorios. Una posible línea de trabajo futuro con respecto a estos sistemas sería la inclusión de estimadores de edad más sofisticados que el propuesto en este trabajo.

Como se comentó en el apartado anterior, las diferentes posturas pueden interferir en la estimación de la edad. Por ello, como trabajo futuro se puede plantear un sistema de predicción de edad que construya un estimador diferente para cada pose. Una posible estrategia es añadir una etapa previa que sea capaz de determinar la postura de una imagen a partir de la localización de una serie de puntos en la cara, tales como los ojos, la nariz y la boca. Tras la determinación de la pose de la cara, utilizar el estimador construido para esa misma postura.

Glosario de acrónimos

MAE: Mean Absolute Error

LOPO: Leave One Person Out

LOIO: Leave One Image Out

AAM: Active Appearance Models

SIFT: Scale-Invariant Features Transform

BIF: Biologically Inspired Features

AGES: AGing PattErn Subspace

PCA: Principal Component Analysis

LDA: Linear Discriminant Analysis

SR: Spectral Regression

SVM: Support Vector Machine

SVR: Support Vector Regression

KNN: K-Nearest Neighbors

ANN: Artificial Neural Network

MVR: MultiVariate linear Regression

Referencias

1. *Automatic Classification of Single Facial Images.* 12, Diciembre 1999, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 21, pp. 1357-1362.

2. **Geng, X., Fu, Y. and Smith-Miles, K.** University of Buffalo. *Computer Science and Engineering.* [Online] Agosto 30, 2010. http://www.cse.buffalo.edu/~yunfu/papers/pricai10_t4.pdf.

3. Age Progression. [Online] 2010. http://www.forensicartist.com/agepro.html.

4. *Estimating Age Privacy Leakage in Online Social Networks.* **Dey, R.** 2011. IT Security for the Next Generation.

5. *Soft Biometric Traits for Personal Recognition Systems.* **Jain, A. K., Dass, S. C. and Nandakumar, K.** Hong Kong : s.n., 2004. Proceedings of International Conference on Biometric Authentication. pp. 731-738.

6. *An age estimation system on the AIBO.* **Fukai, H., Nishie, Y., Abiko, K., Mitsukura, Y., Fukumi, M. y Tanaka, M.** Seoul : s.n., 2008. International Conference on Control, Automation and Systems, 2008. ICCAS 2008. pp. 2551-2554.

7. *Alcohol servers' estimates of young people's ages.* **Willner, P. and Rowe, G.** 2001, Drugs: Education, Prevention, and Policy, 8, pp. 375–383.

8. *Selling tobacco to minors: Can merchants accurately determine a customers age?* **Jason, L. A., Pokorny, S. B., Sherk, J. L., Helzing, M. D. y Rebus, P. J.** 2003, Journal of Human Behavior in the Social Environment, 8, pp. 67-73.

9. *A review of the literature on the aging adult skull and face: Implications for forensic science research and applications.* **Albert, A.K., Ricanek Jr., K. and Patterson, E.** 172(1), s.l. : Forensic Science International, 2007, pp. 1-9.

10. *Modeling the Process of Ageing in Face Images.* **Lanitis, A., Taylor, C. J. and Cootes, T. F.** 1999. The Proceedings of the Seventh IEEE International Conference on Computer Vision. Vol. 1, pp. 131-136.

11. *Human Age Estimation Using Bio-Inspired Features.* **Guo, G., Mu, G., Fu, Y. y Huang, T.S.** Miami, FL : s.n., 2009. IEEE Conference on Computer Vision and Pattern Recognition, 2009. pp. 112-119.

12. *Image-Based Human Age Estimation by Manifold Learning and Locally Adjusted Robust Regression.* **Guo, G., Fu, Y., Dyer, C.R. y Huang, T.S.** 7, Julio 2008, IEEE Transactions on Image Processing, Vol. 17.

13. *Automatic Age Estimation Based on Facial Aging Patterns.* **Geng, X., Zhou, Z. H. and Smith-Miles, K.** 12, Diciembre 2007, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 29, pp. 2234-2240.

14. *Ranking with Uncertain Labels.* **Yan, S., Wang, H., Huang, T.S., Yang, Q y Tang, X.** Beijing : s.n., 2007. IEEE International Conference on Multimedia and Expo, 2007. pp. 96-99.

15. *Locating facial features for age classification.* **Kwon, Y. and Lobo, N. V.** 1993. Proceedings of SPIE - the Intl. Society for Optical Engineering. Vol. 2055, pp. 62-72.

16. *Age Classification from Facial Images.* **Kwon, Y. and Lobo, N. V.** 1, Abril 1999, Computer Vision and Image Understanding, Vol. 74, pp. 1-21.

17. **Alley, T. R.** *Social and Applied Aspects of Perceiving Faces.* s.l. : Erlbaum Assoc., 1988.

18. *Active Appearance Models.* **Cootes, T. F., Edwards, G. J. and Taylor, C. J.** 1998, ECCV.

19. *Comparing Different Classifiers for Automatic Age Estimation.* **Lanitis, A., Draganova, C. and Christodoulou, C.** 1, Febr. 2004, IEEE Transactions on Systems, Man and Cybernetics - Part B: Cybernetics, Vol. 34.

20. *Age Estimation using Active Appearance Models and Support Vector Machine Regression.* **Luu, K., Bui, T.D., Ricanek Jr., K. y Suen, C.Y.** Washington D.C. : s.n., Septiembre 2009, Proceedings of the IEEE Conference on Biometrics: Theory, Applications, and Systems (BTAS).

21. *Spectral Regression Based Age Determination.* **Luu, K., Bui, T. D., Ricanek Jr., K. y Suen, C. Y.** 2010.

22. *Robust Object Recognition with Cortex-Like Mechanisms.* **Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M. y Poggio, T.** 3, Marzo 2007, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 29, pp. 411-426.

23. *Human age estimation using enhanced bio-inspired features.* **El Dib, M.Y. and El-Saban, M.** Hong Kong : s.n., 2010. 17th IEEE International Conference on Image Processing. pp. 1589-1592.

24. *Regression from Patch-Kernel.* **Yan, S, Zhou, X., Liu, M., Hasegawa-Johnson, M. y Huang, T.S.** Anchorage, AK : s.n., 2008. IEEE Conference on Computer Vision and Pattern Recognition, 2008. págs. 1-8.

25. *Learning from Facial Aging Patterns for Automatic Age Estimation.* **Geng, X., Zhou, Z.H., Zhang, Y., Li, G. y Dai, H.** Santa Barbara, California, USA : s.n., 2006. MM'06.

26. **Jolliffe, I.T.** *Principal Component Analysis.* s.l. : Springer, 2002.

27. *A Study on Automatic Age Estimation Using a Large Database.* **Guo, G., Mu, G., Fu, Y., Dyer, C. y Huang, T.S.** 2009. IEEE Conf. Computer Vision, 2009.

28. **Cai, D. and He, X. Han, J.** *Spectral Regression for Dimensionality reduction.* Department of Computer Science, U. of Illinois at Urbana-Champaign. Augsburg : s.n., 2007. Technical Report Nº 2856.

29. *Support Vector Method for Function Approximation, Regression Estimation and Signal Procesing.* **Vapnik, V., Golowich, S. E. and Smola, A. J.** 1996, NIPS, pp. 281-287.

30. *Statistical learning theory.* **Vapnik, V.** New York : s.n., 1998, John Wiley and Sons.

31. *Kernel Regression for Image Processing and Reconstruction.* **Takeda, H., Farsiu, S. and Milanfar, P.** 2, 2007, IEEE Transcations on Image Processing, Vol. 16, pp. 349-366.

32. FG-NET Aging Database. [Online] http://www.fgnet.rsunit.com/.

33. *MORPH: A Longitudinal Image Database of Normal Adult Age-Progression.* **Tesafaye, T. and Ricanek Jr., K.** Southampton, UK : s.n., 2006. IEEE 7th International Conference on Automatic Face and Gesture Recognition. pp. 341-345.

34. *Age Regression Based on Local Images Features.* **Mokadem, A, Charbit, M and Chollet, G.** Singapur : s.n., 2010. Symposium on Image and Video Technology (PSIVT), 2010 Fourth Pacific-Rim. pp. 88-93.

35. **Sinha, P. K.** *Image Acquisition and Preprocessing: Foundations for Machine Vision.* s.l. : Spie Press.

36. *Active shape models - their training and application.* **Albert, M., Ricanek Jr., K. and Patterson, E.** 1, 1995, Computer Vision Graphics and Image Understanding, Vol. 65, pp. 38-59.

37. *Object recognition from local scale-invariant features.* **Lowe, D. G.** 1999. Proceedings of the International Conference on Computer Vision. Vol. 2, pp. 1150-1157.

38. *A 3-Dimensional SIFT Descriptor and its Application to Action Recognition.* **Scovanner, P, Ali, S and Shah, M.** Augsburg : s.n., Septiembre 23-28, 2007, MM, 07.

39. *Locally Adjusted Robust Regression for Human Age Estimation.* **Guo, G., Fu, Y., Huang, T.S y Dyer, C.R.** Copper Mountain, CO : s.n., Enero 7-9, 2008, IEEE Workshop on Applications of Computer Vision, 2008, pp. 1-6.

40. **Anderson, J. A.** *An Introduction to Neural Networks.* s.l. : Hardcover, 1995.

41. **Dasarathy, B. V.** *Nearest Neighbor (NN) Norms: NN Pattern Classification Techniques.* 1991.

42. *A comparison of techniques for robust gender recognition.* **Rojas-Bello, R.N., Lago-Fernandez, L.F., Martínez-Munoz, G. y Sánchez-Montañés, M.A.** 2011. Proceedings of the 18th IEEE International Conference on Image Processing (ICIP). pp. 561-564.

43. Wikipedia. [Online] http://es.wikipedia.org/.

44. Scholarpedia. [Online] http://www.scholarpedia.org/.

PRESUPUESTO

1) Ejecución Material

- Compra de ordenador personal (Software incluido)..... ..... 2.000 €
- Material de oficina ..... 200 €
- Total de ejecución material..... 2.200 €

2) Gastos generales

- 16 % sobre Ejecución Material..... 352 €

3) Beneficio Industrial

- 6 % sobre Ejecución Material..... 132 €

4) Honorarios Proyecto

- 1000 horas a 15 € / hora..... 15000 €

5) Material fungible

- Gastos de impresión ..... 100 €
- Encuadernación ..... 20 €

6) Subtotal del presupuesto

- Subtotal Presupuesto..... 17804 €

7) I.V.A. aplicable

- 16% Subtotal Presupuesto..... 2848.6 €

8) Total presupuesto

- Total Presupuesto ..... 20652,6 €

Madrid, Enero de 2012

El Ingeniero Jefe de Proyecto

Fdo.: Antonio Rodríguez Ruiz

Ingeniero de Telecomunicación

PLIEGO DE CONDICIONES

Este documento contiene las condiciones legales que guiarán la realización, en este proyecto, de un sistema estimador de la edad. En lo que sigue, se supondrá que el proyecto ha sido encargado por una empresa cliente a una empresa consultora con la finalidad de realizar dicho sistema. Dicha empresa ha debido desarrollar una línea de investigación con objeto de elaborar el proyecto. Esta línea de investigación, junto con el posterior desarrollo de los programas está amparada por las condiciones particulares del siguiente pliego.

Supuesto que la utilización industrial de los métodos recogidos en el presente proyecto ha sido decidida por parte de la empresa cliente o de otras, la obra a realizar se regulará por las siguientes:

Condiciones generales

1. La modalidad de contratación será el concurso. La adjudicación se hará, por tanto, a la proposición más favorable sin atender exclusivamente al valor económico, dependiendo de las mayores garantías ofrecidas. La empresa que somete el proyecto a concurso se reserva el derecho a declararlo desierto.

2. El montaje y mecanización completa de los equipos que intervengan será realizado totalmente por la empresa licitadora.

3. En la oferta, se hará constar el precio total por el que se compromete a realizar la obra y el tanto por ciento de baja que supone este precio en relación con un importe límite si este se hubiera fijado.

4. La obra se realizará bajo la dirección técnica de un Ingeniero de Telecomunicación, auxiliado por el número de Ingenieros Técnicos y Programadores que se estime preciso para el desarrollo de la misma.

5. Aparte del Ingeniero Director, el contratista tendrá derecho a contratar al resto del personal, pudiendo ceder esta prerrogativa a favor del Ingeniero Director, quien no estará obligado a aceptarla.

6. El contratista tiene derecho a sacar copias a su costa de los planos, pliego de condiciones y presupuestos. El Ingeniero autor del proyecto autorizará con su firma las copias solicitadas por el contratista después de confrontarlas.

7. Se abonará al contratista la obra que realmente ejecute con sujeción al proyecto que sirvió de base para la contratación, a las modificaciones autorizadas por la superioridad o a las órdenes que con arreglo a sus facultades le hayan comunicado por escrito al Ingeniero Director de obras siempre que dicha obra se haya ajustado a los preceptos de los pliegos de condiciones, con arreglo a los cuales, se harán las modificaciones y la valoración de las diversas unidades sin que el importe total pueda exceder de los presupuestos aprobados. Por consiguiente, el número de unidades que se consignan en el proyecto o en el presupuesto, no podrá servirle de fundamento para entablar reclamaciones de ninguna clase, salvo en los casos de rescisión.

8. Tanto en las certificaciones de obras como en la liquidación final, se abonarán los trabajos realizados por el contratista a los precios de ejecución material que figuran en el presupuesto para cada unidad de la obra.

9. Si excepcionalmente se hubiera ejecutado algún trabajo que no se ajustase a las condiciones de la contrata pero que sin embargo es admisible a juicio del Ingeniero Director de obras, se dará conocimiento a la Dirección, proponiendo a la vez la rebaja de precios que el Ingeniero estime justa y si la Dirección resolviera aceptar la obra, quedará el contratista obligado a conformarse con la rebaja acordada.

10. Cuando se juzgue necesario emplear materiales o ejecutar obras que no figuren en el presupuesto de la contrata, se evaluará su importe a los precios asignados a otras obras o materiales análogos si los hubiere y cuando no, se discutirán entre el Ingeniero Director y el contratista, sometiéndolos a la aprobación de la Dirección. Los nuevos precios convenidos por uno u otro procedimiento, se sujetarán siempre al establecido en el punto anterior.

11. Cuando el contratista, con autorización del Ingeniero Director de obras, emplee materiales de calidad más elevada o de mayores dimensiones de lo estipulado en el proyecto, o sustituya una clase de fabricación por otra que tenga asignado mayor precio o ejecute con mayores dimensiones cualquier otra parte de las obras, o en general, introduzca en ellas cualquier modificación que sea beneficiosa a juicio del Ingeniero Director de obras, no tendrá derecho sin embargo, sino a lo que le correspondería si hubiera realizado la obra con estricta sujeción a lo proyectado y contratado.

12. Las cantidades calculadas para obras accesorias, aunque figuren por partida alzada en el presupuesto final (general), no serán abonadas sino a los precios de la contrata, según las condiciones de la misma y los proyectos particulares que para ellas se formen, o en su defecto, por lo que resulte de su medición final.

13. El contratista queda obligado a abonar al Ingeniero autor del proyecto y director de obras así como a los Ingenieros Técnicos, el importe de sus respectivos honorarios facultativos por formación del proyecto, dirección técnica y administración en su caso, con arreglo a las tarifas y honorarios vigentes.

14. Concluida la ejecución de la obra, será reconocida por el Ingeniero Director que a tal efecto designe la empresa.

15. La garantía definitiva será del 4% del presupuesto y la provisional del 2%.

16. La forma de pago será por certificaciones mensuales de la obra ejecutada, de acuerdo con los precios del presupuesto, deducida la baja si la hubiera.

17. La fecha de comienzo de las obras será a partir de los 15 días naturales del replanteo oficial de las mismas y la definitiva, al año de haber ejecutado la provisional, procediéndose si no existe reclamación alguna, a la reclamación de la fianza.

18. Si el contratista al efectuar el replanteo, observase algún error en el proyecto, deberá comunicarlo en el plazo de quince días al Ingeniero Director de obras, pues transcurrido ese plazo será responsable de la exactitud del proyecto.

19. El contratista está obligado a designar una persona responsable que se entenderá con el Ingeniero Director de obras, o con el delegado que éste designe, para todo relacionado con ella. Al ser el Ingeniero Director de obras el que interpreta el proyecto, el contratista deberá consultarle cualquier duda que surja en su realización.

20. Durante la realización de la obra, se girarán visitas de inspección por personal facultativo de la empresa cliente, para hacer las comprobaciones que se crean oportunas. Es obligación del contratista, la conservación de la obra ya ejecutada hasta la

recepción de la misma, por lo que el deterioro parcial o total de ella, aunque sea por agentes atmosféricos u otras causas, deberá ser reparado o reconstruido por su cuenta.

21. El contratista, deberá realizar la obra en el plazo mencionado a partir de la fecha del contrato, incurriendo en multa, por retraso de la ejecución siempre que éste no sea debido a causas de fuerza mayor. A la terminación de la obra, se hará una recepción provisional previo reconocimiento y examen por la dirección técnica, el depositario de efectos, el interventor y el jefe de servicio o un representante, estampando su conformidad el contratista.

22. Hecha la recepción provisional, se certificará al contratista el resto de la obra, reservándose la administración el importe de los gastos de conservación de la misma hasta su recepción definitiva y la fianza durante el tiempo señalado como plazo de garantía. La recepción definitiva se hará en las mismas condiciones que la provisional, extendiéndose el acta correspondiente. El Director Técnico propondrá a la Junta Económica la devolución de la fianza al contratista de acuerdo con las condiciones económicas legales establecidas.

23. Las tarifas para la determinación de honorarios, reguladas por orden de la Presidencia del Gobierno el 19 de Octubre de 1961, se aplicarán sobre el denominado en la actualidad “Presupuesto de Ejecución de Contrata” y anteriormente llamado ”Presupuesto de Ejecución Material” que hoy designa otro concepto.

Condiciones particulares

La empresa consultora, que ha desarrollado el presente proyecto, lo entregará a la empresa cliente bajo las condiciones generales ya formuladas, debiendo añadirse las siguientes condiciones particulares:

1. La propiedad intelectual de los procesos descritos y analizados en el presente trabajo, pertenece por entero a la empresa consultora representada por el Ingeniero Director del Proyecto.

2. La empresa consultora se reserva el derecho a la utilización total o parcial de los resultados de la investigación realizada para desarrollar el siguiente proyecto, bien para su publicación o bien para su uso en trabajos o proyectos posteriores, para la misma empresa cliente o para otra.

3. Cualquier tipo de reproducción aparte de las reseñadas en las condiciones generales, bien sea para uso particular de la empresa cliente, o para cualquier otra aplicación, contará con autorización expresa y por escrito del Ingeniero Director del Proyecto, que actuará en representación de la empresa consultora.

4. En la autorización se ha de hacer constar la aplicación a que se destinan sus reproducciones así como su cantidad.

5. En todas las reproducciones se indicará su procedencia, explicitando el nombre del proyecto, nombre del Ingeniero Director y de la empresa consultora.

6. Si el proyecto pasa la etapa de desarrollo, cualquier modificación que se realice sobre él, deberá ser notificada al Ingeniero Director del Proyecto y a criterio de éste, la empresa consultora decidirá aceptar o no la modificación propuesta.

7. Si la modificación se acepta, la empresa consultora se hará responsable al mismo nivel que el proyecto inicial del que resulta el añadirla.

8. Si la modificación no es aceptada, por el contrario, la empresa consultora declinará toda responsabilidad que se derive de la aplicación o influencia de la misma.

9. Si la empresa cliente decide desarrollar industrialmente uno o varios productos en los que resulte parcial o totalmente aplicable el estudio de este proyecto, deberá comunicarlo a la empresa consultora.

10. La empresa consultora no se responsabiliza de los efectos laterales que se puedan producir en el momento en que se utilice la herramienta objeto del presente proyecto para la realización de otras aplicaciones.

11. La empresa consultora tendrá prioridad respecto a otras en la elaboración de los proyectos auxiliares que fuese necesario desarrollar para dicha aplicación industrial, siempre que no haga explícita renuncia a este hecho. En este caso, deberá autorizar expresamente los proyectos presentados por otros.

12. El Ingeniero Director del presente proyecto, será el responsable de la dirección de la aplicación industrial siempre que la empresa consultora lo estime oportuno. En caso contrario, la persona designada deberá contar con la autorización del mismo, quien delegará en él las responsabilidades que ostente.