

CECAL

INSTITUTO DE COMPUTACIÓN, FACULTAD DE INGENIERÍA

UNIVERSIDAD DE LA REPÚBLICA

MONTEVIDEO, URUGUAY

**PROYECTO DE GRADO
INGENIERÍA EN
COMPUTACIÓN**

**Reconocimiento Facial Robusto al
Envejecimiento**

Demián Arduş, Matías De Horta, Sebastián Melgar
demian.ardus@gmail.com, matias.dh@gmail.com,
sebastian.melgar91@gmail.com

Diciembre de 2016

Tutores del Proyecto:

Sergio Nesmachnow, Universidad de la República
Alicia Fernández, Universidad de la República

Co-Tutores:

Luis Di Martino, Dirección Nacional de Identificación Civil
Javier Preciozzi, Universidad de la República, Dirección
Nacional de Identificación Civil

Reconocimiento Facial Robusto al Envejecimiento
Ardus, Demián
De Horta, Matías
Melgar, Sebastián
Proyecto de Grado
CECAL
Instituto de Computación - Facultad de Ingeniería
Instituto de Ingeniería Eléctrica - Facultad de Ingeniería
Universidad de la República
Montevideo, Uruguay, Diciembre de 2016

Índice

1. Introducción	5
2. Sistemas de reconocimiento facial y el problema del envejecimiento	7
2.1. Introducción	7
2.2. Características de los sistemas de reconocimiento facial	9
2.3. Envejecimiento	11
2.4. Trabajos Relacionados	14
2.4.1. Maximum Entropy Feature Descriptor for Age Invariant Face Recognition	14
2.4.2. Convolutional Neural Networks for Age-Invariant Face Recognition	15
2.4.3. A Study of Face Recognition as People Age	16
2.4.4. Face Recognition and Retrieval Using Cross-Age Reference Coding With Cross-Age Celebrity Dataset	16
3. Descripción de la solución	19
3.1. Normalización y preprocesamiento	19
3.2. Extracción de características	21
3.2.1. LBP - Local Binary Patterns	21
3.2.2. HDLBP - High Dimensional Local Binary Patterns . .	23
3.2.3. PCA - Principal Component Analysis	24
3.2.4. Parches vs landmarks	24
3.3. Clasificación	25
3.4. CARC - Cross-Age Reference Coding	26
4. Implementación	31
4.1. Descripción general	31
4.2. Arquitectura del sistema	32
4.3. Vista de implementación	33
4.3.1. Aplicación Ruby on Rails	34
4.3.2. Base de datos	35
4.3.3. Sistema de archivos	37
4.3.4. Módulos de C++	37
4.4. Módulos principales	38
4.4.1. Preprocesamiento y normalización	38
4.4.2. Extracción de características	39
4.4.3. Clasificación	40

5. Evaluación experimental	41
5.1. Métricas	41
5.1.1. Base de datos	41
5.1.2. Identificación	42
5.1.3. Verificación	43
5.1.4. Protocolos	44
5.2. Base de datos y metodologías	45
5.2.1. Base de datos FERET	45
5.2.2. Base de datos MORPH	46
5.2.3. Base de datos DNIC	48
5.3. Configuraciones de Ageval	49
5.4. Resultados	50
5.4.1. Resultados para la base FERET	50
5.4.2. Resultados para la base MORPH	51
5.4.3. Resultados para la base de datos DNIC	53
5.5. Análisis de desempeño computacional	54
6. Conclusiones y trabajo futuro	58
6.1. Conclusiones	58
6.2. Trabajo futuro	59
A. Especificación casos de uso	62
A.1. Especificación de Casos de Uso Críticos en Administración . .	62
A.2. Especificación de Casos de Uso Críticos en la Plataforma Web	68
B. Resultados en el álbum 1 de MORPH para los conjuntos M, F, A y W	70
B.1. Conjunto M (Hombres)	70
B.2. Conjunto F (Mujeres)	71
B.3. Conjunto A (hombres y mujeres afroamericanos)	72
B.4. Conjunto W (hombres y mujeres con descendencia europea) .	73

1. Introducción

El reconocimiento facial se ha tornado en un tema de estudio muy activo dentro del área de la visión por computadora [1].

El desempeño de los sistemas de reconocimiento facial ha ido incrementando al correr de los años, afrontando desafíos como cambios en la iluminación, así como también en la pose entre las diferentes fotos [1]. Sin embargo, la mayoría de los estudios se basan en reconocimiento facial a nivel general, pero muy pocos basan su estudio en reconocimiento facial robusto al envejecimiento [26]. Por lo tanto, el reconocimiento facial robusto al envejecimiento pasa a ser un gran desafío en el área. Además de que los sistemas de reconocimiento facial tienen una gran utilidad para muchas aplicaciones como lo son la búsqueda de niños perdidos e identificación de criminales, entre otras [10].

Una técnica implementada en el estudio Cross-Age Reference Coding for Age-Invariant Face Recognition and Retrieval realizado por Bor-Chun Chen et al. (2014) logra superar el desempeño de un humano ante la verificación de fotos de personas tomadas en diferentes intervalos de tiempo [4], es decir, el sistema es capaz de reconocer certeramente una persona con mayor precisión que un humano. De todas formas la técnica no es tan precisa si se compara con el promedio de los votos de un conjunto de humanos.

Por otro lado, se sabe que el pasar de los años afecta de forma similar a las personas de un mismo género o etnia, por lo tanto, es de sumo interés estudiar como varían estas personas al pasar los años [5].

A su vez, existen derivaciones de aplicaciones relacionadas del estudio del envejecimiento facial: reconocimiento y verificación facial robusta al envejecimiento; predicción del envejecimiento a través de los años; estimación de edad [5]. Las aplicaciones de reconocimiento/verificación facial son usadas en lugares donde la seguridad es un factor fundamental. Es necesario mantener la base de datos actualizada con las fotos mas recientes de las personas (un ejemplo de uso es en los aeropuertos). Las aplicaciones que predicen el envejecimiento de una persona pueden ser utilizadas para encontrar personas perdidas hace mucho tiempo. Por último las aplicaciones que predicen la edad pueden ser utilizadas para automatizar el ingreso de la edad de una persona por un administrador en el sistema, obteniendo la edad desde la foto. Un ejemplo de una aplicación que realiza la estimación de edad es <https://how-old.net/>, implementado por Microsoft.

Este proyecto se enmarca en las actividades de colaboración del Instituto de Ingeniería Eléctrica con la Dirección Nacional de Identificación Civil (DNIC), Uruguay. La DNIC necesita un sistema que sea capaz de reconocer personas dentro de una base de datos con determinadas características y

desafíos. Un sistema de identificación civil se caracteriza por contar con pocas imágenes de cada individuo, lo que limita las posibilidades de entrenar clasificadores para obtener un mejor desempeño al momento de clasificar caras. Otro desafío es el poder reconocer caras de una misma persona tomadas en intervalos de diez años aproximadamente, para poder afrontar este desafío el sistema de reconocimiento facial a utilizar debe ser robusto al envejecimiento.

Este proyecto continúa trabajos previos de proyectos de fin de carrera en los cuales se diseñaron sistemas de reconocimiento facial, en particular FaceVal [19] que es un sistema de reconocimiento facial implementado en 2012 por Luis Di Martino, Gabriel Lema, y Sebastián Berchesi. En este proyecto se plantea realizar una reingeniería del sistema de reconocimiento facial Faceval y dotarlo de nuevos módulos y algoritmos que permitan mejorar el reconocimiento y la robustez frente al tiempo entre tomas.

En líneas generales este proyecto tiene dos objetivos principales: por una parte, se busca implementar un sistema de reconocimiento facial que permita a los administradores de la DNIC acceder desde una plataforma web; por otra parte, se busca poder extraer características de una cara que sean eficientes y así mejorar el reconocimiento facial de personas con fotos tomadas en diferentes períodos de tiempo. Siguiendo la línea marcada por estos objetivos, las principales contribuciones del trabajo son las siguientes: el análisis de la literatura relacionada al problema del reconocimiento facial robusto al envejecimiento; la implementación de un sistema de reconocimiento facial que sea robusto al envejecimiento; la implementación de la automatización de pruebas que se ejecutan en sistemas de reconocimiento fácil; la implementación de la automatización de la carga de bases de datos de FERET, MORPH 1, MORPH 2, CACD y DNIC; la implementación de una interfaz de acceso para los administradores del sistema; la evaluación experimental del desempeño del sistema implementado sobre las bases FERET, MORPH 1, MORPH 2 y DNIC.

El resto del documento se estructura del modo que se describe a continuación. En el Capítulo 2 se definen los conceptos relacionados y las principales dificultades identificadas en el estado del arte de los sistemas de reconocimiento facial. En el Capítulo 3 se detallan los principales módulos de un sistema de reconocimiento facial, detallando las técnicas implementadas, se describe el problema del reconocimiento facial ante el envejecimiento y se presenta una propuesta de solución al problema. En el Capítulo 4 se describe la implementación del sistema desarrollado presentando la arquitectura, los componentes y su implementación. En el Capítulo 5 se describe la evaluación experimental del sistema desarrollado. Por último, las conclusiones del proyecto y las principales líneas de trabajo futuro se presentan en el Capítulo 6.

2. Sistemas de reconocimiento facial y el problema del envejecimiento

En este capítulo se presentan los conceptos relacionados con los sistemas de reconocimiento facial. La Sección 2.1 presenta una introducción a este tipo de sistemas, sus características principales y las etapas típicas de un proceso de reconocimiento de patrones. La Sección 2.2 presenta las principales características de los sistemas de reconocimiento facial. Finalmente, la Sección 2.3 describe el problema del reconocimiento facial ante el envejecimiento.

2.1. Introducción

Un sistema de reconocimiento facial es una aplicación que identifica automáticamente una persona en una imagen digital. La identificación es posible mediante un análisis de las características faciales de la persona extraídas de la imagen, y su comparación con una base de datos. Estas características pueden ser extraídas mediante técnicas que sean robustas a la iluminación y pose, entre otras [1].

Los sistemas de reconocimiento facial tienen dos principales funcionalidades que son identificación y verificación. Un sistema para realizar la función de identificación recibe como entrada una imagen de una cara. La imagen es procesada por el sistema como se explica a continuación y el sistema retorna un ranking con las imágenes en la base de datos que tengan la cara mas parecida al de la imagen de entrada. En cambio, un sistema para realizar la función de verificación recibe como parámetro de entrada una imagen y una identificación de una persona (por ejemplo, el número de la cédula de identidad). El sistema procesa la imagen y determina si la cara corresponde a la persona con la identidad declarada. En las figuras 1 y 2 se muestran dos esquemas de sistemas de reconocimiento facial realizando verificación e identificación respectivamente. Por lo tanto, el objetivo de un sistema de reconocimiento facial, dependiendo de cual funcionalidad ejecute, es poder reconocer si una foto pertenece a la identidad declarada (verificación) y/o retornar los candidatos y así poder definir la identidad de la foto entrante (identificación).

La implementación de un sistema de reconocimiento facial se divide en tres módulos [21], como se muestra en la figura 3. El primer módulo, llamado normalización y preprocesamiento, se basa en transformar la imagen de entrada (por ejemplo, en una imagen en blanco y negro), luego extraer los puntos característicos de una cara (ojos, boca, nariz, ceja, etc.) y por último normalizarlas llevándolas a un formato estándar (por ejemplo, llevar la posición de los ojos de una cara de entrada a una posición previamente fijada en

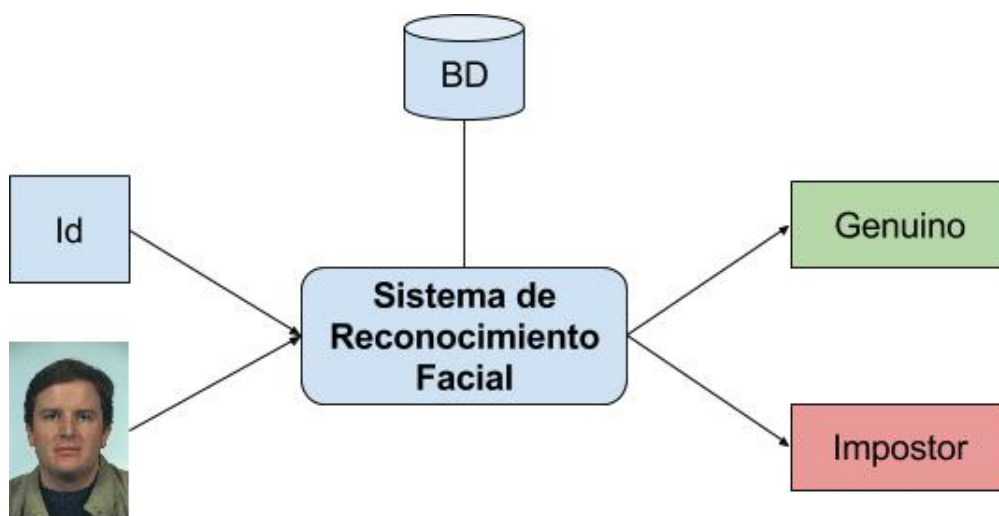


Figura 1: Arquitectura de un sistema de reconocimiento facial aplicando el modo de verificación

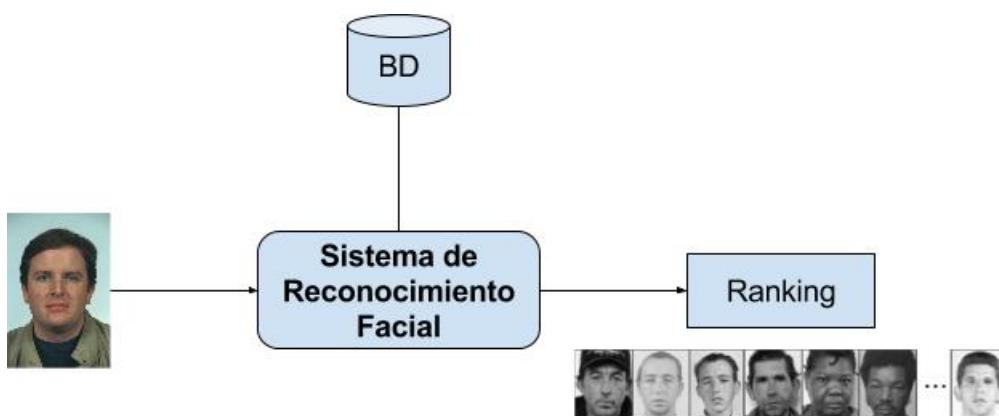


Figura 2: Arquitectura de un sistema de reconocimiento facial aplicando el modo de identificación

el sistema, con el objetivo de que todas las caras queden alineadas en función de los ojos). El siguiente módulo, llamado extracción de características, se encarga de extraer características descriptivas de una imagen para poder utilizarlas por el módulo de clasificación. Las características descriptivas se extraen en puntos específicos de una cara como lo son la nariz, ojos y boca, entre otras, y luego se aplica un algoritmo para obtener el vector característica de la cara. El último módulo, llamado clasificación, procede a realizar la búsqueda de la menor distancia entre el histograma de la foto de entrada

contra los histogramas de las fotos en base de datos.

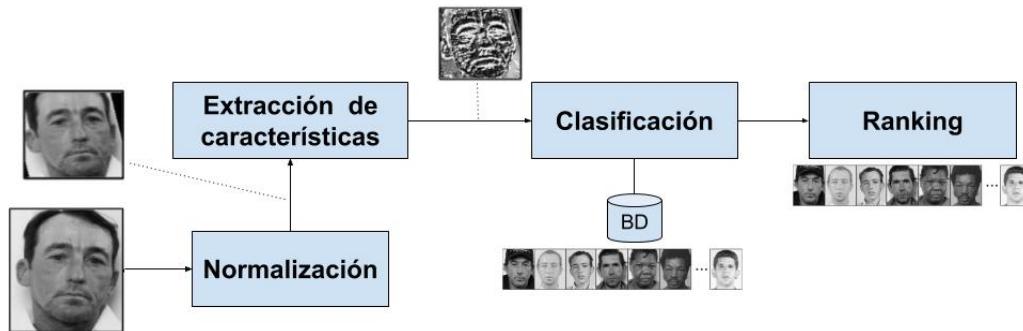


Figura 3: Arquitectura de un sistema de reconocimiento facial. Se muestran los módulos y su salida correspondiente

2.2. Características de los sistemas de reconocimiento facial

El reconocimiento facial presenta grandes desafíos que deben ser enfrentados por los sistemas durante el proceso de reconocimiento de una cara. Las dificultades se presentan desde que se toma la foto hasta que las características de las fotos son extraídas.

Cuando la foto es tomada hay que tener en cuenta el ángulo, la iluminación y los gestos de la persona, entre otros aspectos. Todos los factores mencionados anteriormente pueden degradar la identificación y verificación de un sistema de reconocimiento facial.

El cambio de la iluminación puede afectar a un sistema de reconocimiento facial de tal forma que la foto entrante de una persona puede ser identificada por el sistema como una foto perteneciente a otra persona. Como se menciona en [18], la iluminación es una de las principales causas que degradan el desempeño de sistemas de reconocimiento facial. Existen dos soluciones bien conocidas a este problema. La primera se basa en extraer características invariantes a la iluminación, por lo cual el sistema no dependerá del ambiente en el cual es tomada la foto. La segunda solución se basa en corregir la imagen en la etapa de normalización y preprocesamiento. Una de las técnicas utilizadas para la corrección de la imagen es la ecualización de histogramas, que consiste en aumentar el contraste de la imagen de entrada mediante la manipulación de su histograma. La técnica de ecualización de histogramas permite brindar mayor contraste a las áreas de una foto que tienen bajo

contraste, como se muestra en la figura 4. Un sistema de reconocimiento facial que no implemente ninguna de las técnicas mencionadas, puede que no reconozca como la misma persona a las fotos de la figura 5.

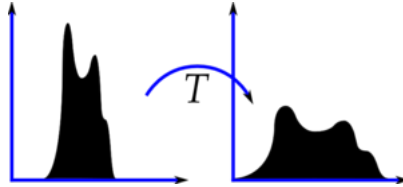


Figura 4: Ecualización de histogramas para incrementar el contraste de fotos en zonas de bajo contraste

Al igual que el cambio en la iluminación, un cambio en la expresión de una persona también perjudica el desempeño del sistema [3] [6]. Una posible solución a este problema es asignar pesos a los descriptores en los puntos más característicos de una cara hallados en la etapa de extracción de características. De esta forma se le pueden asignar pesos a los puntos de la cara con menor influencia en las expresiones y de esta manera hacer que el sistema sea robusto a los cambios de expresiones. En la figura 6 puede observarse dos fotos de la misma persona con diferentes expresiones.

Por otro parte, el envejecimiento es una de las mayores dificultades que afrontan los sistemas de reconocimiento facial, por lo que en la actualidad los investigadores analizan posibles soluciones para resolver este problema [4]. Esta dificultad se presenta cuando las características de una foto no son robustas al envejecimiento y por lo tanto no tienen un buen desempeño. Se debe considerar además que las personas pueden sufrir heridas o alteraciones que afecten la apariencia de su cara con el paso de los años. En la literatura se proponen dos posibles soluciones a este problema. La primera solución consiste en utilizar características que sean robustas al paso del tiempo [4]; la segunda posibilidad es envejecer la cara de la base de datos para así poder compararla con la foto de entrada [11]. Ambos enfoques serán explicados en la siguiente sección. En la figura 7 se observa un ejemplo de envejecimiento.

Por último, el cambio de pose puede afectar el proceso de reconocer los puntos característicos de una cara en la etapa de normalización y preprocesamiento. Por ejemplo, puede suceder que el sistema no encuentre los ojos y por lo tanto no pueda llevar todas las caras a la misma posición, afectando así el desempeño del sistema [9] [12]. Por este motivo, la base de FERET cuenta con fotos de caras tomadas desde diferentes ángulos para medir el desempeño del sistema ante este tipo de problema [16]. En la figura 8 se puede observar un ejemplo de fotos tomadas de diferentes ángulos.



Figura 5: Ejemplo de imágenes de una persona con variación de iluminación



Figura 6: Ejemplo de imágenes de una persona con variación de expresión

Si bien estas dificultades afectan directamente el desempeño de los sistemas de reconocimiento facial, en nuestro proyecto el problema más desafiante es el del envejecimiento pues en la aplicación de DNIC, la pose y las condiciones de iluminación se suponen controladas por el procedimiento de adquisición de las imágenes.

2.3. Envejecimiento

Uno de los principales desafíos que deben afrontar los sistemas de reconocimiento facial es reconocer caras ante la variación del tiempo. En una aplicación de identificación civil, como en DNIC, las tomas consecutivas de imágenes para un mismo sujeto se realizan con intervalos de hasta 10 años.

El desempeño de un sistema de reconocimiento facial, que se explicará con



Figura 7: Ejemplo de imágenes de una persona joven y envejecida



Figura 8: Ejemplo de imágenes de una persona con variación de ángulo (también de pose)

más detalle en la sección 5.1 está definido por su tasa de aciertos identificando una persona por medio de una imagen. Un ejemplo de cómo el envejecimiento en las personas afecta el desempeño de los sistemas de reconocimiento facial se observa en la figura 9, que muestra el desempeño del sistema de reconocimiento facial Faceval [19]. Se puede observar que el desempeño de Faceval se ve afectado en función del tiempo entre foto y foto. También se puede ver que para personas menores a 18 años, el sistema se degrada un 80 % si el

intervalo de tiempo entre foto y foto es de 5 a 10 años. Por último, se puede observar que si el intervalo de tiempo es mayor a 10 años el sistema funciona solamente con un 20 % de precisión y únicamente para personas entre 40 y 49 años, mientras que para los demás casos no llega al 10 %.

Existen dos enfoques para resolver el problema del envejecimiento en el reconocimiento facial. El primer enfoque se basa en envejecer mediante técnicas de procesamiento de imágenes la cara joven de una foto dentro de la base de datos y compararla con la cara envejecida recibida por el sistema como parámetro de entrada, llevando el problema a un problema clásico de reconocimiento facial. El segundo es un enfoque basado en la extracción de características que sean invariante en el tiempo, por lo tanto, se lleva el problema a una clásica situación de un sistema de reconocimiento facial, este enfoque le da más importancia a las características que son invariantes en el tiempo.

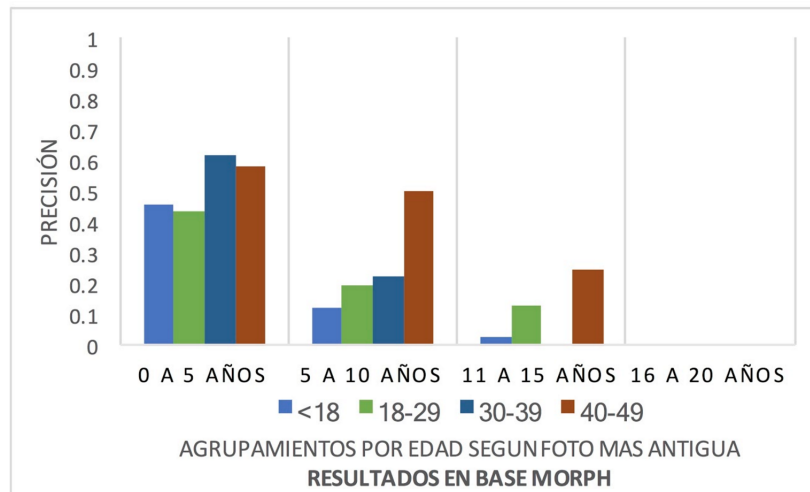


Figura 9: Prueba realizada en el álbum 1 de la base de datos de MORPH con el sistema Faceval [19]

2.4. Trabajos Relacionados

En esta sección se presenta una revisión de los trabajos recientes que han abordado e introducido distintas soluciones, desde la utilización de redes neuronales y operadores para describir texturas, al problema del reconocimiento facial robusto al envejecimiento presentado en este proyecto.

2.4.1. Maximum Entropy Feature Descriptor for Age Invariant Face Recognition

Gong et al. (2015) [10] estudiaron el reconocimiento facial robusto al envejecimiento y propusieron un enfoque para la extracción de características, Maximum Entropy Feature Descriptor (MEFD), y otro enfoque propuesto para la clasificación llamado Identity Factor Analysis (IFA).

Operador MEFD

El operador MEFD resultó de investigar el operador Local Binary Patterns (LBP)[1] con patrones uniformes. Como se profundizará en la sección 3.2.1, este operador se utiliza como descriptor de imágenes y texturas, asignándole a cada pixel un valor, de un total de 256 valores posibles, en función de sus ocho vecinos adyacentes. Existe una variante del operador LBP llamada "LBP con patrones uniformes" que a cada pixel le asigna un valor de un total de 59 valores posibles, lo cual se traduce en una pérdida de información pero una menor dimensión en el vector de características. Como se explica en la sección 3.2.1, los patrones que se descartan, denominados patrones no uniformes, aparecen con una frecuencia de 10 % de todos los patrones según distintos experimentos.

Gong et al. realizaron experimentos sobre 20.000 caras y demostraron que el operador "LBP con patrones uniformes" tiene limitaciones en el problema del reconocimiento facial robusto al envejecimiento. Estas limitaciones surgen porque los patrones no uniformes aparecen con mucha frecuencia dentro de este contexto.

Para superar estas limitaciones, propusieron el operador MEFD, que asigna a cada pixel un valor según sus ocho vecinos adyacentes, utilizando un árbol de decisión, de altura ocho y 59 hojas, previamente entrenado. El objetivo de este árbol es que la frecuencia con que aparece cada valor, de entre los 59 posibles, sea lo más uniforme posible, para que se pierda la menor cantidad de información de la cara, y poder realizar una mejor clasificación.

Framework IFA

El segundo enfoque propuso el uso del framework IFA para la clasificación de características. El enfoque IFA partió de la idea que una característica de una cara con presencia de envejecimiento puede ser descompuesta en los

siguientes 4 vectores: Componente promedio, componente relacionado a la identidad, componente relacionado a la edad, componente de ruido.

Utilizando el algoritmo EM, que se utiliza para hallar los estimadores de máxima verosimilitud de parámetros en modelos que sujetos a variables no observables, se extrae el vector componente relacionado a la identidad y el vector componente promedio. Con estos dos vectores se realiza la clasificación.

En experimentos realizados en la base de datos MORPH 2 [22], utilizando el operador MEFD y el framework IFA, lograron una tasa de desempeño del 92.26 %.

El experimento que se realizó en esta base de datos fue de tomar 10.000 individuos al azar, e identificar la foto más adulta con la foto más joven.

2.4.2. Convolutional Neural Networks for Age-Invariant Face Recognition

Yandong Wen et al. [26] propusieron un sistema de reconocimiento facial robusto al envejecimiento utilizando redes neuronales artificiales convolucionales, en inglés Convolutional Neural Network (CNN).

Una red neuronal artificial está formada por un conjunto de neuronas artificiales interconectadas entre sí. Estas neuronas están distribuidas en distintas capas de neuronas, de manera que las neuronas pertenecientes a una capa se conectan con las neuronas de la capa siguiente de la red. La primera capa de una red neuronal es la entrada de datos, las capas intermedias son las encargadas de realizar las distintas operaciones sobre los datos, y la última capa es la que retorna los resultados. Cada neurona es una unidad de procesamiento de los datos, recibiendo información de la neurona anterior, procesándola según ciertos parámetros entrenables y retornando el resultado a una neurona de la siguiente capa.

Una red neuronal convolucional es un tipo de red neuronal artificial que espera de forma explícita imágenes como entrada, por lo tanto, se adapta mejor a problemas relacionados a imágenes y patrones.

En vez de aplicar redes neuronales sobre el problema del envejecimiento directamente, proponen un nuevo modelo llamado Latent Factor Guided Convolutional Neural Network (LF-CNN) que es un tipo de red neuronal que recibe porciones de una imagen y retorna características de esta invariantes en el tiempo.

Experimentos realizados por los autores en la base de datos MORPH 2 [22], entre otras, lograron una tasa de desempeño del 97.51 %. El experimento que se realizó fue el mismo que en el experimento anterior.

2.4.3. A Study of Face Recognition as People Age

Haibin Ling et al. [18] propusieron un sistema de reconocimiento facial utilizando gradientes orientados a distintas escalas, o como ellos lo llamaron, una pirámide de gradientes orientados.

Al aplicar el operador de gradientes orientados a una imagen, obtenemos una en tonos de gris con una clara distinción de los bordes de las distintas figuras y objetos, ya que el valor gris que le asigna a los píxeles depende del cambio de intensidad los píxeles adyacentes. Este operador es robusto a la iluminación.

Los autores propusieron aplicar este operador a una misma imagen de un rostro en distintos tamaños, para luego concatenarlos. El uso de gradientes orientados ante el cambio de luz en las imágenes fue utilizado exitosamente, por ende, decidieron aplicar esta técnica para afrontar el problema del envejecimiento en las personas. Finalmente a la hora de clasificar utilizan Support Vector Machine (SVM).

En el estudio se menciona que la performance varía en función del tiempo en el cual fueron tomadas las fotos, principalmente los sistemas de reconocimiento facial se ven afectados cuando la diferencia de tiempo entre fotos es mayor a 4 años.

Por último, concluyen que, si bien el envejecimiento es una dificultad para los sistemas de reconocimiento facial, no son mayores a la dificultad con respecto a los cambios en la iluminación o expresiones.

2.4.4. Face Recognition and Retrieval Using Cross-Age Reference Coding With Cross-Age Celebrity Dataset

Chen et al. [4] propusieron un método basado en datos para abordar el problema de reconocimiento facial en distintas edades llamado CARC, y crearon una base de datos de dominio público de celebridades en distintas edades con más de 160.000 fotografías de 2000 celebridades distintas llamada CACD.

Para crear la base de datos CACD y que cumpliera con el requisito de tener muchas fotografías de una misma persona en distintas edades, los autores recolectaron nombres de celebridades cuyas fechas de nacimiento están entre 1951 a 1990. En este período de 40 años, utilizando el sitio www.IMDB.com se quedaron con las 50 principales celebridades populares de cada año, obteniendo 2000 nombres en total. Luego obtuvieron sus fotografías de internet junto a la fecha en que fueron adquiridas las mismas.

El objetivo de CARC es proyectar el vector de características de personas a un espacio de características invariante con la edad, esto significa que dos

vectores de características distintos de la misma persona pero a diferente edad en el espacio de características invariante con la edad serán muy similares. La idea detrás de CARC es multiplicar cada vector por una matriz la cual es creada utilizando los datos de CACD y en la idea de los autores que si dos personas son parecidas lo seguirán siendo con el paso de la edad.

Los autores reportaron que en el mismo experimento utilizado en [10] realizado sobre MORPH 2 se obtuvo una tasa de desempeño de 92.8%.

En este trabajo se decide utilizar el método propuesto en la sección 2.4.4 ya que los resultados reportados pertenecen al estado del arte, al multiplicar cada vector de características por una matriz permite escalar sin mayores dificultades un sistema, se cuenta con la base de datos CACD de dominio público para poder entrenar el algoritmo CARC y es fácil integrar el método CARC en Faceval luego de la etapa de extracción de características.

3. Descripción de la solución

En este capítulo se detallan las principales técnicas que utiliza la solución planteada para cada módulo de un sistema de reconocimiento facial. La Sección 3.1 detalla las técnicas normalización y preprocesamiento. La Sección 3.2 detalla las técnicas de extracción de características. La Sección 3.3 presenta el módulo de clasificación. Por último, en la Sección 3.4 se plantea la propuesta de solución al problema del envejecimiento.

3.1. Normalización y preprocesamiento

La etapa de normalización y preprocesamiento es crucial para convertir las imágenes a un formato previamente definido reduciendo el tamaño, llevando la imagen a escala de grises y a su vez detectando los puntos más importantes de una cara, llamados landmarks. Con respecto a los landmarks, muchos sistemas no son automáticos, es decir, la detección de landmarks se realiza de forma manual, pero en este trabajo se optó por utilizar componentes de software ya entrenadas para realizar esta detección de forma automática.

Asimismo, para incrementar el desempeño de los sistemas de reconocimiento facial, se convierten todas las imágenes a un formato estándar por medio de transformaciones geométricas. Para esto se alinean horizontalmente los ojos y se llevan a una posición (x, y) predefinida. Un ejemplo de esta transformación se observa en la figura 10.



Figura 10: Ejemplo de una imagen de entrada (derecha) y su transformación con la posición de los ojos en la posición (x, y) definida previamente por el sistema (izquierda)

Una vez normalizada y preprocesada la imagen, el sistema de reconocimiento facial procede a la detección de landmarks. Como se mencionó previamente, se utilizaron componentes de software para realizar la detección de los landmarks automáticamente, que se describen a continuación.

STASM [20] es una componente de software implementada en C++ para realizar la localización de landmarks en una cara. Recibe como parámetro de entrada una foto con una cara y retorna la posición de los landmarks (ver figura 11) en la foto.

Por otro lado, Dlib [15] es una componente de software implementada en C++. La componente de software Dlib contiene muchas facilidades, como lo es la detección de landmarks en una cara. Al igual que Stasm, Dlib recibe como parámetro de entrada una foto de una cara y retorna la posición de los landmarks de una foto (ver figura 12). Cabe destacar que Dlib utiliza un predictor entrenado que puede ser utilizado a la hora de detectar los landmarks.



Figura 11: Los 77 landmarks y su numeración utilizando Stasm



Figura 12: Los 68 landmarks y su numeración utilizando Dlib

3.2. Extracción de características

El principal objetivo de la extracción de características es obtener descriptores de la imagen normalizada y retornar un vector que será utilizado en el módulo de clasificación. Para obtener este vector, se realizan operaciones sobre la imagen normalizada y luego se calculan los descriptores que serán el resultado de la extracción de características. Por lo tanto en esta etapa se deben resolver dos problemas: qué información extraer, y cómo utilizar la información extraída. Para extraer la información de una imagen hemos utilizado LBP (Local Binary Patterns) y HDLBP (High Dimensional Local Binary Patterns).

3.2.1. LBP - Local Binary Patterns

El operador LBP es conocido por su gran desempeño en texturas a nivel local [1], utilizado en muchas aplicaciones de tratamiento de imágenes y reconocimiento de patrones.

El funcionamiento de LBP consiste en etiquetar cada píxel de acuerdo con el valor de sus vecinos. Para esto se debe definir qué cantidad de vecinos se tendrán en cuenta para etiquetar el píxel en cuestión. Luego se procede asignar a los vecinos valores de 1 o 0. Estos valores se calculan en función del valor del píxel central, en caso que su valor sea mayor que el valor del píxel vecino, el píxel vecino será etiquetado con un valor de 0, en otro caso, será etiquetado con un valor de 1. Por último, se recorren los vecinos en un orden específico y se forma un número binario de igual largo a la cantidad de vecinos que se tuvo en cuenta. El número binario es una representación de un número entero entre 0 y 2^{n-1} siendo n la cantidad de vecinos considerados.

A su vez, el método LBP tiene la ventaja de ser sencillo y rápido de calcular [1], lo cual es muy importante en aplicaciones que requieran respuestas en tiempo real. También cabe destacar su gran ventaja en cuanto al cambio de iluminación en las diferentes imágenes, ya que el etiquetado de los nodos centrales queda en función del valor del píxel vecino, por lo tanto en una foto que tenga mayor iluminación los vecinos tendrán valores más altos y su valor será más alto, en cambio una foto que tenga poca iluminación los vecinos tendrán valores mas bajos como así también el pixel central. Es por esto que se mantiene la proporción sin importar la iluminación.

En las figuras 13 y 14 se muestra un ejemplo donde se aplica el operador LBP a una porción de una imagen en escala de grises y se consideran los 8 vecinos más cercanos del píxel central para hacer el etiquetado. Se genera un número binario al recorrer los valores resultados de aplicar LBP. En el caso del ejemplo anterior el número binario resulta 01011110 es decir 158 en

notación decimal.

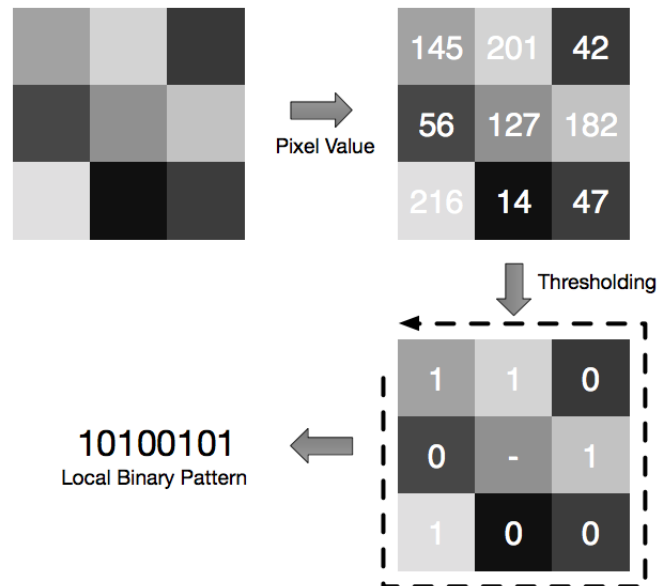


Figura 13: LBP en acción. Inicialmente se tiene un cuadrado con 8 pixeles vecinos al nodo central. Se procede a etiquetar los vecinos con un valor de 0 o 1 calculando cada pixel en caso de ser menor o mayor que el pixel central. Por último se recorren los valores resultados y se genera el número binario.

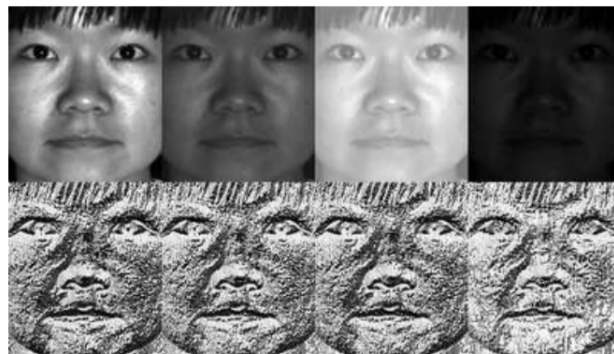


Figura 14: LBP funcionamientos con cambios de iluminación. Si bien la foto cambia de luminosidad, el resultado de aplicar LBP es similar.

A pesar del buen rendimiento de LBP, se implementa el método LBP extendido (ELBP). El método LBP extendido consiste en valorar los N vecinos

distribuidos uniformemente en un círculo de radio R . Luego, en caso de que un vecino no quede contenido completamente dentro del círculo, se obtiene su valor mediante interpolación bilineal. La figura 15 muestra el funcionamiento del operador LBP extendido.

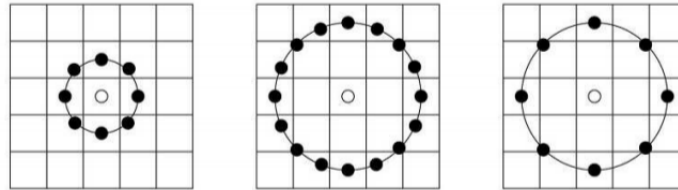


Figura 15: LBP extendido, se aplica un círculo de radio r con centro en el pixel central.

Las palabras obtenidas de LBP pueden ser clasificadas como uniformes o no uniformes. Una palabra LBP es uniforme si su representación binaria no presenta más de 2 cambios de 1 a 0 o de 0 a 1. Por ejemplo la palabra 00100001 no es uniforme ya que presenta tres cambios, mientras que la palabra 00001110 sí lo es ya que presenta dos cambios. Por lo tanto, para una palabra de 8 bits solo existen 58 palabras uniformes. Acorde con que la mayor parte de la información de las texturas en imágenes naturales son codificadas por patrones uniformes, es posible realizar un mapeo no lineal entre las 256 palabras y un conjunto de 59 elementos: 58 dedicados a patrones uniformes y 1 dedicada a los no uniformes, lo cual reduce significativamente la información a tener en cuenta.

Ojala et al. [23] observaron en sus experimentos con imágenes con distintas texturas que los patrones uniformes representan cerca del 90% de todos los patrones obtenidos cuando se utiliza 8 vecinos alrededor de un píxel central al aplicar LBP.

3.2.2. HDLBP - High Dimensional Local Binary Patterns

HDLBP consiste en aplicar LBP a los landmarks a diferentes escalas. La idea es agregar múltiples escalas para así obtener una mejor precisión de la característica de un landmark específico [7].

Para cada landmark, se considera una grilla de 16 bloques ordenados en una matriz de 4×4 y de 10×10 píxeles cada uno para luego aplicar LBP en cada una de los bloques a cinco escalas diferentes como se explica en la sección de LBP. En definitiva la diferencia de HDLBP con LBP es que HDLBP aplica LBP a diferentes escalas.

Con respecto a la implementación de HDLBP, se utilizan 16 landmarks en una grilla de 16 bloques en 5 escalas y 59 palabras por el mapeo lineal entre las 256 palabras, calculando obtenemos un descriptor de dimensión total de 75.520 ($16 \times 16 \times 5 \times 59$).

Como se puede observar, el descriptor tiene una gran dimensión por lo tanto se aplica PCA (Principal Component Analysis) [27] para reducir la dimensión del vector resultante.

3.2.3. PCA - Principal Component Analysis

Dado que los vectores de características obtenidos por medio de HDLBP tienen gran dimensión, tanto el tiempo de clasificación como los espacios de disco y memoria requeridos para almacenarlos se convierten en un problema para un sistema con una base de datos muy grande y que funciona en tiempo real.

Como solución al problema de la dimensión se plantea el uso de PCA. PCA [27] realiza una transformación lineal que escoge un nuevo sistema de coordenadas para el conjunto original de datos en el cual la variación de mayor tamaño del conjunto de datos es capturada en el primer eje ortogonal (llamado el primer componente principal), la segunda varianza más grande es el segundo eje, y así sucesivamente. Para construir esta transformación lineal debe construirse primero la matriz de coeficientes de correlación. Debido a que esta matriz es simétrica, tiene una base completa de vectores propios. La transformación que lleva de las antiguas coordenadas a las coordenadas de la nueva base es precisamente la transformación lineal necesaria para reducir la dimensión de los datos.

PCA fue utilizado en el trabajo de CARC (Cross-Age Reference Coding) [4], que se detalla en la sección 3.4, obteniendo muy buenos resultados. Se observa en dicho trabajo que reduciendo la dimensión de las características HDLBP mediante esta técnica las métricas de performance del sistema no se ven afectadas.

3.2.4. Parches vs landmarks

Como ya se ha explicado anteriormente, LBP se aplica en cuadrados o círculos limitados, por lo tanto se debe definir en que sectores de la imagen se aplicará este método. Existen dos opciones: la utilización de parches que dividen a la imagen en cuadrados de un mismo tamaño o el uso de los landmarks resultado de aplicar el módulo de normalización y preprocesamiento. Ambas opciones se describen a continuación.

Para aplicar LBP a parches de la imagen es necesario que se divida la imagen en cuadrados de igual tamaño. Por ejemplo, supongamos que la imagen se normaliza a 130×130 píxeles, por lo tanto se podrán tomar 10 parches de 13×13 píxeles. Una vez obtenidos los parches se procede a aplicar LBP en cada parche, obteniendo como resultado 10 descriptores, uno para cada parche. Luego se insertan los 10 valores resultados en un vector y se obtiene el vector de características para la foto. El enfoque de parches tiene como desventaja que considera partes de la foto que no aportan información, como por ejemplo el fondo de una imagen.

Por otra parte, el enfoque de la utilización de landmarks [7] asegura que se utiliza información importante de la cara de la imagen de entrada. Para esto se utilizan los landmarks obtenidos en la normalización y preprocesamiento. Se define una ventana donde se aplica LBP, ELBP o HDLPB y se obtiene un descriptor por cada landmark. Para finalizar, se insertan los valores resultados en un vector llamado vector de características, al igual que el enfoque de los parches. El enfoque del uso de landmarks puede también incluir pesos en los landmarks, ponderando los landmarks más importantes en el momento de realizar la clasificación de forma de darle más importancia a aquellos que discriminen mejor las caras.

3.3. Clasificación

La clasificación es el último módulo dentro de un sistema de reconocimiento facial y su objetivo consiste en poder identificar o verificar, en función del modo, una foto en la base de datos.

La clasificación recibe el histograma de la foto de entrada obtenido en el módulo anterior y lo compara con todos los histogramas previamente calculados de las fotos almacenadas en la base de datos. Por último, el módulo de clasificación procede a calcular distancias entre histogramas mediante dos técnicas: Intersección y distancia del coseno.

La distancia por intersección se calcula mediante la ecuación 1, y la distancia del coseno se calcula mediante la ecuación 2 donde H_1 y H_2 son dos histogramas del mismo tamaño n .

$$D(H_1, H_2) = \sum_{i=1}^n \min(h_1(i), h_2(i)) \quad (1)$$

$$D(H_1, H_2) = -1 \times \frac{\sum_{i=1}^n h_1(i) * h_2(i)}{\sqrt{\sum_{i=1}^n h_1(i) * h_1(i)} * \sqrt{\sum_{i=1}^n h_2(i) * h_2(i)}} \quad (2)$$

Una vez obtenida la distancia entre histogramas se ordenan las distancias de menor a mayor. En el modo de verificación, si la distancia es menor a un umbral previamente definido se concluye que el sujeto es genuino, en caso contrario se concluirá que es un impostor. En el modo de identificación, se retorna un ranking con las n fotos con menor distancia a la foto de entrada, donde n es un valor previamente definido en el sistema.

3.4. CARC - Cross-Age Reference Coding

CARC presentada por Chen et al. [4] es una técnica robusta al envejecimiento que basa su estudio en que si dos personas se parecen cuando son jóvenes, tendrán un aspecto parecido cuando envejezcan. Las personas envejecerán de forma tal que seguirán siendo parecidas en un futuro, en otras palabras, el envejecimiento será parecido.

La técnica CARC se puede dividir en cuatro etapas. En la primera etapa se normaliza y se extraen características de las fotos de entrada de una persona. En la segunda etapa se construye el espacio de referencia, que es el espacio sobre el cual se proyecta el vector de características de una imagen para transformarlo en un vector robusto al envejecimiento. Luego, en la tercera etapa se calculan las nuevas características de las fotos en el espacio de referencia. Por último, se obtiene el vector final utilizando la técnica de max-pooling, que consiste en dividir el vector de características en varios sub-vectores, y armar un nuevo vector de características seleccionando el valor máximo de cada sub-vector, con el objetivo de reducir la dimension del vector inicial.

Para obtener la normalización de la foto de entrada se utiliza un extractor de landmarks y una vez obtenida la posición de los ojos, se recorta la foto en un tamaño de $d \times d$, siendo d la distancia entre los ojos. A continuación se aplica la técnica HDLBP explicada en la sección 3.2.2 sobre la imagen normalizada. HDLBP es aplicado sobre cada uno de los 16 landmarks obtenidos por el extractor de características a 5 diferentes escalas, obteniendo así un vector de dimensión 4720. Este vector tiene una gran dimensión, por lo cual se aplica PCA, obteniéndose un vector de dimensión 500.

Para la segunda etapa se debe calcular el espacio de referencia, el cual se construye mediante el cálculo del vector de referencia para cada persona seleccionada dentro del set de referencia, como se observa en la ecuación 3.

$$C_i^{(j,k)} = \frac{1}{N_{i,j}} \sum_{\substack{\text{persona}(x^{(k)})=i \\ \text{año}(x^{(k)})=j}} x^{(k)}, \forall i = 1 \dots n, j = 1 \dots m, k = 1 \dots p \quad (3)$$

Donde $N_{i,j}$ es la cantidad de fotos de una persona i en un año j , n es el numero de personas en el set de referencia, m es el numero de años, p es la cantidad de landmarks y por lo tanto $C_i^{(j,k)}$ calculado en la ecuación 3 es un vector de dimensión $d = 500$ igual que el resultante de HDLBP para la persona i en el año j y el landmark k , el cual representa el promedio de las características de la persona i en un mismo año j , esto es necesario ya que pueden existir más de una foto para una misma persona en un año dado. A su vez, realizando el promedio de las características se reduce el ruido generado por alguna foto en particular.

En la tercer etapa se deben proyectar las características extraídas utilizando la técnica HDLBP sobre el nuevo espacio de referencia. Para esto se considera $C^{j,k}$, una matriz de $d \times n$, siendo d la dimensión de la característica extraída por la técnica HDLBP y n la cantidad de personas en el set de referencia. $C^{j,k}$ se representa de la siguiente forma: $C^{j,k} = [C_1^{j,k} \ C_2^{j,k} \ \dots \ C_n^{j,k}]$, agrupando el set de referencia hallado en la segunda etapa en una única matriz llamada $C^{j,k}$. Luego se define un $\alpha^{(j,k)}$ de dimensión $n \times 1$, el cual representa la relación entre las n personas en el año j para el landmark k . $\alpha_i^{(j,k)}$ será mayor cuanto más cercana sea la característica de prueba $x^{(k)}$ a la persona i , y será cercano a 0 cuanto más diferentes sean las mismas. En la ecuación 4 se presenta una representación resolviendo el problema de mínimos cuadrados con la regularización de Tikhonov.

$$\min_{\alpha^{j,k}} \|x^{(k)} - C^{(j,k)} \alpha^{(j,k)}\|^2 + \lambda \|\alpha^{(j,k)}\|^2 \quad (4)$$

Esta representación plantea hallar los $\alpha^{j,k}$ de forma tal que la norma euclídea de la característica de prueba menos la multiplicación entre $\alpha^{j,k}$ y $C^{(j,k)}$, sumado a un coeficiente $\lambda \|\alpha^{(j,k)}\|^2$ sea mínimo. Esta representación es útil, pero no considera la relación temporal que tienen las representaciones a través de los diferentes años. Es decir, si una persona es parecida a una persona del set de referencia en el año j , esta persona será parecida a la persona del set de referencia en los años adyacentes $j - 1$ y $j + 1$. De forma de considerar la relación temporal de las representaciones en los diferentes años se agrega una matriz constante L definida por la expresión de la ecuación 5 donde m es la cantidad de años considerados:

$$L = \begin{bmatrix} 1 & -2 & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & \dots & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & 1 & -2 & 1 \end{bmatrix} \in R^{(m-2) \times m} \quad (5)$$

Por lo tanto, la matriz L es un operador de suavidad sobre la constante

temporal que aplicada como se muestra en la ecuación 6, genera mas similitud entre los $\alpha^{j,k}$, $\alpha^{j+1,k}$ y $\alpha^{j-1,k}$ resultantes de la minimización. Definiendo una nueva matriz $A^{(k)}$ para todo k siendo k un landmark cualquiera de los 16 considerados. La matriz $A^{(k)}$ es de dimensión $n \times m$ y considerando la matriz L para tener en cuenta la relación temporal entre las personas es posible formular la ecuación 4 de acuerdo a la expresión de la ecuación 6:

$$\min_{A^{(k)}} \sum_{j=1}^n (\|x^{(k)} - C^{(j,k)} \alpha^{(j,k)}\|^2 + \lambda_1 \|\alpha^{(j,k)}\|^2) + \lambda_2 \|LA^{(k)T}\|^2, \forall k \quad (6)$$

Como la variable j varia de 1 a m donde m es el numero de años considerados, se definen dos nuevas matrices para resolver esta representación, $\hat{C}^{(k)}$ definida en la ecuación 7 y $\hat{L}$ definida en la ecuación 8 donde I es la matriz identidad de dimensión $n \times n$:

$$\hat{C}^{(k)} = \begin{bmatrix} C^{(1,k)} & 0 & \dots & 0 \\ 0 & C^{(2,k)} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & C^{(m,k)} \end{bmatrix} \in R^{md \times mn} \quad (7)$$

$$\hat{L} = \begin{bmatrix} I & -2I & I & 0 & \dots & 0 & 0 & 0 \\ 0 & I & -2I & I & \dots & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & I & -2I & I \end{bmatrix} \in R^{(m-2)n \times mn} \quad (8)$$

Luego es posible reescribir la ecuación 6 utilizando las nuevas matrices $\hat{C}^k$ y $\hat{L}$ de acuerdo a la expresión de la ecuación 9

$$\min_{\hat{\alpha}^{(k)}} \|\hat{x}^{(k)} - \hat{C}^{(k)} \hat{\alpha}^{(k)}\|^2 + \hat{\lambda}_1 \|\hat{\alpha}^{(k)}\|^2 + \hat{\lambda}_2 \|\hat{L} \hat{\alpha}^{(k)}\|^2, \forall k \quad (9)$$

La ecuación 9 tiene una solución analítica dada por la función de la ecuación 10, donde se puede observar que la nueva característica será el resultado de multiplicar la característica de HDLBP por la matriz $\hat{P}^{(k)}$.

$$\hat{\alpha}^{(k)} = (\hat{C}^{(k)T} \hat{C}^{(k)} + \hat{\lambda}_1 I + \hat{\lambda}_2 \hat{L}^T \hat{L})^{-1} \hat{C}^{(k)T} \hat{x}^{(k)}, \forall k \quad (10)$$

Escribiendo $\hat{P}^{(k)}$ de la forma que se presenta en la ecuación 11, se puede observar que esta proyección de la matriz se puede obtener previo a calcular la característica de entrada de una foto específica.

$$\hat{P}^{(k)} = (\hat{C}^{(k)T} \hat{C}^{(k)} + \hat{\lambda}_1 I + \hat{\lambda}_2 \hat{L}^T \hat{L})^{-1} \hat{C}^{(k)T}, \forall k \quad (11)$$

Por último para la cuarta etapa, hallados los $\alpha_i^{m,k}$ para todos los k y m , siendo k los landmarks y m los años, se obtiene el α_i^k utilizando max-pooling.

$$\alpha_i^{(k)} = \text{máx}(\alpha_i^{(1,k)}, \alpha_i^{(2,k)}, \dots, \alpha_i^{(m,k)}) \forall i, k \quad (12)$$

A continuación se presenta un pseudocódigo de la construcción del espacio de referencia y la proyección de las características HDLBP de las imágenes, para una base de datos precargada.

Algorithm 1 Pseudocódigo de aplicación de CARC

```
// Extracción de características
global_features = [ ];
for (every photo in database) do
    normalized_img = img.normalize();
    hdlbp_feature = normalized_img.extract_hdlbp();
    global_features << hdlbp_feature.pca_representation();

// Construir C para cada landmark
C = [ ];
for (every landmark) do
    C[landmark] = calculate_c_for_landmark(landmark);

// Construir de la matriz de transformacion Pk
// para cada landmark
lambdaI = 10 * I;
L = build_L_matrix();
P = [ ];
for every landmark do
    Ck = C[landmark];
    lambdas = lambdaI + 10000 * L.transpose() * L;
    intermediate_matrix = (Ck.transpose() * Ck + lambdas).inverse()
    P[landmark] = intermediate_matrix * Ck.transpose();

// Proyectar al espacio de referencia de las características
for (every feature in global_features) do
    for (every landmark) do
        // dup(x) replica el vector x veces
        resized_feature = feature[landmark].dup(years_count);
        new_feature[landmark] = P[landmark] * resized_feature;
    final_representation = new_feature.max_pooling();
    global_features.replace(feature, final_representation);
```

4. Implementación

El presente capítulo describe la implementación del sistema de reconocimiento facial desarrollado en el marco del proyecto. La Sección 4.1 presenta una introducción y los principales casos de uso. La Sección 4.2 presenta la arquitectura del sistema. La Sección 4.3 detalla los principales componentes y su intercomunicación. Por último la Sección 4.4 presenta los módulos principales para el funcionamiento del sistema.

4.1. Descripción general

El sistema desarrollado lleva el nombre de AgeVal y es una plataforma web de verificación e identificación de personas robusta al envejecimiento.

La principal funcionalidad del sistema desarrollado es obtener la identidad de una persona partiendo únicamente de una foto de su rostro, considerando que la persona puede haber envejecido. Se tomó como punto de referencia el sistema de reconocimiento facial Faceval [19] al cual se le realizó un trabajo de reingeniería que incluyó la optimización de performance, modularización del código, conectividad con bases de datos y agregado de nuevos algoritmos y funcionalidades.

La plataforma web cuenta con dos interfaces, la primera para administradores del sistema y la segunda para el acceso a Usuarios registrados por el administrador. Los usuarios pueden acceder a los módulos de verificación e identificación, mientras que los administradores son los encargados de configurar el sistema indicándole al mismo qué algoritmos utilizar y con qué parámetros. La plataforma web también brinda la posibilidad de filtrar la base de datos de imágenes antes de la búsqueda seleccionando el género, siempre y cuando la base precargada cuente con información de género de las personas.

Los administradores del sistema pueden cargar la base de datos con imágenes y datos de otras bases de datos como MORPH [22], FERET [16], CACD [4] y DNIC. También se les brinda la posibilidad de buscar, editar, borrar o dar de alta una persona específica y cargar y establecer distintas configuraciones del sistema.

Para la construcción del sistema se consideraron los siguientes casos de uso, para los cuales se presenta una versión detallada en el anexo A.

Dentro del panel de administración las operaciones que se pueden realizar son: 1) autenticación, 2) alta baja y modificación de usuarios, administradores, personas, fotos y configuraciones, 3) cargado de base de datos externa, 4) extracción de características sobre una base ya cargada y 5) estimación de pesos de parches para cierta configuración.

Dentro de la interfaz web accedida por los usuarios las operaciones que se pueden realizar son: 1) Identificación de una persona, 2) verificación de una identidad y 3) ejecución de experimentos para medir la performance del sistema en distintas bases de datos.

4.2. Arquitectura del sistema

Para la implementación del sistema se estableció un diseño arquitectónico en capas, que se describen en el diagrama de componentes se puede observar en la figura 16. Las capas más altas utilizan servicios que brindan las capas más bajas.

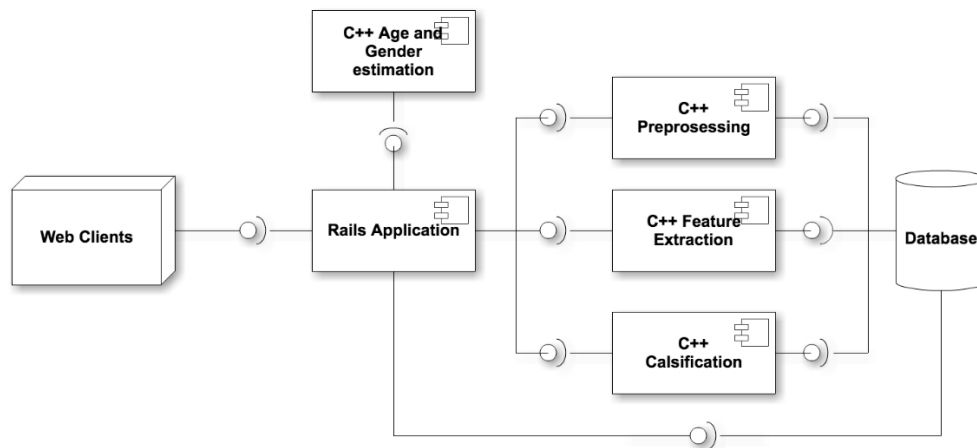


Figura 16: Diagrama de componentes del sistema desarrollado

El sistema de reconocimiento facial implementado en este proyecto está compuesto por cuatro principales subsistemas que son el panel de administración, la plataforma web, un conjunto de módulos desarrollados en C++ y una API encargada de la comunicación con los módulos en C++ desarrollada en Ruby on Rails. A continuación se describen los componentes mencionados.

El panel de administración es el subsistema encargado de proporcionar los casos de uso disponibles para un administrador mediante una interfaz web. Por ejemplo, el panel de administración cuenta con una página de inicio de sesión y luego de que un administrador se autentica se lo dirige a un panel. En el panel el administrador cuenta con diversas pestañas para hacer uso de todas las funcionalidades, ya sea realizar test con determinadas configuraciones, consultar las personas en el sistema y sus respectivas fotos, entre

otras.

La plataforma web es el subsistema encargado de proporcionar los casos de uso disponibles para un usuario mediante una interfaz web. Por ejemplo, cuenta con un panel donde el usuario puede hacer uso de todas las funcionalidades a su disposición, ya sea sacarse una foto o subirla al sistema para luego realizar una búsqueda de las 10 fotos más parecidas en el sistema, verificar si dos fotos pertenecen o no a la misma persona, entre otras.

La aplicación desarrollada en Ruby on Rails es el subsistema encargado de manejar la comunicación entre los componentes web, los módulos en C++ y la base de datos.

El conjunto de módulos desarrollado en C++ forman el subsistema encargado del procesamiento de las imágenes y las tareas con gran costo computacional, como la clasificación y el procesamiento de la base de datos.

La relación entre los distintos subsistemas se presenta en la figura 17.

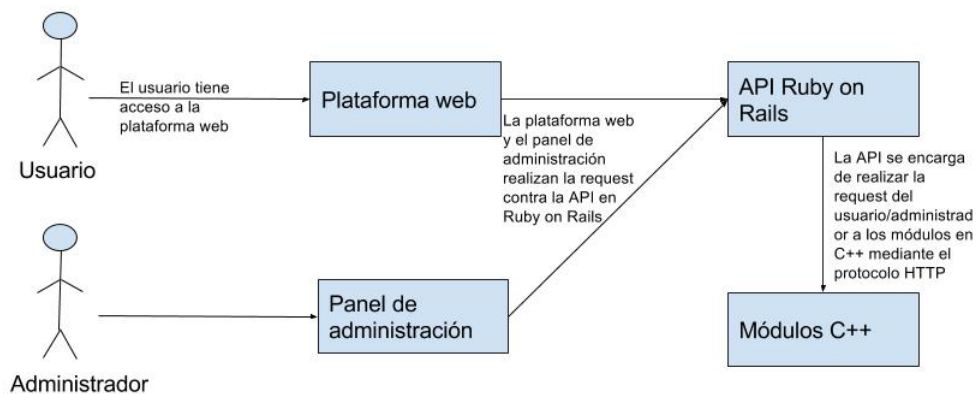


Figura 17: Comunicación entre subsistemas

4.3. Vista de implementación

La implementación del sistema se divide en dos grandes componentes que se comunican entre sí por medio del protocolo HTTP y a su vez comparten la misma base de datos Postgres, como se muestra en la figura 18.

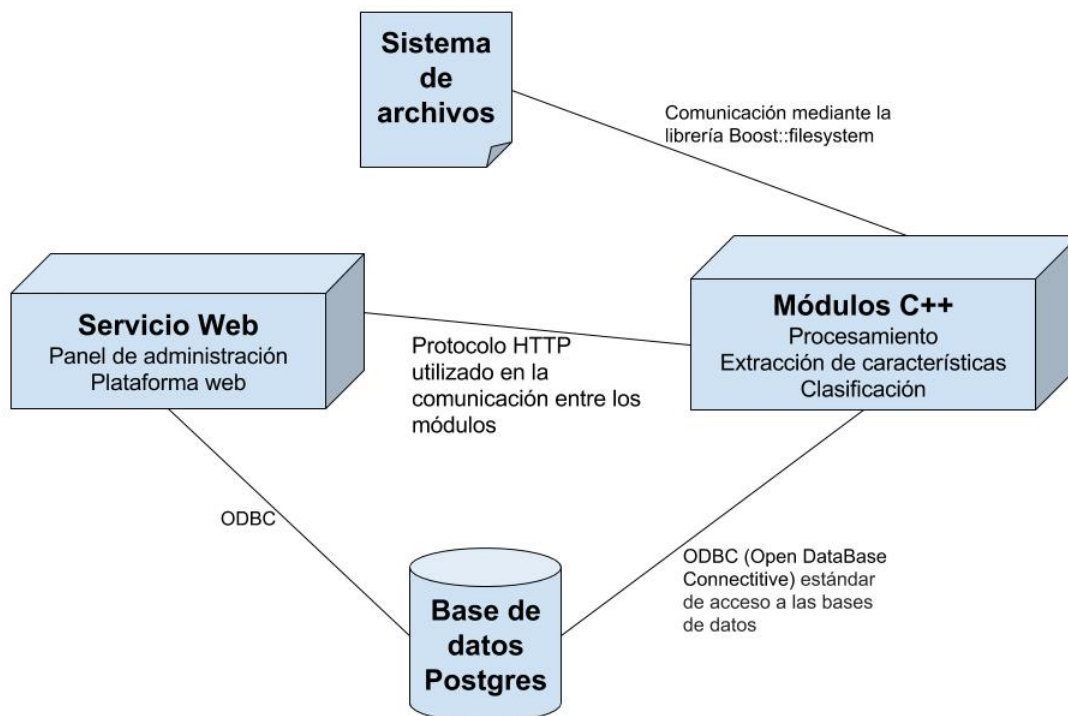


Figura 18: Vista de implementación

4.3.1. Aplicación Ruby on Rails

El frontend del sistema es la aplicación desarrollada en Ruby on Rails, que es la encargada de manejar la comunicación entre los usuarios finales del sistema y el módulo principal de procesamiento (módulo en C++).

Las tecnologías principales utilizadas en el desarrollo del frontend fueron Ruby on Rails, HTML, CSS, Javascript y Postgres SQL. También se utilizaron las siguientes componentes de software para facilitar el desarrollo: PG como interfaz de base de datos, Activeadmin para la creación del panel de administración, Devise para el manejo de sesiones de Administradores, Haml como herramienta para generar código HTML (HyperText Markup Language), Carrierwave como herramienta para el almacenamiento y administración de imágenes.

4.3.2. Base de datos

La base de datos diseñada para el sistema se describe en la figura 19.

Se tomaron en consideración varios aspectos para diseñar la base de datos. Los principales aspectos fueron la capacidad de la base de datos para soportar las tres principales bases externas que pueden ser precargadas (MORPH, FERET y CACD), la flexibilidad ante cambios en los módulos de pre-procesamiento, la extracción de características y clasificación y la capacidad para manejar distintas configuraciones sin necesidad de eliminar todos los datos al cambiarlas.

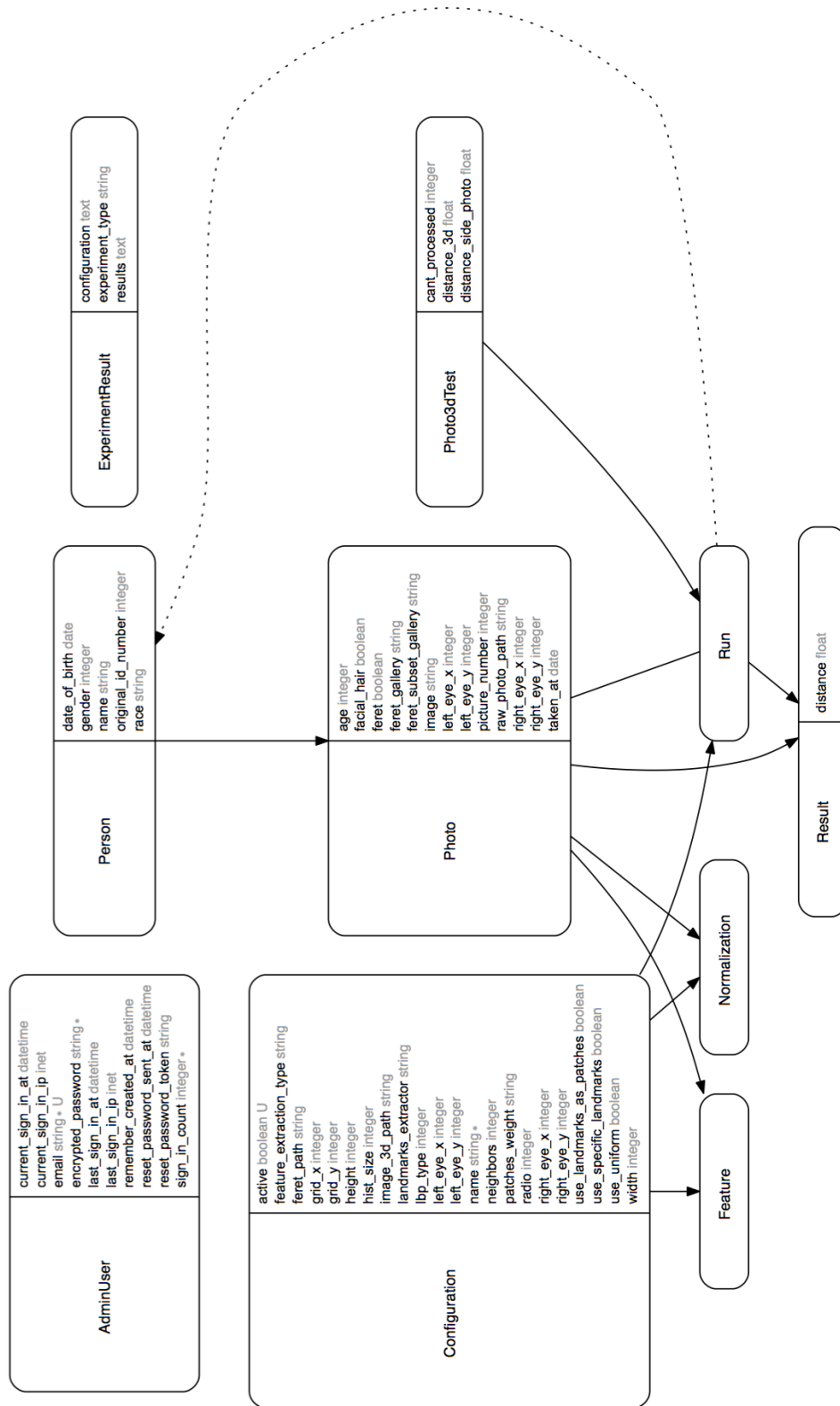


Figura 19: Diseño de base de datos

4.3.3. Sistema de archivos

Si bien el sistema cuenta con una base de datos donde se almacena la mayor parte de la información, también se hace uso del sistema de archivos para almacenar información pesada que se serializa por medio de la biblioteca Boost, así como las imágenes normalizadas en formato JPG.

Boost es una biblioteca para C/C++ que simplifica la comunicación con el sistema de archivos y la serialización de objetos y clases, entre otras.

El sistema crea directorios para guardar la información de forma ordenada. Los directorios creados son normalizaciones, landmarks e histogramas, y se puede encontrar un conjunto de estos directorios por cada configuración presente en el sistema. Existe un directorio de los anteriores por configuración ya que para distintas configuraciones las normalizaciones, landmarks e histogramas de una misma foto pueden variar.

4.3.4. Módulos de C++

Los módulos C++ son uno de los componentes más importante del sistema, ya que brindan herramientas y dan soluciones ya probadas a problemas puntuales. Para el desarrollo se utilizaron las siguientes bibliotecas:

- Boost - serialización de atributos que se almacenan en el sistema de archivos y parseo de JSON.
- Server HTTP - publicación de servicios brindados por el sistema a través de servicios HTTP.
- DLib y Stasm - extracción de landmarks de fotos de caras.
- PThread - paralelización de procesos del sistema utilizando hilos a nivel de sistema operativo.

El sistema construido utilizando estos módulos subsistema es una Interfaz de Programación de Aplicaciones (API, Application Programming Interface) que puede ser consumida por otros sistemas externos por medio de pedidos HTTP. Su interfaz presenta 7 servicios:

- Ejecutar normalización sobre una imagen: recibe como parámetro el identificador de una foto y ejecuta el módulo de normalización sobre la foto recibida.
- Ejecutar extracción de características sobre una imagen: recibe como parámetro el identificador de una foto y ejecuta el módulo de extracción de características sobre la foto recibida.

- Ejecutar clasificación de una imagen: recibe como parámetros el identificador de una foto (obligatorio) y un conjunto de identificadores de otras fotos (opcional) y ejecuta el módulo de clasificación de la foto contra la base de datos, que puede ser filtrada a ciertas fotos por medio del parámetro opcional.
- Calcular parches de la configuración activa: ejecuta el módulo de entrenamiento de pesos de parches para la configuración que se encuentra activa.
- Establecer una configuración como la activa: recibe como parámetro una configuración, la establece como activa y carga en memoria la información necesaria para ejecutar el sistema utilizando dicha configuración.
- Procesar la base de datos: recibe como parámetro un conjunto de identificadores de imágenes y normaliza y extrae características a todas las imágenes cuyo identificador se encuentre en el conjunto.

4.4. Módulos principales

El sistema cuenta con tres módulos principales encargados del preprocesamiento y la normalización de las imágenes, la extracción de características sobre una imagen normalizada y la clasificación. Dichos módulos son completamente independientes, de forma que si se quiere cambiar uno de ellos por uno nuevo, basta con respetar los parámetros de entrada y las salidas del módulo correspondiente para que sea compatible con los otros dos.

4.4.1. Preprocesamiento y normalización

El objetivo de este módulo es llevar todas las imágenes a un formato estándar, transformando la imagen de entrada de forma que los ojos queden alineados, los colores se lleven a escala de grises y el tamaño de la imagen sea siempre el mismo.

Para realizar estas modificaciones el modulo utiliza los siguientes parámetros variables que son establecidos por medio de las configuraciones del sistema:

- height - altura en píxeles de la imagen normalizada.
- width - ancho en píxeles de la imagen normalizada.

- `use_specific_landmarks` - indica al sistema si debe utilizar todos los landmarks del extractor o solo un subconjunto predefinido
- `right_eye_x`, `right_eye_y` - coordenadas de destino del ojo derecho al alinear la imagen.
- `left_eye_x`, `left_eye_y` - coordenadas de destino del ojo izquierdo al alinear la imagen.
- `landmarks_extractor` - indica al sistema qué extractor de landmarks utilizar. Los dos extractores implementados son Stasm y DLib.

4.4.2. Extracción de características

El objetivo de este módulo es extraer descriptores comparables de las imágenes normalizadas retornadas por el módulo anterior. Los tipos de características actualmente implementados en el sistema son ELBP, openfaces y HDLBP, donde cada uno de estos puede ser complementado con el uso de PCA y/o CARC.

Para realizar la extracción de características el módulo utiliza las siguientes configuraciones:

- `feature_extraction_type` - indica al sistema qué descriptores extraer de las imágenes.
- Parámetros de ELBP:
 - `neighbors` - indica al sistema la cantidad de vecinos a considerar por ELBP.
 - `radio` - indica al sistema el radio en pixeles a considerar por ELBP.
 - `lbp_type` - permite indicar el uso de LBP simple o extendido (ELBP).
 - `grid_x` y `grid_y` - indica al sistema el tamaño de ventana en proporción a la imagen.
 - `hist_size` - tamaño del histograma extraído de cada parche.
 - `use_landmarks_as_patches` - indica al sistema si se divide la imagen en parches, o se usan ventanas alrededor de los landmarks únicamente.

4.4.3. Clasificación

El módulo de clasificación es el encargado de comparar los descriptores extraídos en el módulo de extracción de características. La comparación se realiza calculando distancias entre los descriptores con el objetivo de identificar al sujeto de entrada, o poder afirmar si dos imágenes son efectivamente de la misma persona.

El módulo toma como entrada una imagen y un conjunto de imágenes respecto de las cuales clasificar (galería) y como salida guarda en base de datos las referencias a las diez fotos de la galería que son más similares a la imagen de entrada, con sus respectivos indicadores de distancia. El hecho de guardar en base de datos los resultados permite a los usuarios realizar consultas históricas de las clasificaciones.

La única opción configurable con la que cuenta el módulo de clasificación es el peso de los parches, parámetros mediante los cuales el sistema puede dar prioridad a distintas secciones de la imagen, por ejemplo asignándole más peso a los histogramas extraídos de las partes centrales del rostro que al contorno.

El tiempo de procesamiento del módulo de clasificación es el tiempo que transcurre entre que el módulo recibe su entrada y retorna el resultado de la clasificación. Este tiempo es de gran importancia ya que toda operación del sistema ya sea identificación o verificación realizará una comparación entre grandes cantidades de descriptores, si esta comparación es muy lenta el usuario final enfrentará grandes tiempos de espera para obtener los resultados del sistema.

Dada la importancia del tiempo de procesamiento de este módulo para el uso del sistema en tiempo real, se utilizó programación multithreading para reducir el tiempo de ejecución.

La programación multithreading consiste en dividir las tareas en distintos hilos de ejecución de forma de paralelizar tareas independientes. En nuestro sistema las tareas independientes son la comparación entre distintos descriptores, por lo tanto el sistema divide la cantidad total de comparaciones que debe realizar para clasificar un vector en una cantidad predefinida de hilos, luego cada hilo realiza las comparaciones que le fueron asignadas y guarda sus mejores resultados en una variable que todos los hilos comparten para luego seleccionar únicamente los diez mejores en general.

Los resultados experimentales de la optimización mediante el uso de hilos se pueden observar en la Tabla 12 de la sección 5.5.

5. Evaluación experimental

En este capítulo se describe la evaluación experimental del sistema desarrollado. En la sección 5.1 se presentan los conceptos, los indicadores de desempeño y los protocolos utilizados, se introducen los conceptos de base de datos, conjunto galería y de prueba, los distintos indicadores de desempeño para las tareas de identificación y verificación, y se describen brevemente las evaluaciones que se realizan para el sistema. En la sección 5.2 se presentan las bases de datos utilizadas para realizar los experimentos con sus especificaciones y los procedimientos para realizar las evaluaciones en cada una de ellas. En la sección 5.3 se describen las distintas configuraciones utilizadas por Ageval y Faceval en las evaluaciones para obtener los resultados presentados en la sección 5.4. Por último, en la sección 5.5 se presentan resultados de experimentos que evalúan el desempeño computacional del sistema desarrollado.

5.1. Métricas

En el problema del reconocimiento facial es posible distinguir dos tareas: identificación y verificación. Por más que estas tareas se encuentren fuertemente relacionadas, ya que se puede pensar a la tarea de verificación como un caso especial de la tarea de identificación, se necesitan indicadores de desempeño distintos para evaluarlas [14]. También se necesitan muestras biométricas de caras con su identidad para utilizarlas con Ageval y evaluar los resultados que el sistema reporta para poder compararlos con los resultados esperados.

En la sección 5.1.1 se presentan los conceptos de bases de datos que se manejan en el resto del documento, en la sección 5.1.2 se presentan las métricas utilizadas para medir el desempeño del sistema en modo identificación y por último en la sección 5.1.3 se presentan las métricas utilizadas para medir el desempeño del sistema en modo verificación.

5.1.1. Base de datos

Una base de datos de caras, o corpus, está formada por dos tipos de conjuntos de datos:

- Conjunto galería

La galería es un conjunto $G = \{g_1, \dots, g_{|G|}\}$ de muestras biométricas g_i de personas conocidas por el sistema. En el caso de estudio abordado por este proyecto de grado las muestras biométricas son fotografías

y además se conoce la identidad de cada individuo en las distintas fotografías del conjunto G .

- Conjunto de prueba

Un conjunto de prueba es un conjunto $P = \{p_1, \dots, p_{|P|}\}$ de muestras biométricas de personas. En este conjunto se tienen muestras obtenidas en condiciones diferentes a las del conjunto galería. En las muestras pueden existir, por ejemplo, variaciones en la iluminación, la pose o la expresión del individuo, o puede variar el momento de captura de la fotografía respecto a la primera (desde segundos hasta años). Cada muestra biométrica del conjunto de prueba contiene al menos un correspondiente en el conjunto galería G .

A cada muestra biométrica de ambos conjuntos se le puede aplicar la función $id() : muestra_biometrica \rightarrow Identidad$, que dada una muestra biométrica retorna la identidad correspondiente del individuo.

Para evaluar el desempeño de un sistema biométrico se inicializa el sistema con el conjunto galería y se le presenta el conjunto de prueba para que identifique a los individuos de este conjunto.

5.1.2. Identificación

La tarea de identificación trata de encontrar una identidad para una muestra biométrica dada. En los experimentos realizados para medir el desempeño de un sistema de reconocimiento facial, se halla una identidad para cada muestra biométrica del conjunto de prueba. Un sistema debe determinar a cuál muestra p_j del conjunto de prueba corresponde una muestra g_i del conjunto galería. Cuando una muestra p_j es presentada a un sistema, ésta es comparada con la totalidad del conjunto galería. La comparación entre una muestra p_j con cada muestra biométrica g_i genera una calificación de similitud s_{ij} . Cuanto más grande es esta calificación, las muestras biométricas g_i y p_j son más similares (si el clasificador del sistema biométrico utiliza distancias para comparar muestras, se puede obtener la calificación de similitud computando el inverso de la distancia).

Una calificación de similitud s_{ij} es un *acierto* si las muestras biométricas g_i y p_j pertenecen a la misma persona ($id(g_i) = id(p_j)$). El sistema retorna las distintas calificaciones de similitud obtenidas para una muestra p_j en una lista ordenada de forma descendiente según el valor de cada calificación de similitud. Se define la función $pos() : Calificación_de_similitud \rightarrow Z$, que dada una calificación de similitud s_{ij} retorna su posición en la lista creada por el sistema.

Una muestra p_j tiene ranking N si la calificación de similitud s_{ij} con la muestra g_i es un *acierto* y se encuentra entre las primeras N posiciones ($pos(s_{ij}) \leq N$) de la lista retornada por el sistema. Formalmente, se define al ranking N como una función $ranking() : muestrabiometrica \rightarrow N$, dada por la ecuación 13:

$$ranking(s_{ij}) = n : id(g_i) = id(p_j), pos(s_{ij}) \leq n \quad (13)$$

La tasa de desempeño es una métrica que se utiliza para calificar un sistema biométrico en la tarea de la identificación. La tasa de desempeño puede ser calculada utilizando distintos valores para ranking N y esta dada por la ecuación 14.

$$P_I(n) = \frac{|\{p_j : p_j \in P, ranking(p_j) = n\}|}{|P|} \quad (14)$$

5.1.3. Verificación

En la tarea de verificación una persona presenta su fotografía a un sistema y declara una identidad. El sistema tiene que decidir si la muestra biométrica presentada corresponde con la identidad reclamada por la persona.

El sistema computa la calificación de similitud entre la muestra dada y la muestra que contiene el sistema correspondiente a la identidad declarada. El sistema acepta que ambas imágenes corresponden a la misma persona si la calificación de similitud de la muestra presentada es mayor que un umbral de discriminación predefinido τ .

Para medir el desempeño de un sistema en la tarea de verificación son útiles dos indicadores: la tasa de desempeño y la curva ROC. La tasa de desempeño se obtiene calculando todas las calificaciones de similitud y los aciertos entre los elementos del conjunto de prueba P y el conjunto galería G . A continuación, se calcula el cociente del total de los aciertos, considerando el umbral de discriminación τ , sobre la cantidad de elementos de conjunto de Prueba P . Formalmente, la tasa de desempeño se define mediante la ecuación 15.

$$P_v(\tau) = \frac{|\{p_j : s_{ij} \geq \tau, id(g_i) = id(p_j)\}|}{|P|} \quad (15)$$

Las calificaciones de similitud se pueden agrupar en dos clases distintas: verdaderos positivos y falsos positivos. Los verdaderos positivos son muestras identificadas correctamente. En la evaluación de Ageval un verdadero positivo se da cuando la identidad retornada por Ageval es la misma que la de la muestra biométrica. Los falsos positivos son muestras que no son identificadas

correctamente. Un falso positivo sucede cuando la identidad retornada por Ageval no corresponde a la de la muestra biométrica.

Una curva ROC (Receiver Operating Characteristic, Característica Operativa del Receptor) es una representación gráfica que muestra la sensibilidad de un sistema clasificador binario según varía el umbral de discriminación. La utilidad de esta gráfica es la representación del ratio de verdaderos positivos (número de casos que la prueba declara positivos y que son verdaderamente positivo) frente al ratio de falsos positivos (número de casos que la prueba declara positivos y que en realidad son negativos), ya que indica como se comporta el sistema frente a distintos valores de τ .

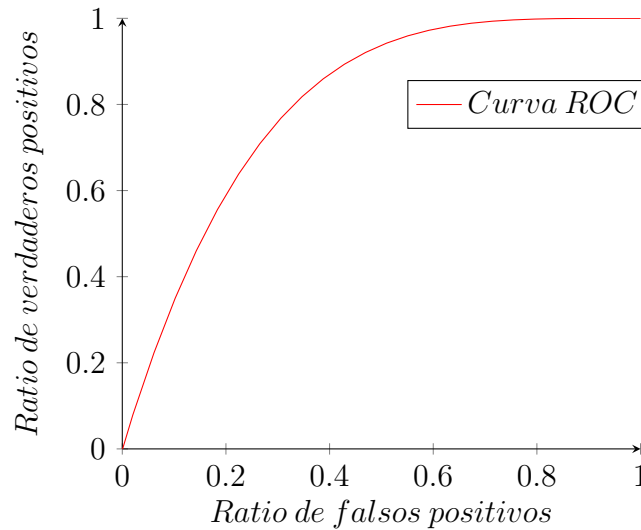


Figura 20: Ejemplo de una curva ROC

Una gráfica de una curva ROC se define con el ratio de falsos positivos y el ratio de verdaderos positivos como los ejes x e y respectivamente. Esta gráfica representa el intercambio entre verdaderos positivos y falsos positivos según el valor de τ . Cuando τ es más cercano a 1, aumentan los verdaderos positivos con el coste de también el aumento de falsos positivos.

5.1.4. Protocolos

Los protocolos son los procedimientos que se deben seguir para evaluar el desempeño de un sistema biométrico en cada base. Los protocolos son definidos por las bases de datos en las cuales se evaluó el sistema y dependen de la base de datos utilizada, ya que evalúan distintas características de un sistema. Por ejemplo, la base de datos MORPH es utilizada para evaluar

cuán robusto a la edad es un sistema biométrico, ya que a diferencia de la base de datos FERET, no posee fotos de caras de tomadas en distintas poses (por ejemplo, de perfil).

5.2. Base de datos y metodologías

En esta sección se presentan las bases de datos que fueron utilizadas para realizar la evaluación experimental de Ageval.

5.2.1. Base de datos FERET

La base de datos FERET es una base de datos de caras promovida por la agencia DARPA (Defense Advanced Research Projects Agency, Agencia de Proyectos de Investigación Avanzada de Defensa) de los Estados Unidos [24]. FERET se creó con el fin de desarrollar técnicas de reconocimiento facial automático para asistir actividades de seguridad e inteligencia. Para construirla se recolectó un conjunto de fotografías de individuos lo suficientemente grande con el objetivo de evaluar los sistemas existentes y otros a desarrollar en el futuro. Las imágenes fueron recopiladas entre agosto de 1993 y julio de 1996 [14] y la base de datos cuenta con un total de 14126 imágenes de 994 individuos en distintas condiciones de iluminación, expresión y tiempo de captura de las fotografías.

En el momento de su publicación, la base de datos FERET se hizo muy popular entre los investigadores por la gran cantidad de personas que incluía y las distintas condiciones controladas, entre otras características.

La base FERET se publicó en dos versiones. La primera versión, lanzada en el año 2001, recibió el nombre “Gray FERET”, y contenía imágenes en escala de grises con una resolución de 256×384 píxeles por imagen. En el año 2003 se publicó una nueva versión de la base que se denominó “FERET Color”, que incluía las siguientes mejoras:

- Las imágenes se incluían en su versión a color.
- Las imágenes incluidas corresponden a las originalmente capturadas y no incluían ningún tipo de pos-procesado, como puede ser la compresión con pérdida para reducir el tamaño de los archivos.
- Se revisó el etiquetado de imágenes y de las coordenadas de los ojos, nariz y boca.
- Se duplicó la resolución de las imágenes llevándolas a un tamaño de 512×768 píxeles.

La base de datos FERET cuenta con los siguientes conjuntos de prueba para su evaluación:

- Fa: es el conjunto galería de imágenes y contiene 994 imágenes de distintas personas.
- Fb: es el conjunto de prueba tomado momentos después de las fotografías del conjunto Fa, donde pueden existir variaciones en la expresión de los individuos respecto al conjunto Fa. El conjunto contiene 992 imágenes de distintas personas.
- Dup1: Conjunto de prueba cuyas fotografías fueron adquiridas hasta 540 días después de la fotografía del conjunto Fa. El conjunto contiene 736 imágenes de distintas personas.
- Dup2: Conjunto de prueba cuyas fotografías fueron adquiridas al menos 540 días después que las fotografías de Fa. El tamaño de este conjunto es de 228 imágenes de distintas personas.

En este trabajo se decidió utilizar la versión “FERET Color” para realizar la evaluación del sistema Ageval. Se toma como conjunto galería el conjunto Fa, que es evaluado frente a Fb y a Dup1 y se calcula la tasa de acierto en los resultados hasta el ranking 5. Esta metodología de evaluación es la misma utilizada para Faceval [19].

5.2.2. Base de datos MORPH

El paso del tiempo afecta la estructura craneofacial de los individuos. Por ejemplo, suceden cambios en la mandíbula en la adultez por el desgaste natural de los dientes [25], o existe pérdida de elasticidad en los músculos que hacen que el rostro se vea modificado. También la pérdida de elasticidad en la piel genera arrugas alrededor de los ojos [8]. La base de datos MORPH surge por la necesidad de evaluar e investigar sistemas biométricos de reconocimiento facial que sean robustos a los cambios mencionados y para ello se necesitan múltiples muestras de caras, tanto masculinas como femeninas, de distintas descendencias y distintas edades.

La primera versión de la base de datos MORPH fue publicada en el año 2003 [22]. Luego se publicaron otras versiones con correcciones y más individuos. La base de datos MORPH es un corpus con caras de múltiples individuos que incluye información sobre ellos, tales como su género, edad, descendencia y estatura.

La base de datos MORPH está organizada en dos álbumes:

- El álbum 1 contiene 1690 fotos digitalizadas de individuos en distintas edades tomadas entre el 26 de octubre de 1962 y el 7 de abril de 1998. De cada foto se conoce la fecha de adquisición y los puntos de referencia.
- El álbum 2 contiene más de 14000 imágenes digitalizadas de 4000 individuos a distintas edades. Las imágenes fueron obtenidas entre mediados de 1990 y 2008.

Las fotos de la base de datos MORPH tienen una dimensión de 400×500 píxeles en ambos álbumes. Las fotos del álbum 1 se encuentran en blanco y negro y las del álbum 2 en color.

La metodología de evaluación en el álbum 1 de MORPH se basa en una publicación realizada por el FBI (Federal Bureau of Investigation, Buró Federal de Investigaciones) en el 2008, que plantea un procedimiento para evaluar un sistema biométrico en la base de datos MORPH [2].

A partir de las imágenes de los individuos del álbum 1, se creó un conjunto T seleccionando 515 personas al azar. Los experimentos para evaluar el impacto de una cara envejecida en el desempeño de reconocimiento facial se llevaron a cabo en un conjunto de individuos T.

A partir del conjunto T, se crearon otros subconjuntos:

- Conjunto M (hombres), con un total de 430 individuos (experimento M).
- Conjunto F (mujeres), con un total de 85 individuos (experimento F).
- Conjunto A (hombres y mujeres afroamericanos), con un total de 403 individuos (experimento A).
- Conjunto W (hombres y mujeres con descendencia europea), con un total de 112 individuos (experimento W).

Para cada conjunto creado, el conjunto galería se formó con las fotos más jóvenes de cada uno de los individuos y el resto de las imágenes de los individuos se utilizaron para crear subconjuntos de imágenes de prueba. Estos subconjuntos se clasificaron según la diferencia de tiempo entre la foto del conjunto de galería y las fotos en el conjunto de prueba. La diferencia de tiempo se tomó en intervalos de 0-5 años, 6 a 10 años, 11 a 15 años y 16 a 20 años. La tasa de acierto utilizada para las pruebas es ranking 1. En la tabla 1 se presentan los detalles acerca de cómo está compuesto el álbum 1.

Tabla 1 Álbum 1 de MORPH - Número de fotos por género y descendencia.

	Afroamericanos	Descendencia europea	Otros	Total
Masculino	1037	365	3	1425
Femenino	216	69	0	285
Total	1253	434	3	1690

La metodología de evaluación en el álbum 2 se basó en utilizar el procedimiento presentado en el trabajo de Li et al. [17]. Se seleccionaron 10000 individuos al azar y se crearon los conjuntos galería y de prueba maximizando la diferencia de tiempo entre las fotografías pertenecientes a los individuos. El conjunto galería esta compuesto por las fotos más jóvenes de los individuos seleccionados y el conjunto de prueba esta formado por las fotos más adultas de los individuos seleccionados. La tasa de desempeño evaluada por el sistema se calcula utilizando ranking 1.

En los conjuntos galería y prueba creados, el promedio de tiempo entre las fotos más jóvenes y mayores de los individuos es de 1,5 años. En la tabla 2 se presentan los detalles acerca de cómo está compuesto el álbum 2.

Tabla 2 Album 2 de MORPH - Numero de fotos por género y descendencia.

	Afroamericanos	Descendencia europea	Otros	Total
Masculino	10283	2650	51	12984
Femenino	1665	550	5	285
Total	11948	3200	56	15204

5.2.3. Base de datos DNIC

La base de datos de DNIC contiene 30549 fotos de 9999 individuos (4949 hombres y 5050 mujeres), en su mayoría con nacionalidad uruguaya. Desde el punto de vista de los objetivos del proyecto, resulta interesante realizar la evaluación del sistema sobre una base de datos con características de la población local del Uruguay.

La base DNIC cuenta con al menos dos fotos de cada individuo, obtenidas entre agosto de 1994 y noviembre de 2016. Las fotografías son todas imágenes en color con resoluciones desde 282×282 hasta 422×422 pixeles, que fueron tomadas en condiciones controladas y no tienen ningun tipo de etiquetado como por ejemplo la posición de los ojos. La cantidad de fotos por rango de edad en la base DNIC se reporta en la tabla 3.

Tabla 3 Base de datos DNIC - Número de fotos por rango de edad.

<18 años	18 a 25 años	25 a 30 años	30 a 35 años	35 a 40 años
104	10952	4609	11626	3258

La metodología de evaluación sobre la base DNIC se basó en utilizar el mismo protocolo de evaluación que MORPH para el álbum 2. El conjunto galería esta compuesto por la foto más joven de cada individuo y el conjunto de prueba por la foto más adulta, obteniendo una distancia de tiempo promedio de 10 años entre las dos imágenes de un mismo individuo en ambos grupos. Para armar estos conjuntos se utilizó la totalidad de la base de datos en lugar de tomar solamente 10000 fotos al azar.

5.3. Configuraciones de Ageval

La configuración adoptada por Ageval en los experimentos realizados utiliza DLib como buscador de puntos de referencia en una cara, HDLBP para extraer las características de los puntos de referencia, PCA para reducir el espacio de características a una dimensión de tamaño 500 y además se utilizan puntos de referencia específicos (solo los de los ojos, nariz y boca) en vez del total de puntos de referencia obtenidos. Además se realizaron experimentos utilizando CARC para normalizar el espacio de características utilizando la configuración previamente mencionada.

Para Faceval, la configuración elegida es la que dio mejores resultados en el trabajo previo [19].

Configuración del modulo de normalización:

- Detección automática: ASM Ojo izquierdo: (44,90)
- Ojo derecho: (44,38)
- Tamaño de imagen: 128×128
- Máscara elíptica

Configuración del módulo extractor de características:

- Vector de características: LBPHS Operador LBP extendido
- Radio: 3
- Vecinos: 8
- Grillado: 9×9

- Cuantización: 256 bins

Configuración del módulo de clasificación:

- Distancia: χ^2
- Asignación de pesos mediante LDA

5.4. Resultados

En esta sección se presentan los resultados obtenidos por Ageval en las distintas bases de datos presentadas así como una comparación con los resultados obtenidos por Faceval.

5.4.1. Resultados para la base FERET

En las Tablas 4, 5 y 6 se presentan los mejores resultados obtenidos por los sistemas Ageval y Faceval. En la Tabla 5 se presentan los resultados de Ageval sin la utilización de CARC, en la Tabla 6 se presentan los resultados de Ageval utilizando CARC y por último en la Tabla 4 se muestran los resultados reportados por Faceval.

Tabla 4 Resultados de Faceval en la base FERET.

	Ranking 1	Ranking 2	Ranking 3	Ranking 4	Ranking 5
Fb	0.987	0.991	0.994	0.994	0.995
Dup1	0.672	0.708	0.734	0.74	0.762
Dup2	0.444	0.53	0.559	0.591	0.614

Tabla 5 Resultados de Ageval sin utilizar CARC en la base FERET.

	Ranking 1	Ranking 2	Ranking 3	Ranking 4	Ranking 5
Fb	0.989	0.993	0.994	0.996	0.996
Dup1	0.717	0.746	0.769	0.791	0.807
Dup2	0.702	0.737	0.772	0.811	0.834

Tabla 6 Resultados de Ageval utilizando CARC en la base FERET.

	Ranking 1	Ranking 2	Ranking 3	Ranking 4	Ranking 5
Fb	0.999	0.999	1.0	1.0	1.0
Dup1	0.872	0.898	0.920	0.929	0.940
Dup2	0.846	0.886	0.912	0.930	0.934

Se puede observar que las tasas de desempeño de Ageval son superiores a los de Faceval para todos los conjuntos (Fb, Dup1 y Dup2) y para todos los rankings considerados (del 1 al 5). También se aprecia que el desempeño de Ageval con CARC supera en un 40 % al de Faceval para el conjunto de prueba Dup2 en Ranking 1. Esto significa que Ageval supera en desempeño a Faceval a la hora de reconocer individuos cuyo intervalo de tiempo entre la fotografía del conjunto galería y la de prueba es mayor a un año y medio.

5.4.2. Resultados para la base MORPH

Siendo MORPH una base de datos con dos álbumes se realizaron pruebas sobre ambos.

■ *Álbum 1*

En las Tablas 7, 8 y 9 se presentan los mejores resultados obtenidos por los sistemas Ageval y Faceval para el álbum 1 de MORPH. En la Tabla 7 se muestran los resultados reportados por Faceval, en la Tabla 8 se presentan los resultados de Ageval sin la utilización de CARC y por último en la Tabla 9 se presentan los resultados de Ageval utilizando CARC. Los resultados en los conjuntos M (hombres), F (mujeres), A (hombres y mujeres afroamericanos) y W (hombres y mujeres con descendencia europea) se encuentran en el apéndice B ya que los resultados obtenidos son similares a los del conjunto T.

Tabla 7 Resultados de Faceval para el conjunto T del álbum 1 de MORPH.

Conjunto Galería, edad de la persona en la foto más joven	Conjunto de prueba de 0 a 5 años	Conjunto de prueba de 5 a 10 años	Conjunto de prueba de 11 a 15 años	Conjunto de prueba de 16 a 20 años
<18	0.410	0.102	0.025	0.000
18-29	0.395	0.179	0.120	0.000
30-39	0.600	0.280	0.000	0.000
40-49	0.667	0.500	0.250	0.000

Tabla 8 Resultados de Ageval para el conjunto T del álbum 1 de MORPH.

Conjunto Galería, edad de la persona en la foto más joven	Conjunto de prueba de 0 a 5 años	Conjunto de prueba de 5 a 10 años	Conjunto de prueba de 11 a 15 años	Conjunto de prueba de 16 a 20 años
<18	0.513	0.334	0.190	0.100
18-29	0.514	0.289	0.219	0.156
30-39	0.638	0.429	0.400	0.500
40-49	0.789	0.286	0.800	1.000

Tabla 9 Resultados de Ageval para el conjunto T del álbum 1 de MORPH utilizando CARC.

Conjunto Galería, edad de la persona en la foto más joven	Conjunto de prueba de 0 a 5 años	Conjunto de prueba de 5 a 10 años	Conjunto de prueba de 11 a 15 años	Conjunto de prueba de 16 a 20 años
<18	0.720	0.589	0.595	0.600
18-29	0.751	0.531	0.538	0.490
30-39	0.860	0.704	0.800	0.667
40-49	0.888	0.714	0.750	1.000

Para todos los experimentos sobre el álbum 1, Ageval obtuvo mejores resultados que Faceval. Se puede observar que Ageval utilizando CARC obtuvo mejores resultados que Ageval sin CARC. También se destaca que se obtuvo un desempeño mayor al 50 % en los conjuntos para las personas cuya primera fotografía fue adquirida antes de los 18 años. Faceval no pudo reconocer ningún individuo cuando la diferencia de años entre la primera foto es mayor a 15 años, lo cual marca que una distancia temporal tan significativa entre la toma de las imágenes disminuye drásticamente el desempeño del sistema. En todos los experimentos la tasa de acierto aumenta a medida que el rango de edades aumenta.

■ *Álbum 2*

La Tabla 5.4.2 reporta una comparación entre los resultados obtenidos por Ageval sin usar CARC, Ageval usando CARC y Faceval sobre el álbum 2 de MORPH.

Tabla 10 Porcentaje de acierto aplicando Ranking 1 de los tres sistemas evaluados para MORPH, álbum 2.

Resultados Ageval	Resultados Ageval - CARC	Resultados Faceval
50.82	69.81	31.71

■ *Comparación con el estado del arte*

La tasa de desempeño de CARC reportada en [4] es de 83.4%, relativamente superior a la reportada por Ageval (69.81 %). A pesar de no reportar resultados similares a los que podemos encontrar en el estado del arte, Ageval logrará mejorar significativamente la tasa de desempeño del sistema anterior Faceval.

La gráfica de la figura 21 muestra las curvas ROC obtenidas por Ageval sin utilizar CARC y utilizándolo. Se puede observar que el ratio de verdaderos positivos para valores cercanos a cero es mayor en la curva con CARC en relación con la que no lo utiliza, por lo tanto, Ageval utilizando CARC tiene un mayor ratio de verdaderos positivos al aumentar el umbral de discriminación τ respecto a cuando no utiliza CARC.

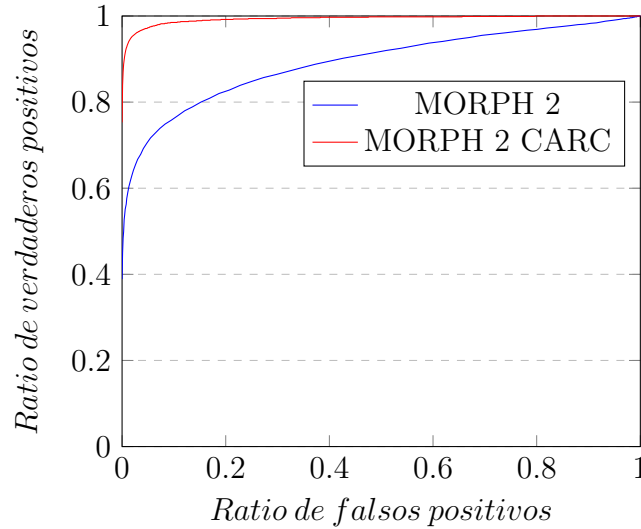


Figura 21: Curva ROC en MORPH 2

5.4.3. Resultados para la base de datos DNIC

La Tabla 11 presenta los resultados obtenidos por Ageval en la base de datos DNIC, utilizando CARC y sin utilizarlo, y una comparación con los

resultados reportados por Faceval.

Tabla 11 Porcentaje de acierto aplicando Ranking 1 sobre la base de datos de la DNIC en los distintos sistemas evaluados.

Resultados Ageval	Resultados Ageval - CARC	Resultados Faceval
16,33	37,92	9,22

Se puede observar que los resultados utilizando CARC mejoran en 28.7% respecto a los reportados por Faceval. Como en el resto de las bases de datos utilizadas en el estudio, la aplicación de CARC mejora de forma significativa los resultados de Ageval.

En la gráfica de la figura 22 podemos encontrar las curvas ROC obtenidas por Ageval sin utilizar CARC y utilizándolo en la base de datos DNIC. Se puede observar que el ratio de verdaderos positivos para valores cercanos a cero es mayor en la curva con CARC respecto a la que no lo utiliza, por lo tanto, Ageval utilizando CARC tiene un mayor ratio de verdaderos positivos al aumentar el umbral de discriminación τ que cuándo no lo utiliza.

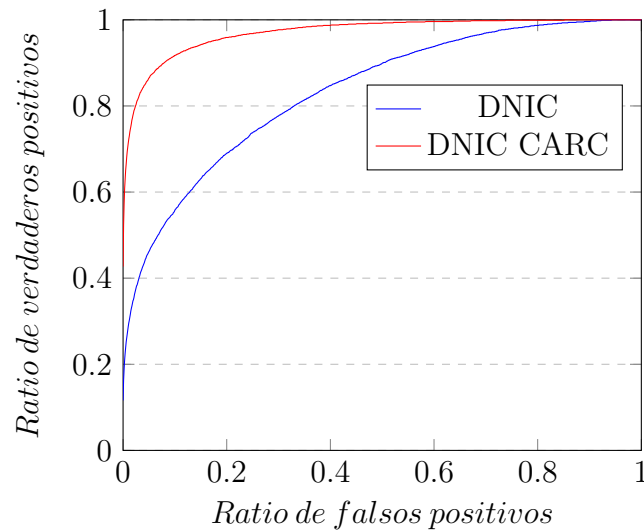


Figura 22: Curva ROC en DNIC

5.5. Análisis de desempeño computacional

El sistema computacional en el que se ejecutó la evaluación experimental de los sistemas de reconocimiento facial es una Macbook Pro Early 2015 que cuenta con las siguientes especificaciones:

- CPU: 2.9 GHz Intel Core i5.
- Memoria: 8 GB 1867 MHz DDR3.
- Almacenamiento: 500 GB Flash Storage.
- Intel Iris Graphics 6100 1536 MB.
- Sistema Operativo: OS X El Capitan, versión: 10.11.5.

La base de datos utilizada para realizar el análisis de desempeño computacional fue MORPH, álbum 2. La cantidad de imágenes iniciales fue de 55.608, de las cuales se utilizaron únicamente 55.159 para el análisis, ya que el sistema no pudo realizar el proceso de extracción de características sobre las 449 restantes debido a que no se pudieron reconocer las caras de los individuos en las mismas.

La primera prueba de desempeño computacional se realizó sobre el módulo de clasificación, independientemente del resto del sistema, variando la cantidad de hilos utilizados. La prueba consiste en clasificar el descriptor de una imagen contra los 55.159 descriptores que se encuentran en la base de datos. Los resultados de eficiencia computacional se pueden observar en la tabla 12.

Para la prueba de desempeño del modulo de clasificación Ageval fue configurado de la siguiente manera:

- Extractor de landmarks: DLIB
- Características: ELBP
- Usando Landmarks específicos

El speedup es una medida de la mejora del rendimiento del sistema al aumentar la cantidad de hilos y se define en la ecuación 16 donde T_N es el tiempo de ejecución con N hilos:

$$speedup = \frac{T_1}{T_N} \quad (16)$$

Tabla 12 Resultados de eficiencia computacional.

Cantidad de hilos	Tiempo de clasificación	Speedup
1	8.30s	1
2	4.23s	1.96
4	3.51s	2.36
8	3.29s	2.52

Como se puede observar en los resultados de la Tabla 12, el tiempo de clasificación disminuye significativamente al aumentar la cantidad de hilos de ejecución pero a su vez el speedup es sublineal.

También se realizaron pruebas de desempeño computacional del sistema en su totalidad, desde que un usuario sube una imagen hasta que puede visualizar los resultados, para evaluar el tiempo de procesamiento, la memoria usada y el espacio en disco necesario para su funcionamiento.

El tiempo de procesamiento para extraer las características y el espacio en memoria y disco necesario varía dependiendo de la configuración usada por el sistema, por lo tanto se tomaron las siguientes tres configuraciones para realizar los experimentos:

Configuración 1:

- Extractor de landmarks: DLIB
- Características: ELBP
- Landmarks específicos

Configuración 2:

- Extractor de landmarks: DLIB
- Características: HDLBP
- Landmarks específicos

Configuración 3:

- Extractor de landmarks: DLIB
- Características: HDLBP + PCA + CARC
- Landmarks específicos

Los resultados se pueden observar en la Tabla 13. Los datos de tiempo de clasificación reportados son en base a promedios de 15 ejecuciones.

Tabla 13 Resultados de experimentos de desempeño computacional

Configuración	Procesamiento	Memoria	Disco	Clasificación
1	1h 30m	1.98 GB	5.6 GB	3.29s $\pm$ 0.32s
2	2h 52m	32.76 GB	86.9 GB	2m 37s $\pm$ 12.1s
3	24h 23m	2,34 GB	23.7 GB	4.5s $\pm$ 0.38s

La diferencia mas grande en el espacio en memoria, espacio en disco y los tiempos de ejecución de la clasificación para la configuración 2, se debe a que los descriptores HDLBP sin el uso de PCA son muy grandes y ademas de ocupar mucho espacio compararlos consume mayor tiempo. La diferencia en tiempo de procesamiento para la configuración 3 se debe a que aplicar PCA y CARC sobre los descriptores es una tarea que requiere mucho tiempo de procesamiento.

6. Conclusiones y trabajo futuro

Esta sección presenta las conclusiones del trabajo realizado en el marco del proyecto y las principales líneas de trabajo futuro.

6.1. Conclusiones

Este proyecto de grado consistió en desarrollar una solución al problema del reconocimiento facial robusto al envejecimiento. Se estudió el problema y se analizaron dos posibles soluciones para afrontarlo; la extracción de características invariantes en el tiempo y el envejecimiento de las caras en la base de datos para poder compararlas con la cara de entrada. Luego de analizar la literatura relacionada al problema, se optó por aplicar el primer enfoque debido a que es más eficiente, ya que el segundo enfoque tiene un mayor margen de error cuando se envejecen las caras de entrada. Para poder llevar a cabo el enfoque de extracción de características invariantes en el tiempo se implementó la técnica HDLBP.

El análisis experimental fue realizado sobre las bases de datos FERET y MORPH utilizadas en la literatura del área, y también se realizó una evaluación del sistema en la base de datos de la DNIC. Las bases FERET y MORPH fueron utilizadas para realizar el análisis experimental ya que otros proyectos de reconocimiento facial reportan resultados sobre estas bases, permitiendo realizar comparaciones en las mismas condiciones entre diferentes sistemas. Los resultados más significativos se obtuvieron en las bases que cuentan con envejecimiento. La base de datos de MORPH cuenta con un envejecimiento promedio por persona de 2.5 años, mientras que la base de datos de DNIC cuenta con un envejecimiento promedio por persona de 10 años. En cambio la base de datos de FERET tiene un intervalo más corto de tiempo entre dos fotos de la misma persona.

Las técnicas implementadas en este proyecto fueron comparadas con el sistema de reconocimiento facial Faceval. Entre los resultados más significativos obtenidos por el sistema implementado en este proyecto se reporta una tasa de desempeño de **86 %** sobre el conjunto de personas de 30-39 años con un intervalo entre fotos de 0 a 5 años sobre MORPH álbum 1, resultado que mejora al sistema Faceval en **26 %**. A su vez, el sistema implementado obtiene una tasa de desempeño de **70.4 %** sobre el conjunto de personas de 30-39 años con un intervalo entre fotos de 5 a 10 años sobre MORPH álbum 1, resultado que mejora al sistema Faceval en **42.4 %**. Debido a los resultados obtenidos se puede observar que Faceval no incorpora técnicas de reconocimiento robustas al envejecimiento, por lo tanto cuanto mayor es el intervalo entre las fotos de entrada, mayor es la mejora de la tasa de desempeño en-

tre el sistema implementado y Faceval. Por otro lado, se obtuvo una tasa de desempeño de **69.81 %** sobre MORPH álbum 2, mejorando así los resultados obtenidos por Faceval en un **38.11 %**. Por último, también se logró mejorar la tasa de desempeño del sistema implementado en este proyecto sobre la base de FERET, obteniendo una tasa de **84.6 %** utilizando ranking 1 y sobre el conjunto de fotos Dup2, mejorando el desempeño de Faceval en un **40.4 %**.

Por otra parte, el sistema implementado en este proyecto, ha mejorado notoriamente los resultados obtenidos por Faceval sobre la base de datos de la DNIC, obteniendo una tasa de desempeño de **37.92 %**, lo cual mejora al sistema Faceval en un **28.9 %**. Este resultado es fundamental para considerar exitoso el proyecto, teniendo en cuenta el envejecimiento de la base de datos de la DNIC.

A partir de los resultados obtenidos se puede concluir que el sistema implementado se desempeña ampliamente mejor ante el envejecimiento que Faceval, aumentando el desempeño en un **36 %** en promedio para todas las bases.

Como otro producto del proyecto, se diseñó e implementó una plataforma web que permite a los usuarios encontrar las caras más parecidas en una base de datos a una cara de entrada, así como también poder verificar si una foto de entrada pertenece a la identificación ingresada por el usuario. Asimismo, el sistema permite cargar distintas bases de datos como FERET, MORPH y CACD de forma automática desde un panel de administración. También es posible ejecutar de forma automática un conjunto de experimentos de evaluación que son un estándar para los sistemas de reconocimiento facial.

Por último, cabe destacar que el proyecto de grado fue presentado en la edición 2016 de Ingeniería de Muestra [13], el evento anual que se organiza en la Facultad de Ingeniería de la Universidad de la República con la Fundación Ricaldoni, donde se presentan proyectos de fin de carrera. En dicha muestra, el proyecto obtuvo el tercer premio del jurado en la categoría “Proyectos de fin de carrera de Ingeniería en Computación”.

6.2. Trabajo futuro

A continuación, se detallan las principales líneas de trabajo futuro que surgen a partir del trabajo realizado en el marco de este proyecto.

Si bien el sistema implementado tiene la opción filtrar por sexo manualmente, es decir, el usuario debe ingresar el sexo de la persona a la que corresponde una foto, una posible mejora es agregar la funcionalidad de poder identificar el sexo automáticamente. Esta funcionalidad podría ser de utilidad para filtrar la base de datos previamente a una búsqueda extensiva sobre todas las personas del sistema. Por otra parte, también se podría agregar la

posibilidad de identificar la edad de la persona de la foto de entrada. Esta funcionalidad podría ser de utilidad en caso de querer implementar el enfoque de envejecer las caras en base de datos, ya que se cuenta con la información sobre la edad en que fue tomada la foto en la base de datos y la edad actual de la persona. Debe tenerse en cuenta que ambas mejoras pueden deteriorar el desempeño del sistema, ya que se introducen errores estimando la edad o el sexo de una persona.

En caso de ingresar una foto de perfil, el sistema no es capaz de reconocer los puntos más importante de la cara y por lo tanto falla en el reconocimiento. Una posible mejora es incluir técnicas que permitan identificar caras que no sean tomadas de frente. También se podría hacer énfasis en reconocer caras con diferentes expresiones. Solucionando estos problemas se puede contar con un sistema completo de reconocimiento facial. Teniendo en cuenta esto, se pudieron haber implementado ambas mejoras pero escapaban el alcance y aplicación de interés para este proyecto.

Por último, podría ser de interés extender el sistema de reconocimiento facial desarrollado para identificación civil a un sistema forense que permita reconocer a partir de imágenes de vídeo o fotos en condiciones no controladas, incluyendo diferencias en pose, iluminación, cantidad de caras y la presencia de oclusiones.

A. Especificación casos de uso

Los actores de los casos de uso son:

1. **Administrador:** Personas que tienen acceso al panel de administración sin poder acceder a la plataforma web.
2. **Usuario:** Personas que utilizan la plataforma web sin poder acceder al panel de administración.

A.1. Especificación de Casos de Uso Críticos en Administración

1. Autenticación

Nombre:	Autenticación del administrador
Objetivo:	Autenticar al Administrador en el panel de Administración
Actores:	Administrador
Precondiciones:	No debe existir una sesión activa en el Panel de Administración.
Descripción:	El Administrador ingresa su email y password, el Sistema valida los datos.
Flujo Normal:	1. El Administrador ingresa su email y password. 2. El Sistema verifica que los datos ingresados sean correctos y retorna un token de autenticación.
Flujos Alternativos:	2A. La combinación email/password proporcionada es incorrecta. - El Sistema despliega un mensaje acorde. - Vuelve a 1.
Post-condiciones:	El Administrador queda autenticado en el sistema.

2. CRUD de Administradores

Nombre:	Alta, modificación y baja de Administradores
Objetivo:	Dar de alta, modificar y eliminar Administradores del Sistema.
Actores:	Administrador
Precondiciones:	Para modificar o dar de alta un Administrador el mismo debe existir en el sistema.
Descripción:	El Administrador puede dar de alta un nuevo Administrador en el sistema. También puede modificarlo o eliminarlo.
Flujo Normal:	1. El Administrador selecciona la acción a realizar: a. En caso de crear un nuevo Administrador i. Ingresa los datos del nuevo Administrador ii. El Sistema da de alta el Administrador b. En caso de modificar un Administrador existente i. Modifica los datos de interés ii. El Sistema actualiza los datos del Administrador seleccionado c. En caso de eliminar un Administrador i. El Administrador selecciona un Administrador. ii. El Sistema da de baja el Administrador previamente seleccionado.
Flujos Alternativos:	
Post-condiciones:	Se da de alta, baja o se modifica un Administrador.

3. CRUD de Personas

Nombre:	Alta, modificación y baja de Personas
Objetivo:	Dar de alta, modificar y eliminar Personas del Sistema.
Actores:	Administrador
Precondiciones:	Para modificar o dar de alta una Persona la misma debe existir en el sistema.
Descripción:	El Administrador puede dar de alta una nueva Persona en el sistema. También puede modificarla o eliminarla.
Flujo Normal:	1. El Administrador selecciona la acción a realizar: a. En caso de crear una nueva Persona i. Ingresa los datos de la nueva Persona ii. El Sistema da de alta la Persona b. En caso de modificar una Persona existente i. Modifica los datos de interés ii. El Sistema actualiza los datos de la Persona seleccionada c. En caso de eliminar una Persona i. El Administrador selecciona una Persona. ii. El Sistema da de baja la Persona previamente seleccionada.
Flujos Alternativos:	
Post-condiciones:	Se da de alta, baja o se modifica una Persona.

4. CRUD de Fotos

Nombre:	Alta, modificación y baja de Fotos
Objetivo:	Dar de alta, modificar y eliminar Fotos del Sistema.
Actores:	Administrador
Precondiciones:	Para modificar o dar de alta una Foto la misma debe existir en el sistema.
Descripción:	El Administrador puede dar de alta una nueva Foto en el sistema. También puede modificarla o eliminarla.
Flujo Normal:	1. El Administrador selecciona la acción a realizar: a. En caso de crear una nueva Foto i. Ingresa los datos de la nueva Foto ii. El Sistema da de alta la Foto b. En caso de modificar una Foto existente i. Modifica los datos de interés ii. El Sistema actualiza los datos de la Foto seleccionada c. En caso de eliminar una Foto i. El Administrador selecciona una Foto. ii. El Sistema da de baja la Foto previamente seleccionada.
Flujos Alternativos:	
Post-condiciones:	Se da de alta, baja o se modifica una Foto.

5. CRUD de Configuraciones

Nombre:	Alta, activación y baja de Configuraciones
Objetivo:	Dar de alta, activar y eliminar Configuraciones del Sistema.
Actores:	Administrador
Precondiciones:	Para activar o dar de alta una Configuración la misma debe existir en el sistema.
Descripción:	El Administrador puede dar de alta una nueva Configuración en el sistema. También puede activarla o eliminarla.
Flujo Normal:	1. El Administrador selecciona la acción a realizar: a. En caso de crear una nueva Configuración i. Ingresa los datos de la nueva Configuración ii. El Sistema da de alta la Configuración b. En caso de activar una Configuración existente i. Selecciona una Configuración ii. El Sistema comenzará a usar la Configuración seleccionada para el procesamiento y clasificación de imágenes. c. En caso de eliminar una Configuración i. El Administrador selecciona una Configuración. ii. El Sistema da de baja la Configuración previamente seleccionada.
Flujos Alternativos:	
Post-condiciones:	Se da de alta, baja o se activa una Configuración.

6. Cargar base externa

Nombre:	Cargar base externa
Objetivo:	Poder cargar diferentes bases de datos en el sistema.
Actores:	Administrador
Precondiciones:	Los archivos deben tener un formato determinado.
Descripción:	El administrador selecciona los archivos requeridos con su formato determinado y el sistema comienza a cargar los datos en la base.
Flujo Normal:	1. El administrador selecciona los archivos de su sistema de archivo y los sube al sistema. 2. El Sistema interpreta los archivos y carga la base de datos.
Flujos Alternativos:	2A. Un archivo no contiene un formato predeterminado. - El Sistema despliega un mensaje acorde. - Vuelve a 1.
Post-condiciones:	Se cargan los datos de la base específica.

7. Extracción de características sobre una base cargada.

Nombre:	Extracción de características sobre una base cargada
Objetivo:	Procesar las imágenes de una base precargada para poder usarlas para identificación y verificación.
Actores:	Administrador
Precondiciones:	El sistema ya debe contar con una Configuración activada.
Descripción:	El Administrador puede procesar una base de datos precargada.
Flujo Normal:	1. El Administrador indica que quiere extraer las características de los datos cargados en la base. 2. El Sistema procesa todas las imágenes. 3. El Sistema informa al usuario las imágenes que no pudieron ser procesadas.
Flujos Alternativos:	
Post-condiciones:	Los datos en la base están prontos para ser usados por los módulos de identificación y verificación.

8. Estimación de pesos de parches para cierta configuración

Nombre:	Estimación de pesos de parches para cierta configuración .
Objetivo:	Estimar de forma automática los mejores pesos para los distintos parches.
Actores:	Administrador
Precondiciones:	Debe existir al menos una Configuración, y una base de datos ya procesada.
Descripción:	El Administrador puede estimar de forma automática los mejores pesos para los distintos parches usados por el sistema.
Flujo Normal:	1. El Administrador selecciona la Configuración e indica que quiere estimar los pesos. 2. El Sistema calcula los mejores pesos y actualiza la Configuración con los mismos.
Flujos Alternativos:	
Post-condiciones:	Se actualizan los pesos de los parches de la Configuración seleccionada.

A.2. Especificación de Casos de Uso Críticos en la Plataforma Web

1. Identificación

Nombre:	Identificación
Objetivo:	Identificar una cara en la base de datos.
Actores:	Usuario
Precondiciones:	
Descripción:	El Usuario ingresa una foto, el Sistema realiza la búsqueda en la base y retorna el top 10 de parecidos.
Flujo Normal:	1. El Usuario ingresa una foto al Sistema. 2. El Sistema realiza la búsqueda y despliega los resultados.
Flujos Alternativos:	1A. El usuario ingresa el filtro de género para podar la búsqueda. - El Sistema aplica los filtros y poda la búsqueda. - Continúa en 2.
Post-condiciones:	El Sistema despliega el rank 10 de las caras más similares.

2. Verificación

Nombre:	Verificación
Objetivo:	Verificar si una foto corresponde a una Persona determinada del Sistema.
Actores:	Usuario
Precondiciones:	La Persona debe existir en el sistema.
Descripción:	El Usuario ingresa una foto y selecciona una posible Persona. El Sistema determina si la Persona es o no la de la foto.
Flujo Normal:	1. El Usuario selecciona una foto del su sistema de archivos. 2. El Usuario selecciona una Persona del Sistema. 3. El Sistema verifica si la Persona seleccionada es la misma que la persona de la foto.
Flujos Alternativos:	
Post-condiciones:	El Sistema despliega un mensaje indicando si la persona y la foto coinciden o no.

3. Ejecución de experimentos de performance

Nombre:	Ejecución de experimentos de performance
Objetivo:	Calcular indicadores de performance sobre determinada base de datos.
Actores:	Usuario
Precondiciones:	La base de datos seleccionada debe estar precargada y procesada en el sistema.
Descripción:	El Usuario selecciona la base de datos y el sistema calcula indicadores de performance para la misma.
Flujo Normal:	1. El Usuario selecciona una base de datos (Morph o Feret). 2. El Sistema calcula indicadores de performance y se los despliega al usuario.
Flujos Alternativos:	
Post-condiciones:	El sistema despliega los indicadores de performance calculados.

B. Resultados en el álbum 1 de MORPH para los conjuntos M, F, A y W

A continuación se detallan los resultados obtenidos en los conjuntos M (Hombres), F (Mujeres), A (hombres y mujeres afroamericanos) y W (hombres y mujeres con descendencia europea) del álbum 1 de MORPH 1.

B.1. Conjunto M (Hombres)

Tabla 14 Resultados de Faceval para el conjunto M del álbum 1 de MORPH.

Conjunto Galería, edad de la persona en la foto mas joven	Conjunto de prueba de 0 a 5 años	Conjunto de prueba de 5 a 10 años	Conjunto de prueba de 11 a 15 años	Conjunto de prueba de 16 a 20 años
<18	0.356	0.193	0.080	0.133
18 a 29	0.293	0.152	0.098	0.148
30 a 39	0.469	0.260	0.067	0.333
40 a 49	0.579	0.545	0.200	0.500

Tabla 15 Resultados de Ageval para el conjunto M del álbum 1 de MORPH.

Conjunto Galería, edad de la persona en la foto mas joven	Conjunto de prueba de 0 a 5 años	Conjunto de prueba de 5 a 10 años	Conjunto de prueba de 11 a 15 años	Conjunto de prueba de 16 a 20 años
<18	0.571	0.319	0.175	0.134
18 a 29	0.526	0.289	0.235	0.179
30 a 39	0.667	0.316	0.429	0.500
40 a 49	0.750	0.286	0.800	1.000

Tabla 16 Resultados de Ageval para el conjunto M del álbum 1 de MORPH utilizando CARC.

Conjunto Galería, edad de la persona en la foto mas joven	Conjunto de prueba de 0 a 5 años	Conjunto de prueba de 5 a 10 años	Conjunto de prueba de 11 a 15 años	Conjunto de prueba de 16 a 20 años
<18	0.781	0.574	0.610	0.455
18 a 29	0.791	0.522	0.590	0.537
30 a 39	0.865	0.632	0.800	0.667
40 a 49	0.867	0.714	0.750	1.000

B.2. Conjunto F (Mujeres)

Tabla 17 Resultados de Faceval para el conjunto F del álbum 1 de MORPH.

Conjunto Galería, edad de la persona en la foto mas joven	Conjunto de prueba de 0 a 5 años	Conjunto de prueba de 5 a 10 años	Conjunto de prueba de 11 a 15 años	Conjunto de prueba de 16 a 20 años
<18	0.250	0.333	1.000	0.400
18 a 29	0.276	0.195	0.167	0.444
30 a 39	0.833	0.444	1.000	-
40 a 49	1.000	-	-	-

Tabla 18 Resultados de Ageval para el conjunto F del álbum 1 de MORPH.

Conjunto Galería, edad de la persona en la foto mas joven	Conjunto de prueba de 0 a 5 años	Conjunto de prueba de 5 a 10 años	Conjunto de prueba de 11 a 15 años	Conjunto de prueba de 16 a 20 años
<18	0.500	0.429	0.500	0.600
18 a 29	0.529	0.355	0.333	0.333
30 a 39	1.000	0.778	1.00	-
40 a 49	1.000	-	-	-

Tabla 19 Resultados de Ageval para el conjunto F del álbum 1 de MORPH utilizando CARC.

Conjunto Galería, edad de la persona en la foto mas joven	Conjunto de prueba de 0 a 5 años	Conjunto de prueba de 5 a 10 años	Conjunto de prueba de 11 a 15 años	Conjunto de prueba de 16 a 20 años
<18	0.364	0.778	1.000	1.000
18 a 29	0.679	0.622	0.308	0.625
30 a 39	0.834	1.000	-	-
40 a 49	1.000	-	-	-

B.3. Conjunto A (hombres y mujeres afroamericanos)

Tabla 20 Resultados de Faceval para el conjunto A del álbum 1 de MORPH.

Conjunto Galería, edad de la persona en la foto mas joven	Conjunto de prueba de 0 a 5 años	Conjunto de prueba de 5 a 10 años	Conjunto de prueba de 11 a 15 años	Conjunto de prueba de 16 a 20 años
<18	0.286	0.188	0.102	0.056
18 a 29	0.265	0.154	0.095	0.102
30 a 39	0.488	0.308	0.182	0.333
40 a 49	0.600	0.556	0.333	1.000

Tabla 21 Resultados de Ageval para el conjunto A del álbum 1 de MORPH.

Conjunto Galería, edad de la persona en la foto mas joven	Conjunto de prueba de 0 a 5 años	Conjunto de prueba de 5 a 10 años	Conjunto de prueba de 11 a 15 años	Conjunto de prueba de 16 a 20 años
<18	0.560	0.325	0.229	0.111
18 a 29	0.509	0.246	0.250	0.154
30 a 39	0.571	0.348	0.600	0.500
40 a 49	0.667	0.200	0.667	1.000

Tabla 22 Resultados de Ageval para el conjunto A del álbum 1 de MORPH utilizando CARC.

Conjunto Galería, edad de la persona en la foto mas joven	Conjunto de prueba de 0 a 5 años	Conjunto de prueba de 5 a 10 años	Conjunto de prueba de 11 a 15 años	Conjunto de prueba de 16 a 20 años
<18	0.659	0.590	0.645	0.538
18 a 29	0.752	0.496	0.557	0.476
30 a 39	0.875	0.667	0.900	0.667
40 a 49	0.846	0.667	0.667	1.000

B.4. Conjunto W (hombres y mujeres con descendencia europea)

Tabla 23 Resultados de Faceval para el conjunto W del álbum 1 de MORPH.

Conjunto Galería, edad de la persona en la foto mas joven	Conjunto de prueba de 0 a 5 años	Conjunto de prueba de 5 a 10 años	Conjunto de prueba de 11 a 15 años	Conjunto de prueba de 16 a 20 años
<18	0.444	0.238	0.000	1.000
18 a 29	0.314	0.136	0.115	0.286
30 a 39	0.500	0.500	0.000	-
40 a 49	0.714	0.500	0.500	1.000

Tabla 24 Resultados de Ageval para el conjunto W del álbum 1 de MORPH.

Conjunto Galería, edad de la persona en la foto mas joven	Conjunto de prueba de 0 a 5 años	Conjunto de prueba de 5 a 10 años	Conjunto de prueba de 11 a 15 años	Conjunto de prueba de 16 a 20 años
<18	0.464	0.357	0.286	1.000
18 a 29	0.571	0.655	0.125	0.500
30 a 39	0.834	0.800	0.200	-
40 a 49	1.000	1.000	1.000	1.000

Tabla 25 Resultados de Ageval para el conjunto W del álbum 1 de MORPH utilizando CARC.

Conjunto Galería, edad de la persona en la foto mas joven	Conjunto de prueba de 0 a 5 años	Conjunto de prueba de 5 a 10 años	Conjunto de prueba de 11 a 15 años	Conjunto de prueba de 16 a 20 años
<18	0.839	0.647	0.545	1.000
18 a 29	0.767	0.722	0.619	1.000
30 a 39	1.000	0.834	0.600	-
40 a 49	1.000	1.000	1.000	1.000

Referencias

- [1] T. Ahonen, A. Hadid y M. Pietikainen. «Face description with local binary patterns: Application to face recognition». *IEEE transactions on pattern analysis and machine intelligence* 28 (12) (2006), págs. 2037-2041.
- [2] A. Albert y K. Ricanek. «The MORPH database: investigating the effects of adult craniofacial aging on automated face-recognition technology». *Forensic Science Communications* 10 (2) (2008).
- [3] M. Alsmadi. «Facial recognition under expression variations.» *International Arab Journal of Information Technology* 13 (1A) (2016), págs. 133-141.
- [4] W. Hsu B. Chen C. Chen. «Face recognition and retrieval using cross-age reference coding with cross-age celebrity dataset». *IEEE Transactions on Multimedia* 17 (6) (2015), págs. 804-815.
- [5] N. Boulgouris, K. Plataniotis y E. Micheli-Tzanakou. *Biometrics: theory, methods, and applications*. Vol. 9. John Wiley & Sons, 2009.
- [6] K. Chang, K. Bowyer y P. Flynn. «Multiple Nose Region Matching for 3D Face Recognition under Varying Facial Expression». *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28 (10) (2006), págs. 1695-1700. ISSN: 0162-8828.
- [7] B. Chen, C. Chen y W. Hsu. *Review and Implementation of High-Dimensional Local Binary Patterns and Its Application to Face Recognition*. Inf. téc. TR-IIS-14-003. Institute of Information Science, Academia Sinica, 2014.
- [8] S. Coleman y R. Grover. «The anatomy of the aging face: volume loss and changes in 3-dimensional topography». *Aesthetic surgery journal* 26 (2006), S4-S9.
- [9] C. Ding, C. Xu y D. Tao. «Multi-task pose-invariant face recognition». *IEEE Transactions on Image Processing* 24 (3) (2015), págs. 980-993.
- [10] D. Gong, Z. Li, D. Tao, J. Liu y Xuelong Li. «A maximum entropy feature descriptor for age invariant face recognition». En: *2015 IEEE Conference on Computer Vision and Pattern Recognition*. 2015, págs. 5289-5297.
- [11] C. Hill, C. Solomon y S. Gibson. «Aging the human face - a statistically rigorous approach». En: *IEEE International Symposium on Imaging for Crime Detection and Prevention*. 2005, págs. 89-94.

- [12] F. Huang, Z. Zhou, H. Zhang y T. Chen. «Pose invariant face recognition». En: *Fourth IEEE International Conference on Automatic Face and Gesture Recognition*. 2000, págs. 245-250.
- [13] Facultad de Ingeniería. *Imágenes Ingeniería de Muestra Facultad de Ingeniería*. [online] <https://www.fing.edu.uy/galerias/area-de-comunicacion/ingenieria-demuestra-2016>. Accessed: 2017-02-13.
- [14] A. Jain K y S. Li. *Handbook of face recognition*. 2005.
- [15] D. King. «Dlib-ml: A machine learning toolkit». *Journal of Machine Learning Research* 10 (2009), págs. 1755-1758.
- [16] Information Technology Laboratory. *The Facial Recognition Technology (FERET) Database*. [online] http://www.itl.nist.gov/iad/humanid/feret/feret_master.html.
- [17] Z. Li, U. Park y A. Jain K. «A discriminative model for age invariant face recognition». *IEEE transactions on information forensics and security* 6 (3) (2011), págs. 1028-1037.
- [18] H. Ling, S. Soatto, N. Ramanathan y D. Jacobs. «A study of face recognition as people age». En: *IEEE 11th International Conference on Computer Vision*. 2007, págs. 1-8.
- [19] L. Di Martino, G. Lema y S. Berchesi. «Faceval: un software de reconocimiento facial» (2012).
- [20] S. Milborrow y F. Nicolls. «Active shape models with SIFT descriptors and MARS». En: *International Conference on Computer Vision Theory and Applications*. 2014, págs. 380-387.
- [21] H. Moon y J. Kim. «Biometrics identification and verification using projection-based face recognition system». En: *International Workshop on Information Security Applications*. 2003, págs. 380-394.
- [22] University of North Carolina Wilmington. *MORPH Non-Commercial Release Whitepaper*. [online] <http://people.uncw.edu/vetterr/MORPH-NonCommercial-Stats.pdf>.
- [23] T. Ojala, M. Pietikäinen y D. Harwood. «A comparative study of texture measures with classification based on featured distributions». *Pattern recognition* 29 (1) (1996), págs. 51-59.
- [24] J. Phillips, H. Moon, S. Rizvi y P. Rauss. «The FERET evaluation methodology for face-recognition algorithms». *IEEE Transactions on pattern analysis and machine intelligence* 22 (10) (2000), págs. 1090-1104.

- [25] M. Taister, S. Holliday y H. Borrmman. «Comments on facial aging in law enforcement investigation». *Forensic science communications* 2 (2) (2000).
- [26] Y. Wen, Z. Li e Y. Qiao. «Latent factor guided convolutional neural networks for age-invariant face recognition» (2016), págs. 4893-4901.
- [27] S. Wold, K. Esbensen y P. Geladi. «Principal component analysis». *Chemometrics and intelligent laboratory systems* 2 (1-3) (1987), págs. 37-52.