



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA ELÉCTRICA

IMPACTO DE LAS MÁSCARAS DE OCLUSIÓN EN LA CLASIFICACIÓN DE GÉNERO EN BASE A IRIS

TESIS PARA OPTAR AL GRADO DE MAGÍSTER EN CIENCIAS DE LA INGENIERÍA,
MENCION ELÉCTRICA

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL ELÉCTRICO

CLAUDIO ALEJANDRO YÁÑEZ MENDOZA

PROFESOR GUÍA:
CLAUDIO PÉREZ FLORES

PROFESOR CO-GUÍA:
JUAN TAPIA FARÍAS

MIEMBROS DE LA COMISIÓN:
CESAR AZURDIA MEZA
PABLO ZEGERS FERNÁNDEZ

Este trabajo ha sido parcialmente financiado por la Comisión Nacional de Investigación Científica y Tecnológica (CONICYT-ANID) a través del FONDECYT Iniciación 11170189

SANTIAGO DE CHILE

2023

RESUMEN DE LA TESIS PARA OPTAR
AL GRADO DE MAGÍSTER EN CIENCIAS
DE LA INGENIERÍA, MENCIÓN ELÉCTRICA,
MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL ELÉCTRICO.
POR: CLAUDIO ALEJANDRO YÁÑEZ MENDOZA
FECHA: 2023
PROF. GUÍA: CLAUDIO PÉREZ FLORES
PROF. CO-GUÍA: JUAN TAPIA FARÍAS

IMPACTO DE LAS MÁSCARAS DE OCLUSIÓN EN LA CLASIFICACIÓN DE GÉNERO EN BASE A IRIS

El iris humano es una fuente importante de rasgos biométricos, como el patrón y el color del iris. Trabajos anteriores han clasificado el género de un sujeto a partir de imágenes del iris, con distintos grados de éxito. Para trabajar con imágenes del iris, éste es segmentado desde una imagen periocular, y luego es normalizado utilizando coordenadas rectangulares. En este proceso, se genera una imagen que enmascara las zonas que no son parte del iris, como párpados, reflejos o pestañas. Estudios anteriores han mostrado que estas máscaras podrían contener información. Si se busca clasificar género usando exclusivamente la textura del iris, la información contenida en las máscaras y potenciales errores al generarlas podrían introducir sesgos en los resultados. En este trabajo se propone una nueva metodología que parea máscaras de género opuesto y asigna una máscara única a cada par, independizándola del género. Adicionalmente, se corrigen manualmente las máscaras para eliminar elementos no enmascarados que no son parte del iris. Se usaron las bases de datos *Gender From Iris* y *ND-CrossSensor-Iris-2013*, las que contienen 3.000 y 29.986 imágenes de iris respectivamente. Al eliminar la información de género en las máscaras, se reduce en un 7.3% la tasa de clasificación de género.

*A remarkable, glorious achievement is just what a long series
of unremarkable, unglorious tasks looks like from far away.*

Tim Urban

Agradecimientos

A mi familia: Cecilia, Juan Claudio, Paulina. Gracias por su apoyo incondicional, por ser un refugio, por todo su amor. Mi mayor fortuna es tenerlos a ustedes.

A Claudia. Gracias *space* por estar conmigo durante todo este proceso, por apoyarme, por escucharme y aconsejarme, y sobre todo por amarme. Me faltan idiomas para expresar todo lo que siento por ti.

A la Tuna Universitaria de la Universidad de Chile y todos sus integrantes, Tunos y (ex-)pardillo(s). La mejor decisión que tomé durante toda mi vida universitaria fue unirme a la Tuna. Gracias por todos los aprendizajes, los viajes, la música, los amigos y hermanos que hice. Gracias por tanto y perdón por tan poco.

A mis amigos de la Universidad: Óscar, René, Camilo, Nicolás y Constantino. Gracias por compartir este difícil proceso, y por darle humor y desahogo al paso por la carrera y el magíster.

A mi primo Dante. Gracias por esas largas conversaciones en que hablamos desde lo más humano a lo más abstracto, por los buenos recuerdos, y por toda la confianza. También a mis primos Maximiliano y Diego, a quienes estimo enormemente, y que siempre recuerdo a pesar de la distancia y la infrecuencia. Gracias a todos los bandi'os.

A mis profesores, Claudio Pérez y Juan Tapia. Gracias por su experiencia y sus enseñanzas, y por la oportunidad de haber investigado con ustedes. Gracias también a la gente del Laboratorio de Visión Computacional: Juan Pablo, Schulz, Diego, Jorge, Daniel, otro Jorge, otro Daniel, y todos los que antaño pasaron por este Lab. Ojalá no hubiera habido pandemia para poder compartir más con ustedes.

A mis amigos de infancia, adolescencia y la vida: Matías, Felipe, Benjamín, y otros que por cosas de la vida no aparecerán nombrados pero que recuerdo con cariño. Gracias por estar ahí conmigo, por toda su buena onda, y por aguantarme durante estos 3.5 largos años (más todos los otros años, y los muchos más que se vienen). Gracias también a mis amigos de rol: Francisco, Joaquín, José, Mona, Nico, Paul. Veintes naturales para todos ustedes.

A mi psicólogo, Carlos. Gracias por ayudar a conocerme mejor a mí mismo, y por entregarme las herramientas para superar esta difícil etapa. Y al Dream Team: Carlos, Iván, Julio, Sabrina y Simón. Gracias por la comprensión durante estos últimos meses de magíster. Es un tremendo agrado trabajar con ustedes.

Este trabajo fue financiado parcialmente por la Comisión Nacional de Investigación Científica y Tecnológica (CONICYT-ANID) a través del FONDECYT Iniciación 11170189.

Tabla de Contenido

1. Introducción	1
1.1. Motivación	1
1.2. Máscaras de oclusión en la clasificación de género en base a iris	3
1.3. Hipótesis	5
1.4. Objetivos	5
1.5. Estructura de la tesis	5
2. Marco teórico	7
2.1. Biometría	7
2.2. Segmentación y normalización del iris	8
2.2.1. Segmentación: localización de la pupila	9
2.2.2. Segmentación: detección de contornos de pupila e iris mediante algoritmo de Viterbi	10
2.2.3. Normalización de iris y máscara	12
2.3. Clasificación de género en base a iris	14
2.4. Redes convolucionales (CNNs)	21
2.4.1. <i>VGGNet</i>	21
2.5. Support Vector Machine	23
3. Metodología	24
3.1. Pareo de máscaras	24
3.1.1. Penalización de pares con alto crecimiento	26
3.1.2. Eliminación de pares con alto crecimiento	26
3.2. Corrección de máscaras de oclusión	27
3.3. Bases de datos	29
3.4. Comparación de desempeño en clasificación de iris	30
3.4.1. Carga de imágenes	31
3.4.2. Partición de datos en entrenamiento y prueba	32
3.4.3. Escalamiento y enmascaramiento de imágenes	32
3.4.4. Experimentos de clasificación de género	33
3.4.4.1. Prueba con características VGG y clasificador SVM	34
3.4.4.2. Prueba con características y clasificador VGG	34
3.4.5. Ajuste de parámetros de modelos	35
3.4.6. Análisis estadístico	36
4. Resultados	37
4.1. Efectos del pareo de máscaras en la distribución de máscaras por género	37
4.1.1. Distribución de pares a distintos umbrales de penalización	38

4.2.	Corrección de máscaras de oclusión	42
4.3.	Resultados de clasificación	42
4.3.1.	Experimento 1: Resultados base	43
4.3.2.	Experimento 2: Efectos de penalizar pares con alto crecimiento	43
4.3.3.	Experimento 3: Efectos de eliminar pares con alto crecimiento	44
4.3.4.	Experimento 4: Efectos de usar máscaras pareadas en clasificación	45
4.3.5.	Experimento 5: Efectos de usar máscaras corregidas en clasificación	46
4.3.6.	Experimento 6: Efectos de usar máscaras pareadas y corregidas en clasificación	47
4.3.7.	Experimento 7: Resultados usando la base de datos ND-CrossSensor-Iris-2013	47
4.4.	Discusión de resultados	49
5.	Conclusiones y trabajo futuro	52
	Bibliografía	54
	Anexos	61
A.	Listado de Acrónimos	61
B.	Resultados en imágenes de iris perioculares	62

Índice de Tablas

3.1.	Parámetros de las bases de datos, los que dan un total de 8 combinaciones posibles.	30
3.2.	Arquitectura de capas <i>fully-connected</i> para la red VGGNet-16.	35
4.1.	Resultados base usando VGG+LSVM y VGG-full.	43
4.2.	Resultados obtenidos por Kuehlkamp y Bowyer en [8].	43
4.3.	Resultados usando máscaras pareadas y eliminando pares sobre cierto crecimiento para las pruebas VGG+LSVM y VGG-full. En negrita se indica el <i>accuracy</i> máxima para cada resolución y prueba.	45
4.4.	Resultados usando máscaras pareadas para las pruebas VGG+LSVM y VGG-full. En negrita se indican el <i>accuracy</i> máximo, la mayor diferencia <i>versus</i> los resultados base, y los p-values por debajo de 0.05.	46
4.5.	Resultados usando máscaras corregidas para las pruebas VGG+LSVM y VGG-full. En negrita se indican el <i>accuracy</i> máximo, la mayor diferencia <i>versus</i> los resultados base, y los p-values menores a 0.05.	46
4.6.	Resultados usando máscaras corregidas y pareadas para las pruebas VGG+LSVM y VGG-full. En negrita se indican el <i>accuracy</i> máximo, la mayor diferencia <i>versus</i> los resultados sin parear, la mayor diferencia <i>versus</i> los resultados sin corregir, y la mayor diferencia <i>versus</i> los resultados base.	47
4.7.	Resumen de resultados usando máscaras corregidas y pareadas para las pruebas VGG+LSVM y VGG-full, con la base de datos GFI. En negrita se indica el <i>accuracy</i> máximo por prueba y resolución.	48
4.9.	Resultados usando la base de datos ND-CrossSensor-Iris-2013 para las pruebas VGG+LSVM y VGG-full. En negrita se indica el <i>accuracy</i> máximo.	48
4.8.	P-values obtenidos con prueba ANOVA bidireccional para cada prueba, resolución y factor. En negrita se indican los p-values menores a 0.05.	49
4.10.	Resultados usando mascararas pareadas y la base de datos ND-CrossSensor-Iris-2013 para las pruebas VGG+LSVM y VGG-full. En negrita se indican el <i>accuracy</i> máximo, la mayor diferencia <i>versus</i> los resultados base, y los p-values menores a 0.05.	50
B.1.	Resultados usando imágenes perioculares de la base de datos GFI para las pruebas VGG+LSVM y VGG-full. La diferencia <i>versus</i> iris normalizado se calculó en base a los mejores resultados promedio obtenidos en esa prueba. En negrita se indica el mayor <i>accuracy</i> , la mayor diferencia <i>versus</i> iris normalizado, y los p-values menores a 0.05.	62

Índice de Ilustraciones

1.1.	Etapas del proceso de reconocimiento en base a iris.	2
1.2.	Diferencias en segmentación en presencia (a, c y e) y ausencia (b, d y f) de maquillaje. Las máscaras de oclusión se muestran en rojo.	4
2.1.	<i>Kernel</i> circular usado en el primer criterio de búsqueda de pupila. [21]	9
2.2.	Mejora en extracción de contornos al usar el filtro anisotrópico. (a) imagen original; (b) mapa de gradientes extraído de la imagen original; (c) contorno extraído de (b) usando el algoritmo de Viterbi; (d) imagen procesada con filtro anisotrópico; (e) mapa de gradientes extraído de la imagen filtrada; (f) contornos extraídos de (e). [32]	11
2.3.	Estimación del punto (X_k^p, Y_k^p) . [31]	13
2.4.	La banda <i>S</i> (en morado) es transformada desde la imagen original (izquierda) hacia la imagen normalizada (a la derecha) siguiendo el modelo de banda elástica. [31]	14
2.5.	Imagen de iris periocular, antes (a) y después (b) de normalizar el iris. La imagen periocular contiene información adicional al iris, como la textura de la piel, carúncula, puntos lagrimales, indicios sobre la estructura ósea de la zona periocular, esclerótica, pupila, cejas y pestañas.	15
2.6.	Configuraciones de la red convolucional (en columnas). La profundidad de las configuraciones aumenta de izquierda (A) a derecha (E), a medida que se añaden más capas (las capas añadidas se muestran en negrita). Los parámetros de las capas convolucionales se denotan como “conv[tamaño del campo receptivo]-[número de canales]”. La función de activación ReLU no es mostrada por brevedad. [22]	22
2.7.	Hiperplano (línea punteada) separando muestras de dos clases (círculos y equis), y formado a partir de vectores de soporte (muestras encerradas en cuadrados). [23]	23
3.1.	Ejemplo de pareo de máscaras. Se usaron máscaras con formas diferentes para ilustrar mejor el proceso. 1) Se forma un par de máscaras de distinto género, indicadas en color rojo y azul. Estas máscaras deben ser similares entre sí para reducir la pérdida de información. 2) Se realiza la operación OR entre las máscaras del par generando una máscara pareada, indicada en color verde. 3) Se aplica la nueva máscara a los iris.	25
3.2.	Iris normalizados y enmascarados usando OSIRIS. Sus máscaras, indicadas en verde, no cubren correctamente los párpados.	27
3.3.	Interfaz del programa para corregir máscaras.	28
3.4.	Dos iris antes (arriba) y después (abajo) de la corrección de máscaras.	29
3.5.	Resumen de metodología usada para comparar desempeño en clasificación de género en base a iris.	31

4.1.	Distribución de máscaras por género antes de formar pares. Las columnas se ven moradas cuando ambos géneros se encuentran superpuestos.	38
4.2.	Distribución de máscaras por género tras formar pares. Las columnas se ven moradas cuando ambos géneros se encuentran superpuestos.	38
4.3.	Crecimiento promedio y número de pares con alto crecimiento, para distintos umbrales de penalización.	39
4.4.	Distribución de pares en base a su crecimiento para distintos umbrales de penalización, usando la resolución 240×20 . Las barras de color anaranjado corresponden a los pares con crecimiento mayor al 10%. Se agregó una etiqueta a los valores menores a 0.5% para facilitar su visualización.	41
4.5.	Distribución de máscaras antes y después de corrección.	42
4.6.	Resultados de clasificación obtenidos usando diferentes umbrales de penalización, a distinta resolución. Las resoluciones con “(correg.)” corresponden a uso de máscaras corregidas, y las que no tienen esta palabra corresponden a máscaras originales. La línea recta que atraviesa los resultados es una regresión lineal del promedio de los resultados por umbral.	44

Capítulo 1

Introducción

1.1. Motivación

El iris humano es una importante fuente de rasgos biométricos. Comparado con otras modalidades biométricas (como la huella dactilar o el rostro), hay múltiples ventajas al usar el iris. Éste posee un complejo patrón de pliegues, surcos y pigmentos, más variado y detallado que la huella dactilar [1]. Esto permite que el iris sea utilizado para identificar con mayor precisión a poblaciones grandes [1]. Este patrón se mantiene relativamente constante durante la mayor parte de la vida de la persona, y dada su naturaleza, es difícil de falsificar [2]. Adicionalmente, no se requiere contacto físico para poder extraer su información. Por estas razones, las tecnologías de reconocimiento de iris han sido utilizadas con éxito en países como los Emiratos Árabes Unidos e India [3].

Para trabajar con imágenes de iris, éste debe ser procesado en varias etapas, como muestra la Figura 1.1. En primer lugar, se debe capturar una imagen periocular, y el iris debe ser extraído de esta mediante segmentación. Una vez que el iris y la pupila han sido encontrados, el iris es normalizado transformándolo a un sistema de coordenadas rectangulares; esta forma de normalizar se conoce como “*Daugman’s rubber sheet model*” [2]. Como parte del proceso de segmentación y normalización, una máscara que cubre las regiones que no son parte del iris es obtenida para eliminar reflejos de luz, áreas cubiertas por párpados o pestañas, maquillaje, etc. [2, 4]. Esta **máscara de oclusión** es importante ya que previene que información que no pertenece al iris interfiera con los procesos de clasificación posteriores [4]. Si se desea efectuar reconocimiento en base a iris, la imagen normalizada debe ser codificada. Esta imagen codificada es comparada con otras imágenes para poder identificar al sujeto.

Tomando en cuenta las ventajas del iris para identificar personas, en trabajos anteriores se ha explorado la posibilidad de clasificar el género¹ de una persona a partir de su iris [5–7]. Esto puede realizarse usando tanto la imagen periocular original como la imagen del iris normalizado [8]. A diferencia de la identificación a partir del patrón del iris, el clasificar el género en base imágenes del iris no permite individualizar a un sujeto. A cambio, se puede

¹ Entiéndase como género biológico, clasificable en hombre o mujer, y excluyendo anomalías cromosómicas.



Figura 1.1: Etapas del proceso de reconocimiento en base a iris.

usar de forma complementaria a otros rasgos biométricos [9]. Por ejemplo, se puede usar para etiquetar y caracterizar demográficamente una base de datos de imágenes de iris que previamente no contenía esta información [10]. Además, conociendo el género de un sujeto se puede reducir el tiempo de identificación, ya que sólo será necesario buscar en una mitad de la base de datos para identificarlo [6].

En esta tesis se abordará la clasificación de género en base a iris usando imágenes de iris normalizadas, dada su estabilidad en el tiempo [2]. Es de interés particular el clasificar el género exclusivamente a partir del iris, su textura y patrones, excluyendo cualquier información que puedan introducir otros elementos. El usar exclusivamente el iris para clasificar género permitiría determinar de forma empírica cuánta información de género hay en la textura del iris. Además, la literatura médica ha mencionado diferencias entre iris y zonas perioculares masculinos y femeninos [11–15].

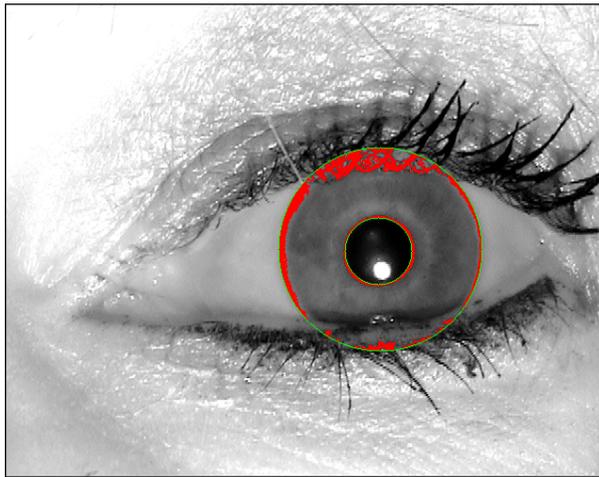
Para efectos de poder compararse con otros trabajos previamente publicados, la métri-

ca usada en este trabajo para evaluar el desempeño en la clasificación de género corresponde al *accuracy* (exactitud). Éste se define como el porcentaje de muestras clasificadas correctamente respecto del total de muestras a clasificar.

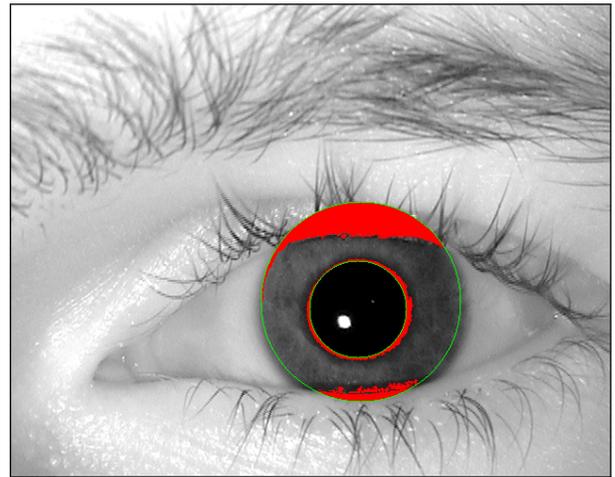
1.2. Máscaras de oclusión en la clasificación de género en base a iris

Se han realizado múltiples trabajos en torno a la clasificación de género usando iris normalizados. Algunos de estos trabajos [5–7, 16–18] reportan resultados cercanos o superiores al 80% de *accuracy*. Otros trabajos [8, 19, 20] en cambio, reportan resultados en torno al 60% de *accuracy*. En el pasado, se han dado posibles explicaciones para esta discrepancia [8, 20]. Algunas de las razones que se mencionan para justificar esta discrepancia son el uso de particiones de entrenamiento y prueba no separadas por sujeto, la presencia de información que no corresponde al iris, y la utilización de sólo una partición de prueba.

El uso de máscaras de oclusión en la clasificación de género en base a iris es necesario, como se mencionó anteriormente, ya que elimina información que no pertenece al iris [4]. Sin embargo, si lo que se busca es clasificar usando exclusivamente la textura del iris, las máscaras de oclusión pueden introducir sesgos en los resultados ya que éstas podrían estar correlacionadas con el género. Anteriormente se ha mostrado que la presencia de maquillaje influye en la generación de las máscaras [20]. Por ejemplo, en la Figura 1.2 se contrasta el proceso de segmentación y normalización de una imagen con maquillaje y una imagen sin maquillaje. La presencia de maquillaje en el párpado inferior ocasionó una segmentación incorrecta de esa zona, y posteriormente, se logra ver el maquillaje en el iris normalizado, aún después de ser enmascarado. Esto no sucede con la imagen sin maquillaje. También se ha mostrado que, usando sólo las máscaras de oclusión sin información de iris, es posible lograr resultados cercanos al 60% de *accuracy* [20]. Esto implicaría que los resultados obtenidos no sólo reflejan la información de género presente en el iris sino que también en las máscaras.



(a) Imagen periocular segmentada, con maquillaje.



(b) Imagen periocular segmentada, sin maquillaje.



(c) Imagen normalizada, con maquillaje.



(d) Imagen normalizada, sin maquillaje.



(e) Imagen normalizada enmascarada, con maquillaje aún visible.



(f) Imagen normalizada enmascarada, sin maquillaje.

Figura 1.2: Diferencias en segmentación en presencia (a, c y e) y ausencia (b, d y f) de maquillaje. Las máscaras de oclusión se muestran en rojo.

Otro sesgo que puede ser introducido indirectamente por las máscaras de oclusión se debe que a las máscaras generadas automáticamente no cubren correctamente las zonas que no contienen textura del iris. Esto podría ocasionar que elementos como el maquillaje, los párpados o las pestañas, influyan en los resultados de clasificación de género basada en iris [8, 20]. En caso de que estas zonas que no son parte del iris estuvieran presentes, los resultados reflejarían no sólo la información de género presente en el iris, sino también la información presente en dichas zonas.

Dado que las máscaras pueden introducir sesgos, y que han habido discrepancias en los resultados obtenidos en la literatura reciente, es necesario diseñar una metodología que permita medir y eliminar el impacto de las máscaras de oclusión en la clasificación de género basada en iris. De este modo se clasificaría usando exclusivamente la información presente en la textura del iris, independizándose de la información de las máscaras. Para esto, se propone un nuevo método denominado pareo de máscaras, que asegura que las distribuciones de máscaras de oclusión son iguales en ambos géneros. Esto permitiría eliminar la información de género presente en las máscaras. El método consiste en parear cada máscara masculina con una máscara femenina que sea similar, y luego generar una nueva máscara a partir de la unión de las dos máscaras originales. El método de pareo reduce la cantidad de información de iris que ese pierde al generar la máscara nueva.

Adicionalmente, se corrigieron manualmente las máscaras generadas automáticamente, lo que permite estudiar el efecto de los errores en las máscaras generadas. Ambos factores, los pares de máscaras y las máscaras corregidas, entregan una nueva perspectiva en el efecto de las máscaras de oclusión en la clasificación de género en base a iris.

1.3. Hipótesis

Se proponen las siguientes hipótesis respecto de la clasificación de género en base a iris:

- Las pestañas, párpados, y otros elementos presentes en las imágenes normalizadas de iris contienen información que un clasificador puede aprender a usar para predecir género.
- Si la distribución de las máscaras de oclusión no es idéntica en ambos géneros, esta información puede ser usada por un clasificador, haciendo que los resultados reflejen información de género presente en las máscaras de oclusión además del iris.
- El *accuracy* en clasificación obtenido al eliminar la información de género presente en las máscaras de oclusión sería significativamente inferior a la obtenida sin eliminar dicha información.

1.4. Objetivos

El objetivo general de esta tesis es definir una metodología para eliminar la información de género presente en las máscaras de oclusión y en los elementos que no son parte del iris; y medir el impacto de las máscaras de oclusión en la clasificación de género en base a iris. Los objetivos específicos para lograrlo son:

1. Desarrollar una metodología que permita eliminar la información de género presente en las máscaras de oclusión en la clasificación de género en base a iris.
2. Corregir manualmente las máscaras para bloquear las estructuras que no corresponden a la textura del iris.
3. Usar máscaras corregidas y la metodología ya mencionada en la clasificación de género en base a iris.
4. Medir el impacto de las metodologías diseñadas y su nivel de significancia en la clasificación de género usando pruebas estadísticas.

1.5. Estructura de la tesis

La presente tesis se estructura de la siguiente forma. En el Capítulo 1 se introduce el trabajo realizado, explicando su motivación, su hipótesis y los objetivos para comprobarla.

En el Capítulo 2 se presenta la fundamentación teórica para entender el trabajo realizado. Además, se describe el algoritmo usado para segmentar y normalizar iris, y se detalla el estado del arte. En el Capítulo 3 se describe la metodología a seguir para comprobar la hipótesis planteada y cumplir los objetivos, incluyendo el método propuesto de pareo de máscaras y la metodología de clasificación de género en base a iris. En el Capítulo 4 se muestran los resultados obtenidos tras aplicar la metodología descrita en el Capítulo 3, incluyendo su análisis estadístico y discusión. Finalmente, en el Capítulo 5 se concluye con respecto a los objetivos e hipótesis planteadas en la Introducción, y se proponen mejoras metodológicas y trabajo a futuro.

Capítulo 2

Marco teórico

En este capítulo se presenta el trasfondo teórico requerido para entender el trabajo realizado. En primer lugar, se habla de biometría y se mencionan las ventajas y desventajas de la biometría *soft*. Luego se describe el proceso de segmentación y normalización del iris realizado por el software OSIRIS [21]. Posteriormente, se describe el estado del arte, resumiendo las publicaciones anteriores que han trabajado en la clasificación de género en base a iris (periocular y normalizado). Finalmente, se describen brevemente los algoritmos de *machine learning* usados, VGGNet-16 [22] y *Support Vector Machine* (SVM) [23].

2.1. Biometría

La biometría es la ciencia de reconocer personas automáticamente basado en características físicas o de comportamiento, tales como rostro, huella dactilar, iris, mano, voz, postura y firma [9]. La biometría permite confirmar o establecer la identidad de una persona en base a “quién es esa persona”, en vez de en base a “qué posee esa persona” (identificación basada en posesión, por ejemplo, una tarjeta de acceso) o “qué recuerda esa persona” (identificación basada en conocimiento, por ejemplo, una contraseña) [24]. Para operar, los sistemas biométricos requieren que previamente se haya inscrito o “enrolado”, proceso en el que se extraen y se almacenan los rasgos biométricos a utilizar, y se los asocia a un individuo [24].

El uso de sistemas biométricos para la identificación de una persona tiene múltiples ventajas en comparación a la identificación basada en posesión o en conocimiento. Es significativamente más difícil copiar, compartir y distribuir rasgos biométricos en comparación a contraseñas o tokens [24]. Los rasgos biométricos no pueden ser perdidos u olvidados, y requieren que la persona a reconocer esté presente [24]. Además, los rasgos biométricos son difíciles de falsificar [24].

Los rasgos biométricos *soft*, un tipo de rasgos biométricos cuya definición como tal fue introducida en el año 2011 [10], son “características humanas físicas, de comportamiento o adheridas, clasificables en categorías pre-definidas y reconocibles por otros humanos” [9]. Una definición más reciente de rasgos biométricos *soft* es “características personales describibles

por humanos que pueden ser usadas para asistir en o efectuar reconocimiento de personas” [25]. Estos rasgos emergen de forma natural entre las personas y les permiten distinguir a sus pares, a diferencia de la biometría clásica [10]. Ejemplos de rasgos biométricos *soft* incluyen, entre otros, color de piel e iris, postura, vestimenta, etnicidad, edad, y género.

El uso de rasgos biométricos *soft* ofrece ciertos beneficios complementarios a los rasgos biométricos clásicos, potenciando la capacidad de reconocimiento de éstos [9]. Por definición, los rasgos biométricos *soft* son comprensibles por los humanos. Esto facilita tareas de interoperabilidad máquina-persona, por ejemplo, en tareas de vigilancia por video [26]. Otro beneficio reportado es la robustez a datos de baja calidad. En situaciones donde los datos no permiten efectuar biometría clásica —*e.g.*, si no fuera posible reconocer a una persona a través de su iris—, podría ser posible extraer rasgos biométricos *soft* —*e.g.*, su color de ojos, su raza o su género [25, 26].

Ciertos rasgos biométricos *soft* pueden ser obtenidos a distancia, por ejemplo, el color y tipo del cabello, la estatura, el género, el tipo de cuerpo/contextura y la vestimenta. Esto permite que sean obtenidos sin requerir el consentimiento o cooperación del sujeto observado [10, 26]. Además, los rasgos biométricos *soft* no son distintivos —*i.e.* no permiten identificar a un sujeto—, lo que es positivo en temas de privacidad a la hora de extraer y almacenar estos rasgos [10, 26].

Esto último es a su vez una desventaja de los rasgos biométricos *soft*. Al no ser distintivos, es posible saber varios rasgos biométricos *soft* de un sujeto y aún así no poder distinguirlo de otros [10]. Otra desventaja es la poca permanencia —*i.e.* qué tan estables son en el tiempo— de ciertos rasgos biométricos *soft*. Rasgos como el color o estilo del cabello y la vestimenta pueden cambiar fácilmente, lo que limita su uso en el tiempo [10]. Finalmente, ciertas categorías de rasgos biométricos *soft* son altamente subjetivas. Categorías binarias como el género son más fáciles de discernir entre sujetos, pero otras como el color del cabello pueden requerir distinciones más sutiles en las que no todos están de acuerdo [10]. Por ejemplo, lo que para una persona es “cabello de color café”, para otra puede ser “cabello de color castaño oscuro”, y una misma persona puede tener distintas tonalidades de café en su cabello.

2.2. Segmentación y normalización del iris

Como se mostró previamente en la Figura 1.1, el procesamiento del iris consta de múltiples etapas: captura, segmentación, normalización, enmascaramiento, codificación e identificación. Para segmentar y normalizar el iris existen distintas metodologías. Recientemente, se han publicado múltiples trabajos donde se realiza segmentación usando *Deep Learning* [27–30]. Sin embargo, en esta tesis se usó el software OSIRIS (*Open Source for IRIS*, [Programa de Código] Fuente Abierta para Iris) [21] para segmentar y normalizar el iris. Este software se usó por varias razones: al inicio de esta tesis, las demás opciones no estaban aún disponibles; al ser un software que lleva años siendo utilizado en la industria y en la academia [31], se puede dar por probados su desempeño y robustez, y puede ser usado como punto base para la segmentación de iris [31]; usar OSIRIS permite estudiar el impacto de que la segmentación sea imperfecta y hayan elementos que no sean parte del iris visibles en el iris normalizado.

La segmentación y normalización del iris usando OSIRIS constan de las siguientes etapas [21, 31, 32]: localización de pupila, detección de contornos de pupila e iris, y normalización. Antes de poder normalizar, se debe adquirir la imagen a procesar. Para esto, se ocuparon sensores infrarrojos especializados que permiten capturar la imagen de un ojo (como el LG2200) o ambos (como el LG4000) [33], y permiten ver la textura de este aún cuando el iris es oscuro. La imagen adquirida corresponderá a una imagen periocular. A continuación se describe la metodología usada por OSIRIS para segmentar y normalizar.

2.2.1. Segmentación: localización de la pupila

El primer paso en la segmentación es localizar la pupila. Como parte de la configuración del programa, se deben configurar valores mínimos y máximos para el radio de la pupila y del iris, acotando los valores de búsqueda posteriores. La metodología usada en OSIRIS se basa en dos criterios:

1. La suma de los valores de los pixeles al interior de la pupila (con forma circular) debe ser cercana a 0.
2. El gradiente \vec{G}_P en cada punto P del contorno de la pupila está alineado con el radio \vec{OP} , donde O es el centro de la pupila. Por lo tanto, la suma de los ángulos θ_P entre \vec{G}_P y \vec{OP} debería ser cercana a 0.

Para acelerar la búsqueda, la resolución de la imagen es reducida y los criterios son aplicados en forma de filtros. Asumiendo que el radio de la pupila es r , el primer criterio equivale a usar un *kernel* circular de tamaño $r \times r$, como se muestra en la Figura 2.1. Tras haber aplicado este filtro, los valores son normalizados entre 0 y 1 e invertidos (0 pasa a ser 1, y viceversa). De esta forma, un buen candidato a pupila tendrá un valor cercano a 1, y un mal candidato un valor cercano a 0.

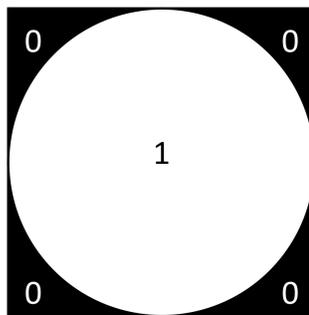


Figura 2.1: *Kernel* circular usado en el primer criterio de búsqueda de pupila. [21]

El segundo criterio puede ser expresado matemáticamente como:

$$\begin{aligned}
\frac{1}{N} \sum_{P=1}^N \cos \theta_P &= 1 \quad , \text{ dado que } \theta_P = 0 \\
\frac{1}{N} \sum_{P=1}^N \frac{\vec{G}_P \bullet \vec{OP}}{\|\vec{G}_P\| \cdot \|\vec{OP}\|} &= 1 \quad , \text{ usando el producto escalar} \\
\frac{1}{N} \sum_{P=1}^N (X_{G_P} X_{OP} + Y_{G_P} Y_{OP}) &= 1 \quad , \text{ asumiendo vectores unitarios} \quad (2.1)
\end{aligned}$$

donde N es el número de puntos en la circunferencia delimitando la pupila. Esta última ecuación puede ser interpretada como la suma de dos operaciones de filtro, una horizontal y otra vertical. Ya que \vec{OP} es un vector unitario, X_{OP} e Y_{OP} pueden ser expresados como:

$$X_{OP} = \frac{x}{\sqrt{x^2 + y^2}} \quad Y_{OP} = \frac{y}{\sqrt{x^2 + y^2}} \quad (2.2)$$

donde x e y son las coordenadas del punto P relativas al centro O .

Para aplicar este criterio se sigue el siguiente procedimiento:

1. Computar los mapas de gradiente horizontal X_{G_P} y vertical Y_{G_P} de la imagen usando el operador Sobel.
2. Generar dos *kernels* con forma de anillo de tamaño $r \times r$, X_{OP} e Y_{OP} .
3. Aplicar el filtro X_{OP} a X_{G_P} y el filtro Y_{OP} a Y_{G_P} .
4. Encontrar la ubicación de los valores máximos en la suma de ambas aplicaciones de filtro.

Esto resulta en valores entre 0 (los gradientes en el contorno no son radiales, por lo tanto es un mal candidato) y 1 (buen candidato).

El resultado final se obtiene al sumar los resultados de ambos candidatos. Ya que el radio r no se conoce a priori, el cálculo se repite para múltiples valores de r , y el máximo corresponderá al centro (O, r) de la pupila.

2.2.2. Segmentación: detección de contornos de pupila e iris mediante algoritmo de Viterbi

Habiendo determinado la ubicación, no necesariamente exacta, del centro de la pupila, se utiliza el algoritmo de Viterbi [34] para encontrar los contornos de la pupila y el iris [32]. Este algoritmo es aplicado a alta resolución para encontrar contornos precisos y enmascarar, y a baja resolución para mejorar la precisión de los contornos de normalización.

Antes de detectar los contornos, la imagen del iris es suavizada usando un filtro anisotrópico [35]. Este filtro permite reducir gradientes locales, lo que reduce el ruido. A su vez,

el filtro preserva las formas y contornos más pronunciados. En la Figura 2.2 se muestra la diferencia al detectar contornos antes y después de filtrar.

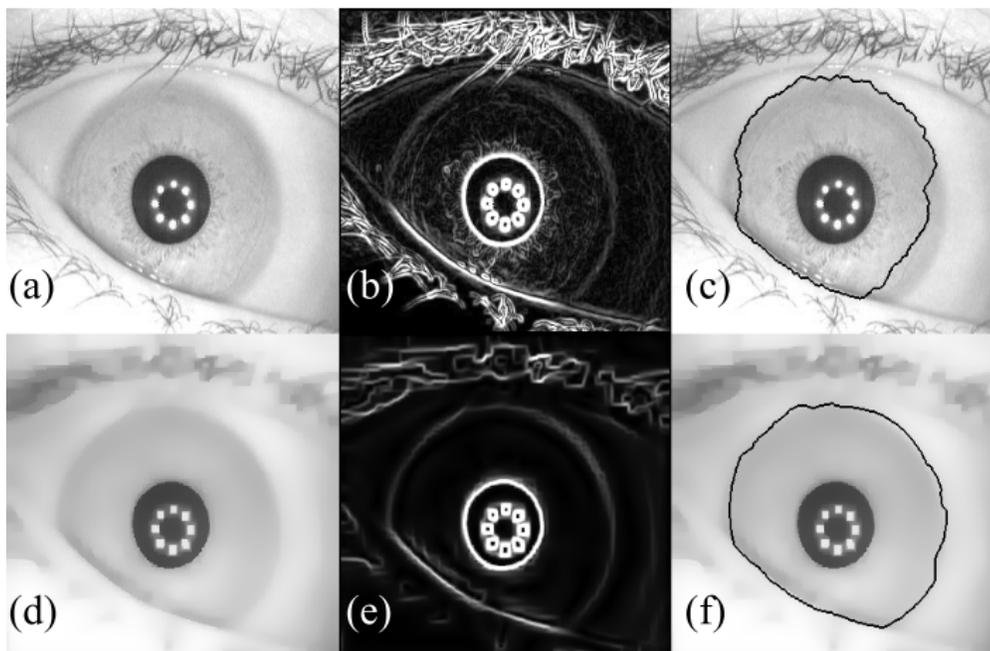


Figura 2.2: Mejora en extracción de contornos al usar el filtro anisotrópico. (a) imagen original; (b) mapa de gradientes extraído de la imagen original; (c) contorno extraído de (b) usando el algoritmo de Viterbi; (d) imagen procesada con filtro anisotrópico; (e) mapa de gradientes extraído de la imagen filtrada; (f) contornos extraídos de (e). [32]

Teniendo la imagen suavizada I , se desea encontrar un camino óptimo global $(R_1, \dots, R_n, \dots, R_N)$ correspondiente a los ángulos $(\theta_1, \dots, \theta_n, \dots, \theta_N)$ cuyos valores varían entre 1° y 360° , mediante un proceso iterativo. El muestreo de los ángulos no es uniforme, pudiendo realizarse a alta o baja resolución. En primer lugar, se calcula un mapa de gradiente radial $G_n(r)$ usando la ecuación:

$$G_n(r) = \sum_{k=1}^5 I(r+k, \theta_n) - I(r-k, \theta_n) \quad n = 1, \dots, N \quad (2.3)$$

donde $I(r, \theta)$ es el valor de intensidad del pixel (representado en coordenadas polares relativas al centro estimado de la pupila). Los valores de $G_n(r)$ son normalizados en relación a su máximo y su mínimo para que estén entre 0 y 1. De esta forma, $G_n(r)$ puede considerarse como la probabilidad de que el pixel situado en (r, θ_n) pertenezca al contorno.

Teniendo el mapa de gradiente, se usa el algoritmo de Viterbi para encontrar el contorno cerrado óptimo, el cual maximiza la probabilidad global. Esto consta de dos operaciones, una de avance y otra de retroceso. La operación de avance consiste en computar una matriz

de gradiente acumulado $C_n(r)$ según:

$$C_n(r) = G_n(r) + \max \begin{bmatrix} C_{n-1}(r-1) \\ C_{n-1}(r) \\ C_{n-1}(r+1) \end{bmatrix} \quad (2.4)$$

La operación de retroceso consiste en retroceder en la matriz de gradiente acumulado, seleccionando el mejor radio previo R_{n-1} dado el radio actual R_n , mediante:

$$R_{n-1} = \operatorname{argmax}_{r \in [R_n-1, R_n, R_n+1]} [C_{n-1}(r)] \quad (2.5)$$

El proceso es inicializado escogiendo el máximo en la matriz de gradiente acumulado para $n = N$ para todos los radios posibles. Durante la operación de retroceso se agrega una restricción para asegurarse de que el contorno sea cerrado:

$$|R_n - R_N| \leq n \quad , \quad n \leq N \quad (2.6)$$

2.2.3. Normalización de iris y máscara

La normalización realizada por OSIRIS está basada en el modelo de banda elástica de Daugman [2]. Mediante este proceso, el área del iris delimitada por los contornos de la pupila y el iris —detectados en el paso previo— es convertido a una forma de representación invariante a la rotación y relativamente invariante a la dilatación pupilar ².

Para normalizar, OSIRIS usa los contornos gruesos detectados a baja resolución usando el algoritmo, a diferencia de metodologías previas en que se usan círculos para normalizar [32]. Los contornos finos detectados a alta resolución son usados para generar la máscara de ocusión.

Sean W y H el ancho y alto, respectivamente, de la imagen normalizada deseada. Siguiendo la metodología de banda elástica de Daugman, se computa un muestreo regular de ángulos θ_k donde k va desde 0 a W , de modo que $\theta_0 = 0$ y $\theta_W = 2\pi$:

$$\theta_k = \frac{k}{W} 2\pi \quad (2.7)$$

Sean (x_p, y_p, φ_p) y (x_i, y_i, φ_i) las coordenadas de un punto en los contornos gruesos de la pupila y el iris respectivamente, donde (x, y) son las coordenadas del radio relativas al centro estimado de la pupila, y φ el ángulo muestreado.

Se estima el punto (X_k^p, Y_k^p) (mostrado en la Figura 2.3), en el contorno de la pupila y correspondiente al ángulo θ_k , interpolando los dos puntos más cercanos j y $j+1$ del contorno

² Si bien originalmente se decía que el modelo de Daugman era invariante a la dilatación pupilar, investigaciones más recientes [36, 37] muestran que la deformación sufrida por el iris producto de la dilatación pupilar es no lineal, y puede afectar al reconocimiento de iris.

grueso de la siguiente forma:

$$X_k^p = (1 - \alpha) \cdot x_p^j + \alpha \cdot x_p^{j+1} \quad (2.8)$$

$$Y_k^p = (1 - \alpha) \cdot y_p^j + \alpha \cdot y_p^{j+1} \quad (2.9)$$

con:

$$\alpha = \frac{\theta_k - \varphi_p^j}{\varphi_p^{j+1} - \varphi_p^j} \quad (2.10)$$

De la misma forma, se estima el punto (X_k^i, Y_k^i) del contorno del iris.

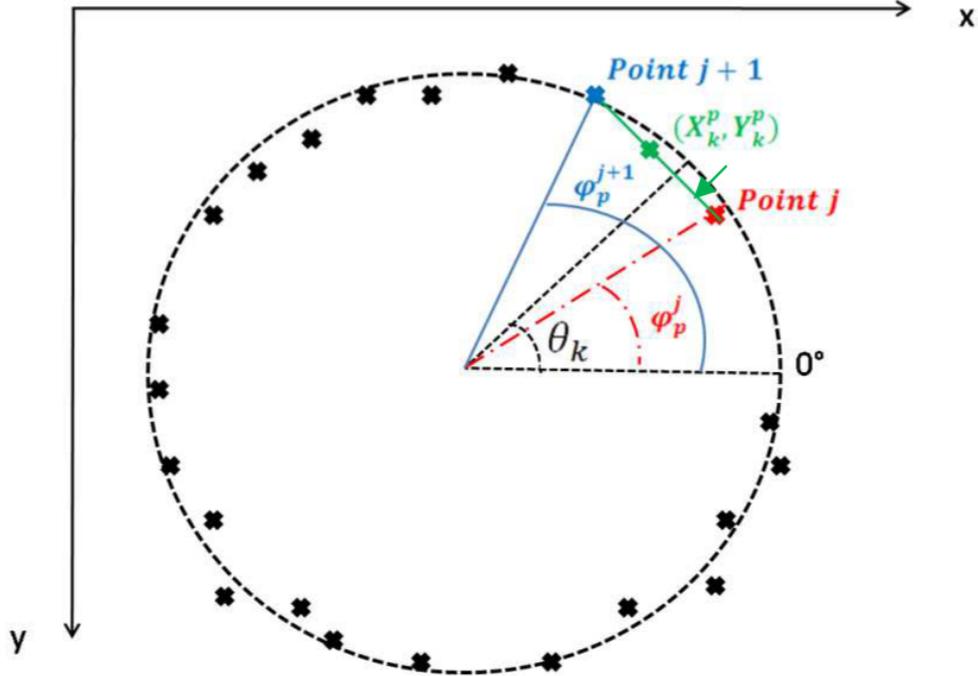


Figura 2.3: Estimación del punto (X_k^p, Y_k^p) . [31]

Luego, sea S el segmento formado por (X_k^p, Y_k^p) y (X_k^i, Y_k^i) . Este segmento es escalado para que quepa en dentro del alto H de la imagen normalizada. En la imagen normalizada, el pixel de la fila h y la columna k tomará el mismo valor que el pixel localizado en $(x_{k,h}, y_{k,h})$ en la imagen original:

$$x_{k,h} = \left(1 - \frac{h}{H}\right) X_k^p + \frac{h}{H} X_k^i \quad (2.11)$$

$$y_{k,h} = \left(1 - \frac{h}{H}\right) Y_k^p + \frac{h}{H} Y_k^i \quad , \quad h \in [0, H] \quad (2.12)$$

Este proceso queda ilustrado en la Figura 2.4, donde los extremos del segmento S (en morado) corresponderán a los puntos (X_k^p, Y_k^p) y (X_k^i, Y_k^i) :

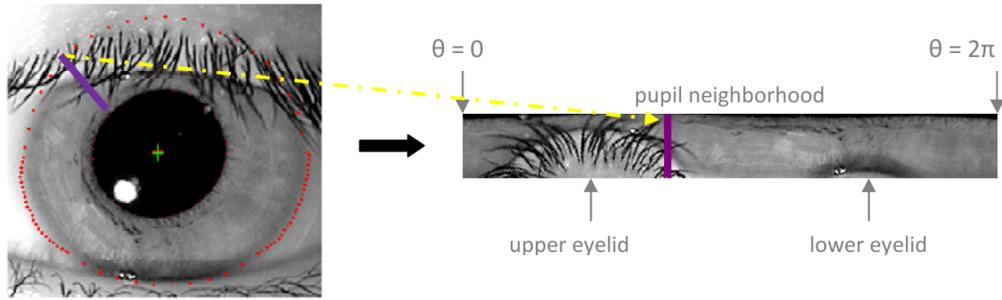


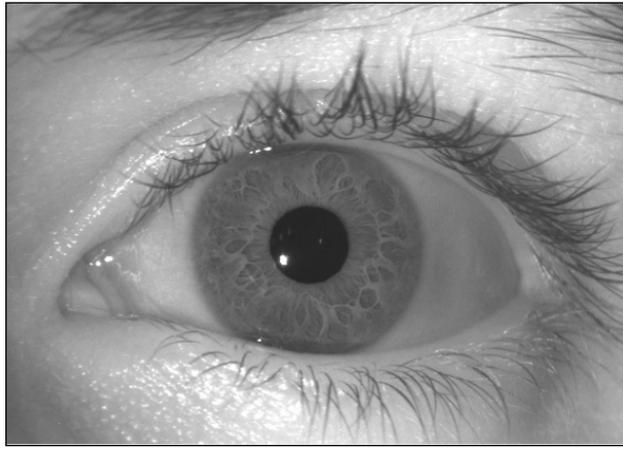
Figura 2.4: La banda S (en morado) es transformada desde la imagen original (izquierda) hacia la imagen normalizada (a la derecha) siguiendo el modelo de banda elástica. [31]

Esta misma transformación es aplicada a la máscara de oclusión que se generó a partir de los contornos finos.

2.3. Clasificación de género en base a iris

La clasificación de género a partir de imágenes de iris ha sido abordada en múltiples estudios siguiendo dos metodologías principales [8]: clasificación usando imágenes perioculares de iris, y clasificación usando imágenes normalizadas de iris. La clasificación usando imágenes de iris perioculares usualmente entrega resultados cercanos y por sobre el 80% de *accuracy*, y ha sido repetida de forma exitosa en el tiempo [8, 19, 38–42]. En cambio, al clasificar usando iris normalizados se han reportado resultados mixtos [5–8, 16–20]. El uso de imágenes perioculares para clasificar género se beneficia de información de género adicional que no está presente en el iris. Por ejemplo, en la Figura 2.5.a se muestra una imagen periocular, la cual contiene información adicional al iris, como la textura de la piel, carúncula, puntos lagrimales, indicios sobre la estructura ósea de la zona periocular, esclerótica, pupila, cejas y pestañas; la imagen normalizada (Figura 2.5.b) contiene sólo la textura del iris, y pequeñas zonas del párpado y las pestañas.

El primer trabajo que se ha encontrado en el que se clasificó género en base a iris fue el de Thomas *et al.* [5], en el año 2007. En este trabajo se usaron más de 50.000 imágenes de iris normalizados de tamaño 240×20 píxeles originalmente adquiridas usando un sensor LG-2200. A estas imágenes se les aplicó la transformada de Fourier y un filtro Gabor 1D, lo que resultó en una imagen compleja (con parte real e imaginaria). De esta imagen compleja se tomó sólo su parte real para computar múltiples características, tanto geométricas como de textura. Las características geométricas corresponden a la diferencia en los ejes X e Y entre el centro de la pupila y del iris, la distancia euclidiana entre el centro de la pupila y el del iris, las áreas de la pupila y el iris, y la diferencia y la razón entre estas áreas. Las características de textura corresponden a valor medio de cada fila, desviación estándar de cada fila, varianza local por fila usando ventanas de 1×5 , y variación local por columna. Para clasificar usaron Bagging de árboles de decisiones de tipo C4.5 [43], reportando un *accuracy* al clasificar género mayor al 80%.



(a) Imagen periocular



(b) Imagen del iris normalizada

Figura 2.5: Imagen de iris periocular, antes (a) y después (b) de normalizar el iris. La imagen periocular contiene información adicional al iris, como la textura de la piel, carúncula, puntos lagrimales, indicios sobre la estructura ósea de la zona periocular, esclerótica, pupila, cejas y pestañas.

Un segundo trabajo en que se realizó clasificación de género en iris es el de Lagree y Bowyer [44]. En él se usaron 600 imágenes adquiridas con un sensor LG-4000, y normalizadas a 240×40 píxeles usando el software IrisBee. De las imágenes se extrajeron características usando 6 detectores de puntos y líneas, y 3 mediciones de textura de Laws [45]. Usando el algoritmo de clasificación SMO (Optimización Mínima Secuencial), obtuvieron un *accuracy* del 62 % al clasificar género.

Bansal *et al.* [16] experimentaron con 300 imágenes de iris adquiridas con un sensor dual Cross Match SCAN-2. Las imágenes fueron obtenidas de 150 sujetos, 100 hombres y 50 mujeres. De las imágenes se eliminaron elementos que no son parte del iris (por ejemplo, párpados y pestañas) usando un algoritmo de enmascaramiento no especificado. En este trabajo usaron características estadísticas y de *wavelets*, y se clasificó usando una *Support Vector Machine* (SVM), reportando un *accuracy* del 83.06 %.

Tapia *et al.* [18] usaron diferentes implementaciones de patrones binarios locales (*local binary patterns*, LBPs en inglés) para extraer características del iris normalizado. En este trabajo se usaron 1.500 imágenes, las que debían ser separadas en particiones *subject-disjoint*, es decir, un mismo sujeto no debía aparecer tanto en la partición de entrenamiento como la de prueba. Sin embargo, posteriormente se informó que las particiones usadas no eran *subject-disjoint*, puesto que la base de datos tenía más de una imagen por sujeto, con un promedio de seis imágenes por sujeto [6]. Esto podría provocar que el clasificador estuviera aprendiendo a reconocer los sujetos en vez de identificar el género de éstos. En ese trabajo se reporta un *accuracy* al clasificar género de hasta un 91 %, pero como las particiones no eran *subject-disjoint*, los resultados podrían estar sesgados.

Da Costa-Abreu *et al.* [46] compararon distintos tipos de características y clasificadores. Las características usadas corresponden a características geométricas (como las coordenadas del centro de la pupila y el iris) y características de textura (obtenidas usando filtros log-Gabor 1D), las que además fueron combinadas mediante concatenación en una tercera prueba. Los clasificadores usados fueron: perceptrón multi-capas (*Multi-Layer Perceptron*, MLP), SVM, poda incremental de errores reducidos (*Incremental Reduced Error Pruning*, IREP) optimizada, árbol de decisiones, y *K-Nearest Neighbours* (KNN). Al combinar las características, obtuvieron un *accuracy* de hasta 81.43 %.

Tapia *et al.* [6] generaron una base de datos que eliminó el problema de [18]. En este nuevo trabajo, usaron el mismo código binario que se extrae del iris habitualmente para realizar identificación a partir de imágenes de iris. Este código fue usado para clasificar género en base a iris. En este trabajo, Tapia *et al.* descubrieron que las características del iris se encuentran repartidas por el iris, en vez de estar localizadas en bandas concéntricas particulares. También determinaron que mediante selección de características se obtenían mejores resultados que al usar todas las características. Usando esta metodología, obtuvieron un *accuracy* del 89 %.

Bobeldyk y Ross [19] usaron cuatro regiones diferentes de las imágenes de iris periocular y normalizado para clasificar género: imagen periocular completa, región sólo de iris de la imagen periocular, región externa al iris de la imagen periocular, e iris normalizado. Para extraer características de las regiones ya mencionadas usaron el operador de textura BSIF (Binarized Statistical Image Feature). Los mejores resultados fueron obtenidos con la imagen periocular completa (84 %), mientras que los peores resultados fueron obtenidos con el iris normalizado (65 %).

Singh *et al.* [39] usaron un *autoencoder* supervisado para extraer características. Éstas fueron usadas para clasificar género y etnicidad. En su trabajo introdujeron la codificación de la etiqueta de clase en el entrenamiento del *autoencoder*, lo que le da el carácter de supervisado. El *accuracy* obtenido al clasificar imágenes periorculares género usando esta técnica fue de entre un 80 % y un 83 %.

Tapia y Viedma [47] usaron imágenes tanto del espectro visible como del infrarrojo cercano (*Near-Infrared*, NIR) para clasificar género en base a imágenes periorculares. Las características usadas fueron la intensidad de las imágenes, LBPs uniformes (ULBPs) e histogramas de gradientes orientados (*Histogram of Oriented Gradients*, HOG). Además se usó información mutua (*Mutual Information*, MI) para seleccionar las características más relevantes. En este trabajo, los autores mostraron que al combinar las características de ambos espectros, y al usar características extraídas a múltiples escalas, se obtienen mejores resultados que al usar sólo las de uno de los dos espectros. También mostraron que las características más relevantes se encuentran principalmente fuera del iris.

Rattani *et al.* [48] usaron imágenes de la zona periocular capturadas usando dispositivos móviles para clasificar género. Para esto, usaron diferentes descriptores de textura y clasificaron estas características usando SVM y MLP. Como las imágenes provienen de dispositivos móviles, éstas corresponden al espectro visible, a diferencia de otros sensores especializados que abarcan el espectro NIR. Los mejores resultados fueron obtenidos usando HOG y MLP (91 %). En base a lo mencionado en las conclusiones, es posible que las parti-

ciones no hayan sido *subject-disjoint*, lo que podría ocasionar resultados más favorables de lo normal.

Tapia y Aravena [17] clasificaron género en base a iris normalizados usando *deep learning*. En este trabajo se usan dos métodos para clasificar, uno semi-supervisado y otro supervisado. El método semi-supervisado usa una *Deep Belief Network* (DBN) que combina una etapa de aprendizaje no supervisado con una de aprendizaje supervisado, con las que se busca compensar una potencial falta de imágenes etiquetadas. El método supervisado usa una *Convolutional Neural Network* (CNN). Usando el método semi-supervisado obtuvieron un *accuracy* de hasta un 77.79%, mientras que con el método supervisado alcanzaron un *accuracy* de hasta un 83.00%, el que aumentó a un 84.66% al usar *data augmentation*.

Kuehlkamp *et al.* [20] abordan problemas que trabajos anteriores sobre clasificación de género en base a iris podrían haber pasado por alto. Uno de estos problemas es la presencia de maquillaje en las imágenes. En este trabajo, los autores mostraron que si se usan sólo las máscaras de oclusión o si sólo se aplica un umbral para clasificar —factores en los que influye la presencia de maquillaje—, se puede alcanzar una tasa de clasificación del 60%. Un segundo problema abordado por este trabajo es el de las particiones aleatorias. En este trabajo se mostró que si se realiza una única partición de entrenamiento y prueba, los resultados pueden variar entre el 40% y el 100%. Por esta razón, los autores recomiendan promediar múltiples pruebas para obtener resultados que reflejen mejor el desempeño real del modelo.

Rattani *et al.* [49] ahondaron en la clasificación de imágenes perioculares obtenidas con dispositivos móviles. En este trabajo los autores usaron múltiples modelos para clasificar, incluyendo CNNs, y además usaron particiones *subject-disjoint*. Los mejores resultados se obtuvieron usando CNNs pre-entrenadas para extraer características, y clasificando en base a los resultados combinados de dos clasificadores, en este caso, SVM y MLP.

Tapia y Aravena [40] usaron CNNs para clasificar género en base a imágenes perioculares. En este trabajo entrenaron modelos combinando características tanto del ojo izquierdo como el derecho, y los compararon con modelos entrenados usando sólo uno de los ojos. Los resultados obtenidos muestran que la CNN se beneficia al combinar características de ambos ojos, volviéndola más robusta ante degradaciones en resolución y oclusiones, y obteniendo un *accuracy* de hasta un 87.26% al clasificar género.

Viedma y Tapia [50] usaron CNNs para clasificar género, y para determinar la ubicación de las zonas más relevantes en las imágenes de iris perioculares para la clasificación de género. Para esto, usaron tres CNNs con distinto entrenamiento: CNN no entrenada para clasificación de iris, CNN con *fine-tuning* en clasificación de género en base a iris, y CNN entrenada desde cero en clasificación de género en base a iris. Usando *Gradient-weighted Class Activation Mapping* (Grad-CAM), visualizaron los mapas de activación de la CNN con *fine-tuning*. Los resultados muestran que al usar *fine-tuning* se obtienen mejores resultados que con los otros dos métodos. Además, los mapas de activación revelan que las zonas más relevantes son las que rodean el iris, y no el iris en sí.

Bobeldyk y Ross [38] usaron imágenes perioculares para clasificar género y raza. En este trabajo se menciona literatura médica que corrobora la relación entre el género y la textura del iris, las glándulas lagrimales y las córneas. Para la clasificación de género, se

usaron los descriptores LBP y BSIF para extraer características de textura del iris periocular. También se comparó el desempeño en clasificación al usar una región rectangular circunscrita al iris, en contraste con una imagen periocular en la que se excluye dicha región. Los resultados obtenidos muestran que al excluir el iris se obtiene un mejor desempeño (82 %) que al usar la región circunscrita al iris (79 %).

Bobeldyk y Ross [51] estudiaron si era posible clasificar imágenes perioculares al reducir las a menor resolución. Para esto, escalaron las imágenes de las bases de datos BioCOP2009 y Cosmetic Contact usando interpolación bicúbica. A partir de imágenes de 480×640 píxeles, se generaron imágenes de 340×400 , 170×200 , 85×100 , 42×50 , 21×25 , 10×12 , 5×6 , y 2×3 píxeles. Luego usaron dos métodos para clasificar género: BSIF+SVM y CNN. Los resultados obtenidos muestran que al usar sólo 30 píxeles (en la imagen de 5×6 píxeles) se puede lograr un *accuracy* promedio del 72 % usando BSIF+SVM y del 77 % usando una CNN, en la clasificación de género.

Tapia y Pérez [7] clasificaron género en base a iris normalizados usando filtros cuaterniónicos en cuadratura 2D (*2D quadrature quaternionic filter*, QQF) y selección de características. Los filtros QQF consisten en representar la información de fase de una imagen en un código binario bidimensional a partir de filtros Gabor 2D. Para seleccionar características se usó un método llamado *Weighted Conditional Mutual Information Maximization* (W-CMIM), en el que se van seleccionando características que añaden información que no haya sido previamente añadida por las características ya seleccionadas. Usando estas técnicas y el clasificador *GentleBoost* se lograron tasas de clasificación del 93 %.

Kuehlkamp y Bowyer [8] realizaron pruebas para evaluar el poder discriminativo de la textura del iris normalizado para la clasificación de género, y contrastarlo con las imágenes perioculares. Para esto, los autores clasificaron usando tres combinaciones de características y clasificadores: características “hechas a mano” clasificadas con una SVM lineal, características extraídas con una red convolucional *VGGNet-16* (en adelante sólo VGG por simplicidad) y clasificadas con una SVM lineal, y características tanto extraídas como clasificadas con una red convolucional VGG. En estas pruebas usaron la base de datos *Gender For Iris* (GFI), la cual etiquetaron manualmente para indicar presencia de maquillaje. Adicionalmente, generaron una nueva base de datos (GFI-C), con un mayor número de imágenes, y previamente etiquetadas para indicar presencia de maquillaje. Los resultados obtenidos corroboran que hay más información de género en las imágenes perioculares que en las normalizadas. Además, muestran que el maquillaje incide entre un 2 % y un 6 % en los resultados. Por último, los autores estudiaron el uso de máscaras probabilísticas, en las que se van enmascarando píxeles en base a qué tan probable es que estén enmascarados. Cuando no se enmascara, el *accuracy* obtenido es del 64.2 %, mientras que al enmascarar con una probabilidad del 70 % el *accuracy* decae al 56.8 %.

Manyala *et al.* [52] estudiaron la presencia de rasgos de género en elementos de la región periocular. Para esto, recortaron manualmente la zona de las cejas en las imágenes perioculares, y trabajaron tanto sobre las imágenes perioculares sin cejas como sobre las imágenes de las cejas sin el resto de la imagen periocular. Luego extrajeron sus características usando una CNN, y las clasificaron usando una SVM o la misma CNN, la que fue ajustada (*fine-tuned*) para clasificar género. Los autores reportaron mejores resultados usando sólo las cejas que usando las imágenes perioculares sin cejas.

Viedma *et al.* [42] usaron el algoritmo *XGBoost* para seleccionar las características más relevantes en imágenes perioculares. Estas características corresponden a intensidad de imagen (valores de píxeles en bruto), ULBPs, HOG, y HOG inverso. Para clasificar se usó principalmente una SVM, pero también se usaron otros 9 clasificadores para validar los resultados. Adicionalmente, estudiaron el impacto de cubrir la zona del iris de color negro. Los resultados obtenidos muestran que al seleccionar características usando *XGBoost* se obtiene un *accuracy* de hasta 89.22 % al clasificar género. Al cubrir el iris, este *accuracy* decae a un 84.35 %.

Tapia *et al.* [41] estudiaron la clasificación de género en imágenes perioculares obtenidas de distintas bases de datos, simulando escenarios más difíciles. Las bases de datos abarcan múltiples longitudes de onda, tanto en el espectro visible como en el NIR. En las pruebas realizadas usando sensores cruzados pero dentro de un mismo espectro obtuvieron un *accuracy* del 85 % en promedio. Al entrenar y probar en espectros diferentes, el *accuracy* decae al 82 % en promedio. Finalmente, al usar múltiples espectros tanto en entrenamiento como en clasificación, obtuvieron un *accuracy* del 83 %.

Tapia y Arellano [53] usaron redes generativas adversarias (*Generative Adversarial Networks*, GANs) para crear imágenes de iris perioculares artificiales etiquetadas por género. Estas imágenes fueron usadas para entrenar una CNN en la clasificación de género en base a imágenes de iris perioculares. Los autores usaron dos metodologías de GANs ya existentes, y propusieron una nueva metodología en la que se seleccionan píxeles del espacio latente en base a su relevancia para codificar género usando CMIM. Los resultados obtenidos muestran que las metodologías ya existentes no codifican correctamente información de género, mientras que la metodología propuesta permite lograr tasas de clasificación comparables, e incluso mejores, a las obtenidas usando bases de datos tradicionales.

Tapia y Arellano [54] crearon una versión modificada de BSIF (M-BSIF) usando texturas tomadas de imágenes oculares, para clasificar género en base a iris normalizados. Se probaron distintas configuraciones de M-BSIF, y luego se clasificó las características extraídas usando múltiples algoritmos de clasificación. Los resultados obtenidos muestran una mejora en el desempeño al clasificar usando M-BSIF, en contraste con el descriptor BSIF original.

Tapia *et al.* [55] estudiaron la clasificación de género en base a imágenes de iris de espectro visible escaladas usando super-resolución. Para esto, tomaron imágenes de baja resolución (150×150 píxeles) tomadas usando smartphones, y las escalaron al doble y triple de su resolución original usando CNNs de súper-resolución (Super-Resolution Convolutional Neural Networks, SRCNNs). Los resultados obtenidos muestran una mejora de hasta un 22 % en la clasificación de género al aumentar la resolución de las imágenes.

Eskandari y Sharifi [56] estudiaron el desempeño en la clasificación de género en base a iris en contraste con el rostro completo. Para esto, usaron las bases de datos de rostros *CASIA-Iris-Distance* y *MBGC*. Al trabajar en iris, este fue segmentado a partir de la imagen del rostro, y normalizado a 20×240 píxeles. De las imágenes se extrajeron características usando el descriptor ULBP, del que se seleccionó un 60 % de las características usando BSA (*Bat Search Algorithm*). Los resultados muestran que al usar sólo características de la región del iris las tasas de clasificación van desde el 54 % de *accuracy* (con un sólo iris) al 66 % (usando ambos iris), mientras que al combinar estas características con las del rostro y la

zona periocular se alcanzan resultados en torno al 90% de *accuracy*.

Con la pandemia del virus SARS-CoV-2, el estudio de rasgos biométricos, tanto *soft* como tradicionales, en rostros cubiertos con mascarillas se ha vuelto más activo [57–61]. La mayoría de los trabajos mencionados a continuación se enfocan en la extracción de rasgos biométricos del rostro, y en particular, aprovechan la información periocular cuando las oclusiones (en este caso, de mascarillas) dificultan el acceso a los demás rasgos.

Alonso-Fernandez *et al.* [62] estudiaron el poder predictivo de distintas regiones del rostro en la clasificación de género, edad y etnicidad. El estudio fue realizado usando rostros de la base de datos *Labeled Faces in the Wild*, los que fueron obtenidos en entornos no controlados. De estos rostros se extrajeron las regiones perioculares y los labios, y se obtuvo características usando distintos tipos de redes convolucionales. Las características fueron clasificadas usando SVMs. En la clasificación de género se reportaron resultados en torno al 90% al usar un ojo o los dos. Los autores mencionan como punto a mejorar el desbalance de clases presente en su base de datos: ésta contiene un 77.6% imágenes masculinas y 22.4% imágenes femeninas. Esto podría distorsionar los resultados obtenidos.

Raja *et al.* [63] investigaron la clasificación de género en base a imágenes de iris perioculares usando imágenes de múltiples espectros. Estas imágenes fueron capturadas usando un sensor que opera en ocho bandas de frecuencia distintas, abarcando los espectros visible y NIR. De estas imágenes se extraen características (denominadas GIST) multi-escala usando filtros basados en escenas naturales, y se las concatena. Las imágenes son clasificadas usando *Kernel Weighted Collaborative Representation Classification* (KWCRC). Usando esta técnica se logró un *accuracy* en clasificación de género del 81.6%.

Sable y Rajput [64] clasificaron género en base a iris normalizados usando *wavelets* discretos 2D. Las imágenes usadas fueron descompuestas en cuatro sub-bandas usando *wavelets* discretos 2D. La componente de baja frecuencia fue descompuesta y se le aplicó la transformada Radon, resultando en una matriz de características de 49×181 . A partir de esta matriz, se calculó la mediana, la desviación estándar y la varianza, dando lugar a un vector de tres características. Cabe destacar que las últimas dos características son cuadráticamente proporcionales entre sí. Las imágenes fueron separadas en conjuntos de entrenamiento y prueba *subject-disjoint*, con la siguiente composición: 408 imágenes masculinas y 412 femeninas en el conjunto de entrenamiento, y 217 imágenes masculinas y 28 femeninas en el conjunto de prueba. Los autores reportan un 95.9% de *accuracy* al clasificar género. Los resultados llaman la atención, considerando las características usadas y el desbalance en el conjunto de pruebas.

Alonso-Fernandez *et al.* [65] ahondan en su investigación previa, esta vez usando una base de datos más balanceada. Las imágenes de esta base de datos fueron obtenidas usando smartphones, y como en el trabajo anterior, fueron tomadas en entornos no controlados. Los autores ahondaron en el uso de redes convolucionales más pequeñas, centrando su potencial uso en dispositivos móviles. El mayor *accuracy* obtenido usando un solo ojo fue de 76.6%, y usando ambos ojos fue de 78.9%.

Vetrekar *et al.* [66] estudiaron la clasificación de género en base a imágenes perioculares ocluidas por lentes ópticos. La base de datos usada usó un sensor similar al que generó la

base de datos usada en [63]. Esta base de datos contiene imágenes perioculares de sujetos con y sin lentes ópticos. De las ocho bandas de frecuencia disponibles se seleccionaron las cuatro con mayor entropía. A las imágenes se les extrajeron características usando filtros Gabor 2D, y luego se las clasificó usando *Probabilistic Collaborative Representation-based Classifier* (ProCRC). Las imágenes con lentes ópticos fueron usadas exclusivamente en el conjunto de prueba, y se comparó el desempeño al clasificar estas imágenes con el desempeño al clasificar las imágenes sin lentes. Los resultados muestran degradación en los resultados (aproximadamente un 3%) al usar lentes ópticos.

Khan *et al.* [67] clasificaron género en base a imágenes de iris en el espectro visible. Para esto, separaron las imágenes de iris en sus 3 canales (rojo, verde y azul), y extrajeron características usando momentos de Legendre, momentos de Zernike, y el descriptor HOG. Estas tres características fueron concatenadas y clasificadas usando una SVM, obteniendo un *accuracy* del 96%. Las imágenes usadas corresponden a la base de datos *CVBL Color*, la que cuenta con tres imágenes de cada ojo. Los autores mencionan que se usaron dos imágenes para entrenar y una para probar, por lo que las particiones no son *subject-disjoint*. Como ya se mencionó previamente [6, 20], esto podría ocasionar que se reporten resultados más favorables de lo que en realidad son.

2.4. Redes convolucionales (CNNs)

Las redes neuronales convolucionales (*Convolutional Neural Networks*, CNNs) [68] son un tipo de modelo de aprendizaje de máquinas inspirado en la corteza visual del cerebro humano [69]. Éstas pueden ser entrenadas para clasificar distintos tipos de información, en particular imágenes. Las CNNs están conformadas por múltiples “capas convolucionales”, consistentes en filtros de distinto tamaño, los que transforman la información que ingresa a ellos mediante la operación de convolución [70]. Los filtros que componen las CNNs son ajustados durante un proceso de entrenamiento, en el que se hace uso del algoritmo de retro-propagación [71]. De esta forma, la red aprende a reconocer patrones y extraer características, los que después son usados para clasificar [22].

2.4.1. VGGNet

VGGNet es una arquitectura de CNN presentada por primera vez en el desafío ILSVRC-2014 [72] y luego en el paper “Very Deep Convolutional Networks for Large-Scale Image Recognition” [72]. La abreviación VGG significa “Visual Geometry Group” (Grupo de Geometría Visual en inglés), correspondiente al grupo del *Department of Engineering Science* de la Universidad de Oxford. El número que suele acompañar al nombre (por ejemplo, VGGNet-16) indica el número de capas de la red.

Esta arquitectura innovó frente a otras arquitecturas de ese entonces al reducir el tamaño de los filtros convolucionales a 3×3 , y aumentar la profundidad de la red hasta 19 capas [22]. En su paper demuestran que la mayor profundidad de la red influye positivamente en los resultados de clasificación, y que el modelo VGGNet es capaz de generalizar para un amplio rango de tareas y bases de datos [22].

La entrada a la red es una imagen RGB (canales rojo, verde y azul) de 224×224 píxeles, a la que se le ha restado a cada pixel la media de su canal correspondiente; esta media es calculada en el conjunto usado para entrenar la red. La imagen pasa a través de un conjunto de capas convolucionales, donde cada filtro es de 3×3 píxeles. Los filtros son aplicados de modo que se preserve el tamaño original de la entrada, usando pasos (*strides* en inglés) y *padding* de tamaño 1. Después de algunos de los filtros convolucionales se agregan capas de *max pooling* de tamaño 2×2 píxeles y *stride* 2. Al conjunto de capas convolucionales le siguen tres capas *Fully Connected*. Las primeras dos tienen 4.096 canales cada una. La última tiene 1.000 canales, uno por cada clase de la base de datos de ILSVRC. Finalmente, la última capa corresponde a la función *soft-max*. En la Figura 2.6 se detallan las distintas configuraciones presentadas en [22].

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224×224 RGB image)					
conv3-64	conv3-64 LRN	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 conv1-256	conv3-256 conv3-256 conv3-256	conv3-256 conv3-256 conv3-256 conv3-256
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

Figura 2.6: Configuraciones de la red convolucional (en columnas). La profundidad de las configuraciones aumenta de izquierda (A) a derecha (E), a medida que se añaden más capas (las capas añadidas se muestran en negrita). Los parámetros de las capas convolucionales se denotan como “conv[tamaño del campo receptivo]-[número de canales]”. La función de activación ReLU no es mostrada por brevedad. [22]

2.5. Support Vector Machine

La “Máquina de Vectores de Soporte” o “Support Vector Machine” (SVM) es un modelo de aprendizaje supervisado para clasificación y regresión, introducido en 1995 por Cortes y Vapnik [23]. Este modelo opera creando un hiperplano de decisión que separe las muestras a clasificar en dos categorías. Para esto, la SVM es entrenada usando un conjunto de datos de entrenamiento. Estos datos son proyectados a un espacio de características de mayor dimensionalidad, y luego se definen *vectores de soporte* a partir de algunas de las muestras (ver Figura 2.7).

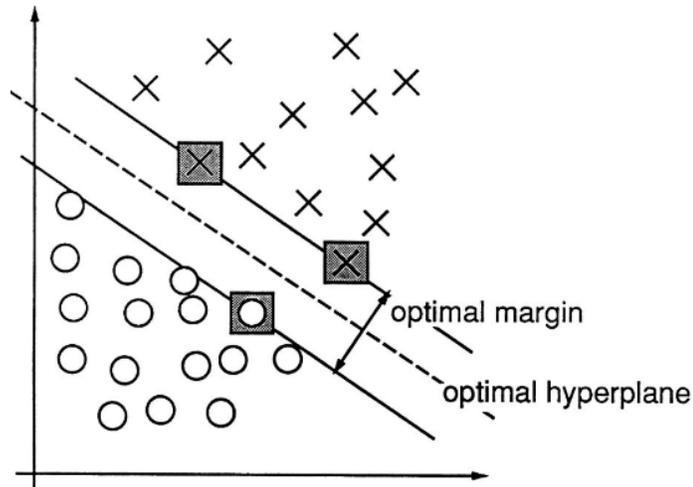


Figura 2.7: Hiperplano (línea punteada) separando muestras de dos clases (círculos y equis), y formado a partir de vectores de soporte (muestras encerradas en cuadrados). [23]

Capítulo 3

Metodología

En este capítulo, se describe la metodología a seguir para comprobar las hipótesis planteadas y cumplir los objetivos. En primer lugar, se define el método propuesto de pareo de máscaras. Luego se describe la metodología y el software usados para corregir manualmente las máscaras. Posteriormente, se caracterizan las bases de datos usadas en este trabajo. Finalmente, se describe la metodología utilizada para clasificar género en base a iris, incluyendo pre-procesamiento, particiones y clasificadores a usar.

3.1. Pareo de máscaras

Como se mencionó anteriormente, es posible que las máscaras de oclusión contengan información de género. De ser así, al entrenar clasificadores en imágenes de iris con máscaras aplicadas, podrían aprender a utilizar la información de las máscaras para clasificar género. Sería beneficioso poder eliminar este sesgo en la clasificación de género en base a iris si lo que se quiere determinar es la cantidad de información presente específicamente en la textura del iris.

Para minimizar este problema, se igualarán las distribuciones de máscaras mediante una técnica que hemos llamado pareo de máscaras. Los pasos de esta técnica son los siguientes:

1. Juntar las máscaras en pares hombre–mujer según cierto criterio de similitud.
2. Realizar la operación OR entre las máscaras de cada par para obtener una máscara pareada.
3. Aplicar estas nuevas máscaras a los pares de iris. De esta forma, la máscara de cada par de iris será idéntica, y cubrirá los elementos no deseados de cada iris.

Al hacer que cada par de iris hombre–mujer tenga la misma máscara, la distribución resultante de máscaras es idéntica para ambos géneros. En la Figura 3.1 se muestra un ejemplo de esta operación para uno de los pares. En este ejemplo se usaron máscaras que no son similares para ilustrar mejor el proceso.

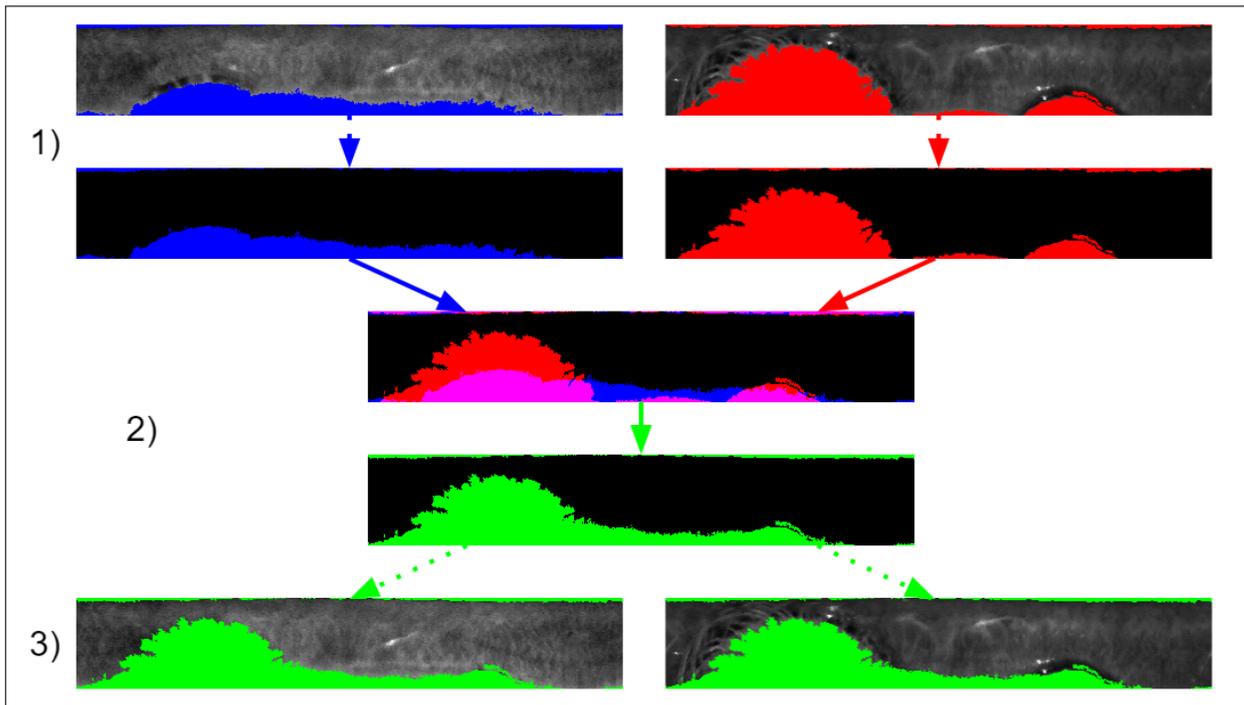


Figura 3.1: Ejemplo de pareo de máscaras. Se usaron máscaras con formas diferentes para ilustrar mejor el proceso. 1) Se forma un par de máscaras de distinto género, indicadas en color rojo y azul. Estas máscaras deben ser similares entre sí para reducir la pérdida de información. 2) Se realiza la operación OR entre las máscaras del par generando una máscara pareada, indicada en color verde. 3) Se aplica la nueva máscara a los iris.

El primer paso es juntar las máscaras en pares hombre–mujer según cierto criterio de similitud. Esto tiene por objetivo reducir la cantidad de información que se pierde al crear pares de máscaras y aplicar esta nueva máscara. La Figura 3.1 (2) ilustra cómo se puede perder información durante el pareo. Los pixeles magenta son aquellos que ya se encontraban enmascarados en ambas máscaras originales, por lo que no incurrirán en pérdida de información. En cambio, los pixeles rojos añadirán máscara al iris de la izquierda, y los pixeles azules añadirán máscara al iris de la derecha. Se deben formar los pares de tal modo que las máscaras originales crezcan lo menos posible y se pierda la menor cantidad de pixeles de iris. Para esto, se definirá una métrica que refleje este factor.

Si tenemos dos máscaras de iris m_f y m_m (una masculina y otra femenina), la nueva máscara generada a partir de estas corresponderá a la operación OR entre las máscaras m_f y m_m . Esta operación puede escribirse como $m_f \cup m_m$, y a la máscara resultante la denominaremos **máscara pareada** entre m_f y m_m . Si $\|m\|$ corresponde a la cantidad de pixeles en la máscara m , entonces la cantidad de pixeles adicionales que tiene la máscara pareada entre m_f y m_m con respecto a m_f estará dada por $\|m_f \cup m_m\| - \|m_f\|$. El **máximo crecimiento de máscara** (*maximum mask growth* en inglés) con respecto a las dos máscaras originales será denotado por M_G y se calcula como:

$$M_G(m_f, m_m) = \max \left(\frac{\|m_f \cup m_m\| - \|m_f\|}{N}, \frac{\|m_f \cup m_m\| - \|m_m\|}{N} \right), \quad (3.1)$$

donde N es el número de píxeles en cada máscara. Bajo esta definición, un par (m_f, m_m) será mejor cuanto menor sea $M_G(m_f, m_m)$. Este valor puede ser expresado como un porcentaje del área de la imagen del iris. Esta será la medida de *fitness*, es decir, la métrica que describirá qué tan bueno es un par.

Habiendo definido la medida de *fitness*, se la debe usar para formar pares hombre-mujer. Asumiendo que la cantidad de máscaras masculinas (M_m) y femeninas (M_f) es la misma, podemos generar una matriz $M_{M_G} = M_G(M_m, M_f)$ con el crecimiento de las máscaras de todas las posibles combinaciones. Determinar los mejores pares en base a esta matriz corresponde a un problema de optimización combinatoria conocido como *problema de asignación* [73].

Si lo que se busca es minimizar la suma total del crecimiento de las máscaras de todos los pares, se pueden usar algoritmos tales como el *algoritmo Húngaro* [74] o el *algoritmo Jonker-Volgenant* [75, 76]. Estos algoritmos permiten encontrar la combinación óptima de forma más eficiente que usando fuerza bruta. Por ejemplo, el algoritmo Húngaro es de orden $O(N^3)$.

3.1.1. Penalización de pares con alto crecimiento

Al minimizar la suma total de los crecimientos, se observa que el algoritmo genera algunos pares con un alto crecimiento (ver Sección 4.1). Ya que el algoritmo optimiza la suma total y no los pares de forma individual, es posible que algunos pares sean formados con un alto crecimiento. En los pares con alto crecimiento, al menos uno de los iris perderá una cantidad considerable de información de iris tras aplicársele la máscara pareada.

Para prevenir que se formen pares con un alto M_G , los valores en la matriz M_{M_G} que excedan un determinado umbral son penalizados multiplicándolos por 10^5 . Esto ocasiona que el algoritmo evite formar pares con alto M_G , a menos que sea estrictamente necesario. En este último caso, el algoritmo tenderá a favorecer a los valores más bajos de entre los que fueron penalizados.

Dado que la elección de umbral de penalización podría impactar los resultados en la clasificación, se explorarán múltiples valores de umbral. Se estudiará el efecto de la elección de umbral en la distribución de máscaras en la base de datos, y en los resultados obtenidos al clasificar.

3.1.2. Eliminación de pares con alto crecimiento

La penalización mencionada en la sección anterior no previene que se formen pares penalizados, sino que sólo reduce su ocurrencia. Como ya se mencionó, esto ocasiona que hayan iris normalizados con poca información de iris a causa del método de formación de pares. Los iris con poca información podrían tener efectos (positivos o negativos) en los resultados.

Para medir este potencial efecto, se eliminarán aquellos pares con más alto de creci-

miento. Esta eliminación se realizará fijando distintos umbrales de crecimiento, y se eliminarán del conjunto de entrenamiento aquellos pares que superen dicho valor. En cada una de estas pruebas se medirá el desempeño en la clasificación, y se evaluará la significancia estadística de los resultados usando ANOVA.

Los posibles resultados son dos: que al eliminar los peores pares los resultados no cambien con respecto a no eliminar; o que los resultados sí cambien, ya sea que estos mejoren o empeoren. En el segundo caso, habrá que considerar este efecto al momento de estudiar el impacto de usar pares de máscaras.

3.2. Corrección de máscaras de oclusión

Las máscaras que genera OSIRIS (Sección 2.2.3) permiten cubrir los elementos no deseados en las imágenes de iris. Sin embargo, éstas a menudo no cubren la totalidad de los elementos que no son parte del iris. Por ejemplo, en la Figura 3.2 se muestran cuatro imágenes de iris normalizados, donde las máscaras fueron incorrectamente generadas por OSIRIS, y en las dos imágenes de abajo hay presencia de maquillaje en los párpados. Como ya fue mencionado anteriormente, la presencia de elementos que no son parte del iris puede producir sesgos en los resultados de clasificación de género en base a iris [20].

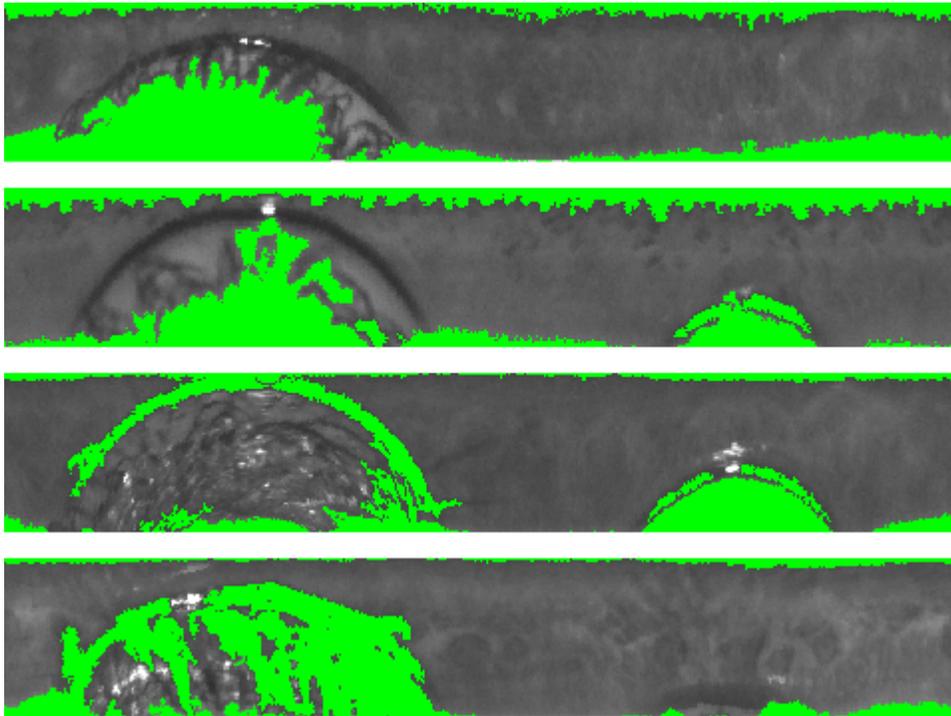


Figura 3.2: Iris normalizados y enmascarados usando OSIRIS. Sus máscaras, indicadas en verde, no cubren correctamente los párpados.

Para estudiar el efecto de las máscaras de oclusión mal generadas, se corrigió manualmente cada máscara de la base de datos GFI para eliminar los elementos no deseados.

Las máscaras originales fueron conservadas para evaluar el impacto de la corrección en el desempeño de los clasificadores.

Para corregir las máscaras, se desarrolló un programa que permite inspeccionar y corregir manualmente los errores de segmentación de las máscaras. Este programa fue desarrollado en Python, usando las librerías *PySimpleGUI*, *numpy*, *pandas* y *Pillow*. En la Figura 3.3 se muestra la interfaz del programa. Entre las características del programa se encuentran: opciones para cambiar la opacidad de la máscara y visualización de la imagen periocular, facilitando su corrección; medir el tiempo que se tarda en corregir cada máscara para estimar el tiempo restante; capacidad de cambiar el radio de dibujo de la máscara con la rueda, y cambio a modo de dibujo/borrar con click derecho.

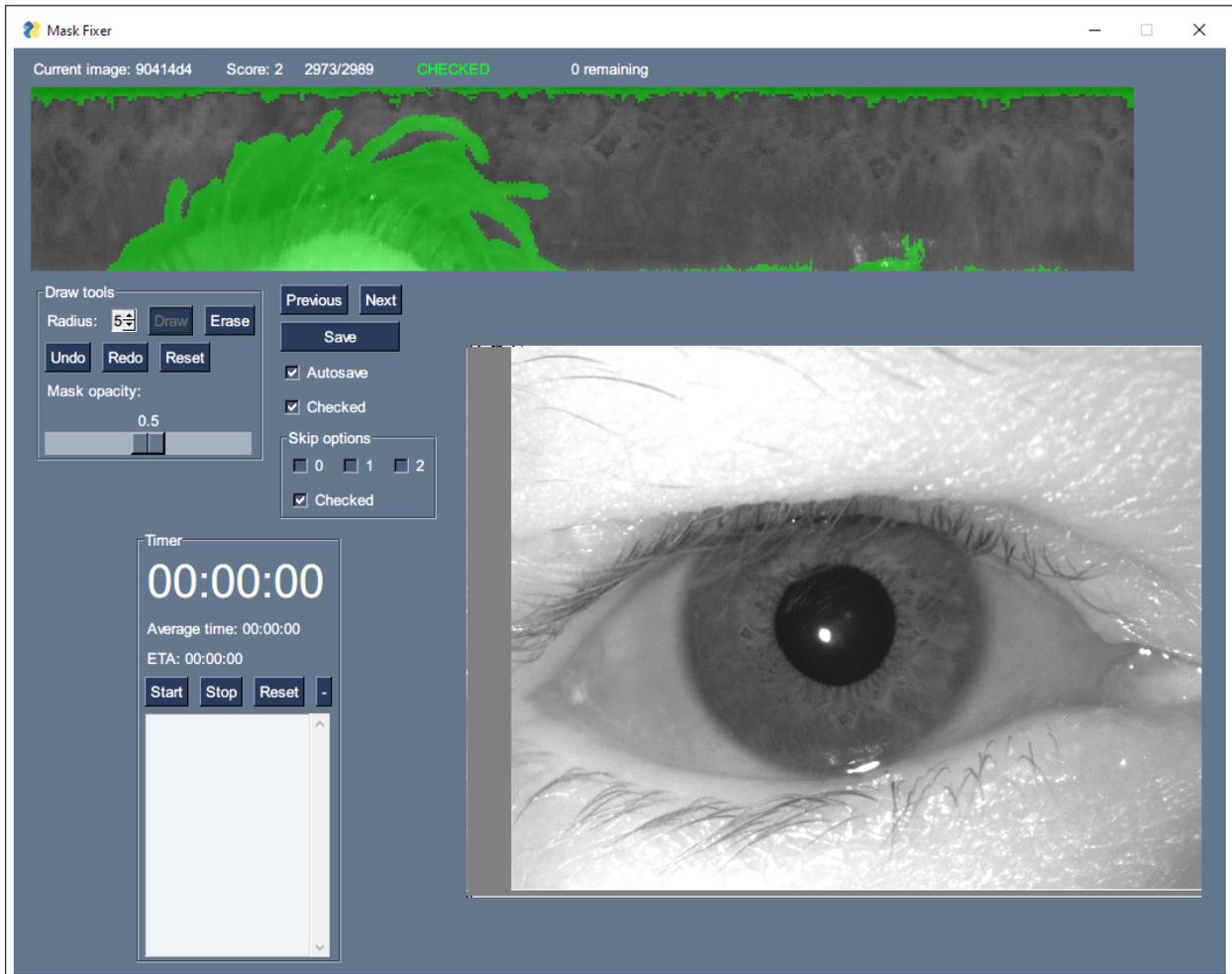


Figura 3.3: Interfaz del programa para corregir máscaras.

Usando este programa, se cubrieron todos los elementos que no son parte del iris tales como pestañas, párpados, pupila, cabellos y esclerótica. Adicionalmente, no se eliminaron pixeles de máscara de las máscaras originales. Ejemplo de ello se ve en la Figura 3.4, donde se muestran dos iris antes y después de la corrección de sus máscaras. Tras haber corregido todas las máscaras, se inspeccionó visualmente cada imagen para corroborar que no quedaran elementos no deseados. Dado que esta corrección fue hecha por un operador humano, es

posible que se introduzcan sesgos o errores en las máscaras.

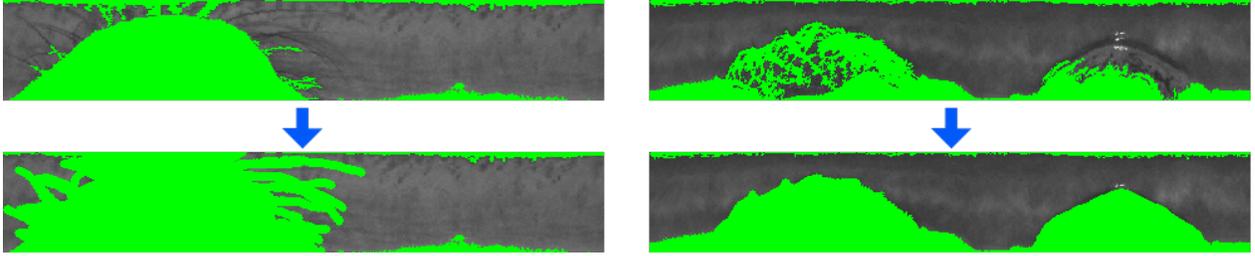


Figura 3.4: Dos iris antes (arriba) y después (abajo) de la corrección de máscaras.

En el resto de este trabajo, se usará el término **máscara corregida** para referirse a las máscaras procesadas usando la metodología descrita en esta sección, y **máscara original** para referirse a las máscaras sin corregir.

3.3. Bases de datos

Las imágenes de iris usadas en este trabajo corresponden a dos bases de datos: *Gender From Iris* (GFI) [6], y *ND-CrossSensor-Iris-2013* [77]. Las imágenes de ambas bases de datos fueron normalizadas usando OSIRIS [21], produciendo imágenes de 480×80 píxeles, cada una con su máscara correspondiente. Las máscaras de la base de datos GFI fueron corregidas usando la metodología descrita en la Sección 3.2, conservando las máscaras originales para estudiar su impacto en el desempeño. A partir de las imágenes normalizadas se generaron imágenes de resoluciones más pequeñas, 240×20 y 240×40 píxeles, usando interpolación bicúbica. Esto incluyó el escalamiento de las máscaras, tanto las originales como las corregidas, mediante la operación *max-pooling*, para garantizar que los elementos no deseados se mantuvieran enmascarados a resoluciones más pequeñas.

La base de datos GFI está compuesta por 3.000 imágenes de iris de 480×640 píxeles adquiridas con el sensor LG-4000 usando iluminación infrarroja cercana. Las imágenes corresponden a fotografías de 1.500 sujetos, 750 mujeres y 750 hombres, a los que se les fotografiaron ambos ojos. La mayoría de las pruebas realizadas en esta investigación usaron imágenes de la base de datos GFI, ya que ha sido utilizada en más estudios, permitiendo comparar resultados directamente.

La base de datos *ND-CrossSensor-Iris-2013* está compuesta de 29.986 imágenes de iris adquiridas con el sensor LG-4000, y 116,564 imágenes de iris adquiridas con el sensor LG-2200. Estas imágenes fueron obtenidas de 676 sujetos, en 27 sesiones realizadas a lo largo de 3 años. De esta base de datos sólo se usaron las imágenes del sensor LG-4000 para reducir las fuentes de variabilidad. Adicionalmente, se limitó el número de imágenes por sujeto a 20 (10 imágenes por ojo). Estas imágenes fueron usadas para verificar que la metodología desarrollada funcionara en otra base de datos.

Para la realización de las pruebas de clasificación, se usaron las imágenes de resolución

240×20 y 240×40 pixeles. Se usaron estas resoluciones ya que son resoluciones habitualmente usadas en procesamiento de iris [6, 17–20, 54]. En cada resolución se usaron las imágenes de ambos ojos. Adicionalmente, se trabajó por separado con las máscaras originales y con las corregidas, y se probó con y sin pareo de máscaras. Estos parámetros, resumidos en la Tabla 3.1, dan un total de 8 combinaciones distintas.

Tabla 3.1: Parámetros de las bases de datos, los que dan un total de 8 combinaciones posibles.

Resolución [px]	Máscaras	Pares
240×20	Originales	Sí
240×40	Corregidas	No

3.4. Comparación de desempeño en clasificación de iris

En esta sección se define la metodología con que se comparará el desempeño de los distintos métodos para clasificar género en base a iris, y con que se evaluará el impacto de las máscaras. Esta metodología permitirá además estudiar el impacto que tienen los pares de máscaras, la eliminación y la penalización de los pares con alto crecimiento, y las correcciones en las máscaras. En resumen, la metodología constará de los siguientes pasos:

1. Normalización: usando OSIRIS, se normalizan y se generan las máscaras de los iris usando dos resoluciones distintas, 240×20 y 240×40 .
2. Partición entrenamiento/prueba: las imágenes se separan en un 80 % para el conjunto de entrenamiento y 20 % para el conjunto de prueba. Se generan 30 particiones diferentes para obtener resultados que reflejen mejor la realidad.
3. Enmascaramiento y escalamiento: se aplican las máscaras a los iris y se escalan los valores de intensidad de los pixeles de los iris normalizados. Las máscaras corresponderán al valor 0, y los pixeles de iris abarcaran el rango 1 a 255. Para enmascarar, se podrán usar los siguientes tipos de máscara: máscara original, máscara pareada, máscara corregida manualmente, y máscara corregida y pareada.
4. Entrenamiento y *fine-tuning*: se entrenan los modelos, ajustando sus parámetros mediante búsqueda por cuadrícula y *5-fold cross-validation*.
5. Extracción de características y clasificación: se extraen características usando una red neuronal convolucional VGGNet-16. Luego se clasifica el género usando la misma red VGG, o se usa una SVM lineal (LSVM).
6. Evaluación y ANOVA: se evalúan los resultados de clasificación obtenidos usando *accuracy*, y se usa la prueba ANOVA para medir qué tan significativos fueron los resultados.

En la Figura 3.5 se resume la metodología usada.

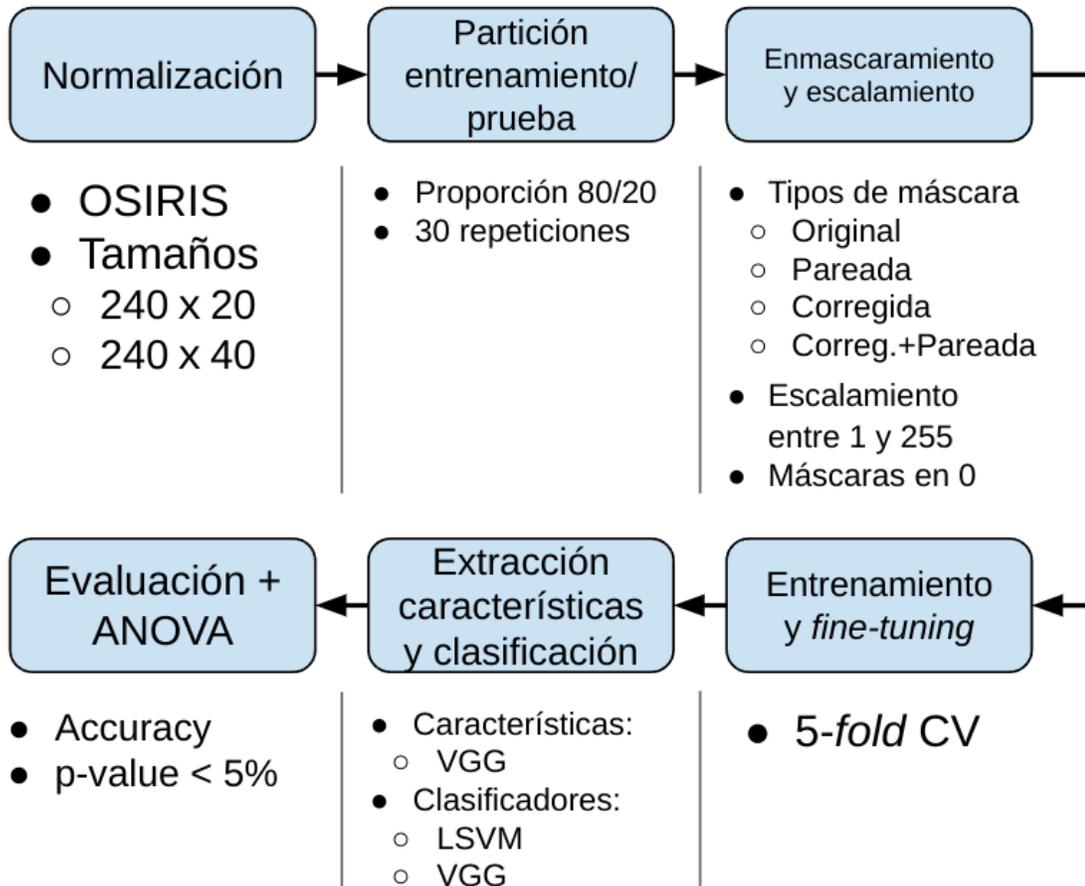


Figura 3.5: Resumen de metodología usada para comparar desempeño en clasificación de género en base a iris.

3.4.1. Carga de imágenes

El primer paso para trabajar en imágenes de iris normalizadas es cargarlas a partir del archivo de imagen. Las imágenes provienen de sensores cercanos al infrarrojo que entregan la información en escala de grises. La profundidad de color de estas imágenes es de 8 bits, por lo que los valores de los píxeles estarán entre 0 y 255. La imagen cargada es representada como una matriz con las mismas dimensiones que la imagen original.

Adicionalmente, al momento de cargar la imagen también se cargan la clase (femenina o masculina) y la máscara de ésta (original o corregida). La clase puede extraerse del nombre de la imagen, del directorio donde se encuentra, o de alguna lista externa. Ésta se representa como un 0 en el caso femenino y como un 1 en el caso masculino. La máscara corresponde a una imagen binaria del mismo tamaño que la imagen original. Ésta indica con 1's los píxeles que deben ser enmascarados por contener elementos no correspondientes al iris.

Para que las imágenes ya cargadas puedan ser usadas por los clasificadores, es necesario convertirlas a partir de la matriz 2-D a un vector 1-D. Este proceso se conoce como vectorización, y puede realizarse “filas primero” (*row-major order*) o “columnas primero” (*column-major order*) [78]. Por consistencia con Python, las imágenes serán vectorizadas

filas primero. Esta misma vectorización es aplicada a las máscaras.

Las imágenes y máscaras vectorizadas son almacenadas en dos matrices: una matriz de datos, que contiene las imágenes, y otra de máscaras. Cada matriz tiene tantas columnas como pixeles tienen las imágenes (número de características), y tantas filas como imágenes sean cargadas (número de muestras). También se genera un vector con la clase de cada imagen, del mismo largo que el alto de las matrices de datos y de máscaras.

3.4.2. Partición de datos en entrenamiento y prueba

Los datos se dividen en un 80 % para el conjunto de entrenamiento y 20 % para el conjunto de prueba. Cada conjunto es balanceado para que contenga la misma cantidad de imágenes de mujeres y hombres. Además, cada sujeto cuenta con una imagen por ojo, con ambos ojos presentes en el mismo conjunto.

Si el experimento fuera realizado en una sola partición entrenamiento-prueba, podría darse que los resultados fueran particularmente favorables o desfavorables sólo por azar [8, 79]. Para asegurarse que los resultados sean representativos del desempeño real que se puede obtener con los datos, al igual que el [8], se generaron 30 particiones entrenamiento/prueba diferentes. Las particiones usadas en [8] no se conocen, por lo que las particiones usadas en esta tesis fueron generadas de forma aleatoria. Estas particiones se conservaron para cada uno de los experimentos.

3.4.3. Escalamiento y enmascaramiento de imágenes

El siguiente paso es escalar y enmascarar las imágenes. Para esto, se usa el siguiente algoritmo:

1. Calcular la mediana de cada fila i , y asignar dicho valor a los pixeles enmascarados de cada fila.
2. Determinar los valores mínimos \min_i y máximos \max_i de cada fila.
3. Escalar y redondear los valores x_{ij} de cada fila según la siguiente fórmula:

$$x_{ij} \leftarrow \left\lfloor \frac{254 \cdot (x_{ij} - \min_i)}{\max_i - \min_i} + 1 \right\rfloor \quad (3.2)$$

4. Aplicar el valor 0 a los pixeles enmascarados.

Mediante este método, los valores no enmascarados quedan en el rango [1–255], y los valores enmascarados quedan en 0. Así, se aprovecha el espectro completo de la escala de grises, y se mantienen los pixeles enmascarados en un mismo valor. Este procedimiento se aplica tanto para las máscaras originales como las máscaras pareadas descritas en la Sección 3.1.

En caso de usarse máscaras pareadas en el experimento, sólo las imágenes del conjunto de entrenamiento son usadas para conformar los pares, ya que no debería poder conocerse *a priori* el género de las imágenes del conjunto de prueba. En el conjunto de prueba se usan máscaras normales o corregidas sin parear. Esto completa el pre-procesamiento de los datos.

3.4.4. Experimentos de clasificación de género

Los protocolos de clasificación usados en este trabajo se basaron en los usados en [8] (ver Sección 2.3). De las pruebas realizadas en [8], se recrearán las dos que usan características extraídas con la red convolucional VGGNet-16 (en adelante VGG); la prueba con características “hechas a mano” requiere algunos datos que no han sido divulgados públicamente, lo que impide recrearla. Las pruebas originales fueron realizadas en dos bases de datos, siendo una de ellas la base de datos GFI (Sección 3.3), lo que posibilita la comparación directa entre los resultados que se obtengan en este trabajo y los presentados en la publicación.

Se intentará replicar aquellos parámetros o métodos que hayan sido descritos en [8]. Parámetros o métodos ambiguos o no descritos serán determinados de modo que permitan obtener los mejores resultados posibles. Para ambas pruebas se usará la implementación de la red VGG que se encuentra en la librería Tensorflow [80]. Al igual que en [8], la red estará pre-entrenada usando la base de datos ImageNet [81].

Los resultados reportados en [8] no están separados por ojo, práctica que es habitual en otras publicaciones del área. Tampoco se hace referencia a esta separación en el texto. Por estas razones, se asumió que se usaron las imágenes de ambos ojos en conjunto, factor que será replicado en estas pruebas. Esta práctica es razonable pues permite contar con más imágenes, potencialmente mejorando los resultados.

Para estudiar la clasificación de género en base a iris, se realizarán siete experimentos. En cada experimento se realizarán dos pruebas, una por cada clasificador. Con estos se evaluarán los efectos de las máscaras pareadas y corregidas en la clasificación de género en base a iris. Los experimentos a realizar son los siguientes:

- **Experimento 1:** Resultados base. Se usan las máscaras originales para obtener los resultados que se usarán como punto de comparación para los demás experimentos.
- **Experimento 2:** Efectos de penalizar pares con alto crecimiento. Se usan máscaras pareadas y se estudia el efecto de variar el umbral de crecimiento en los resultados de clasificación. El experimento es reducido a 5 particiones en vez de 30, y sólo se realizó la prueba VGG+LSVM (ver Sección 3.4.4.1). Este experimento permite decidir el umbral de penalización en la formación de pares de máscaras.
- **Experimento 3:** Efectos de eliminar pares con alto crecimiento. Se usan máscaras pareadas y se estudia el efecto de eliminar los pares con mayor crecimiento en los resultados de clasificación. Este experimento permite determinar si se eliminarán los pares con mayor crecimiento.
- **Experimento 4:** Efectos de usar máscaras pareadas en clasificación. Se usan máscaras pareadas y se estudia el efecto de su uso en la clasificación. En este experimento se

encuentran fijos el umbral de penalización, y se ha decidido previamente si se eliminan o no los pares de mayor crecimiento.

- **Experimento 5:** Efectos de usar máscaras corregidas en clasificación. Se usan máscaras corregidas y se estudia el efecto de su uso en la clasificación.
- **Experimento 6:** Efectos de usar máscaras pareadas y corregidas en clasificación. Se usan máscaras que han sido pareadas y corregidas, y se estudia el efecto de su uso en la clasificación.
- **Experimento 7:** Resultados usando la base de datos ND-CrossSensor-Iris-2013. Se usan las imágenes de iris normalizados de la base de datos ND-CrossSensor-Iris-2013 para ver si la metodología propuesta puede ser aplicada en otra base de datos. Primero se obtienen resultados base con las máscaras originales, y luego se usan máscaras pareadas para estudiar su efecto en los resultados de clasificación.

3.4.4.1. Prueba con características VGG y clasificador SVM

Para esta prueba, se extraen características de los iris normalizados usando la red VGGNet-16, pre-entrenada con la base de datos ImageNet. Estas características son extraídas de la primera capa *Fully-Connected* (FC) de la red, lo que resulta en 4.096 características.

El usar una capa FC de la red pre-entrenada obliga a que las imágenes tengan dimensiones $224 \times 224 \times 3$. Para cumplir este requisito, las imágenes son escaladas usando interpolación lineal, y luego son convertidas de escala de grises a RGB repitiendo la imagen en los tres canales.

Las características ya extraídas son usadas para entrenar y clasificar con una *Support Vector Machine* lineal [23] (o LSVM por su sigla en inglés). Este clasificador posee múltiples parámetros que pueden ser ajustados. Sin embargo, en [8] no se hace mención a éstos.

Para mejorar el desempeño del clasificador, se ajustaron dos parámetros de éste: C , parámetro que define la regularización del modelo; y `max_iter`, parámetro que define el número máximo de iteraciones que se usarán para entrenar el modelo. Estos parámetros son ajustados usando *grid-search* y *k-fold cross-validation*, siguiendo la metodología descrita en la Sección 3.4.5. En adelante, esta prueba será llamada “VGG+LSVM”.

3.4.4.2. Prueba con características y clasificador VGG

Para esta prueba, se usó la red VGGNet-16 *de extremo a extremo* (*end-to-end* en inglés). De esta forma, es la red VGG la que realiza tanto la extracción de características como la clasificación de las muestras. Para esto, los pesos de las capas convolucionales de la red se congelan, y sólo los pesos de las capas FC son ajustados.

Al no necesitar los pesos originales de las capas FC, es posible cambiar el tamaño de las imágenes que ingresan a la red, hasta un mínimo de $32 \times 32 \times 3$. Dependiendo del tamaño de las imágenes que ingresen, será necesario definir una arquitectura para las capas FC de la red con más o menos capas y/o neuronas. En [8] no se hace mención a la arquitectura de

las capas FC, o al tamaño de las imágenes de entrada, ni tampoco a los parámetros usados para entrenar—tales como *learning rate*, optimizador, o función *loss*. Estos elementos fueron ajustados con libertad para obtener mejores resultados.

Cambiar el tamaño de las imágenes de entrada introduce distorsiones en éstas. Además, imágenes más grandes requieren más memoria en la GPU usada para entrenar la red. Por esta razón, se alteró el tamaño de las imágenes de entrada sólo lo necesario para que puedan ser usadas por la red. En el caso de las imágenes de 240×40 píxeles no se hicieron cambios a su tamaño, mientras que a las imágenes de 240×20 píxeles se incrementó su alto hasta el mínimo de 32 usando interpolación lineal.

Para no sobreentrenar las redes, se separó el conjunto de prueba en dos mitades balanceadas de forma aleatoria. Una de estas mitades se usó como conjunto de validación, y la otra se conservó como conjunto de prueba. El conjunto de validación se usó para monitorear el entrenamiento de la red y para realizar detención temprana.

Las capas FC de la red VGG fueron definidas tomando en cuenta la cantidad de características resultantes al final de las capas convolucionales. También se realizaron pruebas con distintas arquitecturas, y se optó por aquella que resultara en un entrenamiento adecuado, esto es, que se observara el incremento progresivo en el *accuracy* de los resultados sobre los conjuntos de entrenamiento y validación a medida que la red es entrenada. Las capas de la red usada finalmente se muestran en la Tabla 3.2.

Tabla 3.2: Arquitectura de capas *fully-connected* para la red VGGNet-16.

Tipo de capa	Neuronas/ <i>Dropout</i>	Función de activación
Dense	1024	ReLU
Dropout	0,5	N/A
Dense	512	ReLU
Dropout	0,5	N/A
Dense	256	ReLU
Dense	2	Softmax

El entrenamiento se realizó durante 100 épocas, con detención temprana tras 20 épocas sin mejoría en el conjunto de validación. Se usó un *batch size* de 32, y un *learning rate* de 10^{-4} . En adelante, esta prueba será llamada “VGG-full”.

3.4.5. Ajuste de parámetros de modelos

Para obtener buenos resultados, se debe ajustar los parámetros de cada modelo antes de poder usarlo. Para esto, se empleó una metodología de *k-fold cross-validation* (validación cruzada) junto con *grid search* (búsqueda de cuadrícula).

Durante el ajuste de parámetros sólo se utiliza la partición de entrenamiento. Esta

partición es sub-particionada para obtener sub-particiones de entrenamiento y de prueba. Como se mencionó en la Sección 3.4.2, si la búsqueda de parámetros se realizara en una sola partición entrenamiento-prueba, los resultados podrían estar sesgados. Para evitar esto se usa *k-fold Cross-Validation* (CV) [79]. Esto consiste en sub-particionar los datos en k partes iguales —en este caso, en 5 partes—, y repetir la búsqueda de parámetros k veces, usando una de las partes como sub-partición de prueba y las demás como sub-partición de entrenamiento.

La búsqueda de cuadrícula consiste en definir múltiples posibles valores para cada parámetro a ajustar, y explorar exhaustivamente cada combinación de valores. Estas combinaciones se usan para entrenar un modelo usando la sub-partición de entrenamiento, y se evalúan usando la sub-partición de prueba. La combinación de parámetros que dio los mejores resultados será la escogida para entrenar el modelo final.

Para reducir el tiempo necesario para encontrar los parámetros, a la vez que se preserva la precisión de éstos, la operación se realiza en dos etapas: gruesa (*coarse*) y fina (*fine*). En la primera etapa, los parámetros a buscar se encuentran más separados entre sí, permitiendo abarcar un rango amplio de parámetros sin tener que realizar tantas pruebas. En la segunda etapa, la búsqueda se centra en torno a los valores obtenidos en la etapa anterior, reduciendo la separación entre los valores e incrementando la precisión.

3.4.6. Análisis estadístico

En este trabajo, se usó la prueba ANOVA (análisis de varianza, *Analysis Of Variance* en inglés) [82] para medir qué tan significativos fueron los resultados. Si el p-value obtenido es inferior al 5 %, los resultados se considerarán significativos. En cada uso, se verificaron los supuestos que permiten usar esta prueba: 1) que las muestras sean aleatorias e independientes entre sí, 2) que los resultados sean obtenidos de una población distribuida normalmente, y 3) que la varianza de las distribuciones sean similares. El primer supuesto se cumple, pues las particiones usadas son efectivamente aleatorias e independientes entre sí.

El segundo supuesto se verificó de dos formas: usando gráficos Q-Q (cuantil-cuantil), y usando la prueba de Shapiro-Wilk [83]. En el caso de los gráficos Q-Q, estos indican normalidad si las muestras graficadas en estos siguen una línea recta. La prueba de Shapiro-Wilk, por otro lado, propone como hipótesis nula que la distribución a probar está distribuida normalmente, y esta hipótesis se rechaza si el valor entregado por el estadístico de esta prueba es pequeño.

Para realizar estas pruebas y la prueba ANOVA se usaron múltiples librerías. La librería *statsmodels* [84] permite generar modelos estadísticos, y se usó para realizar la prueba ANOVA unidireccional y bidireccional. La librería de computación científica y matemática *scipy* [85] fue usada para generar los gráficos Q-Q y realizar la prueba de Shapiro-Wilk. Finalmente, la librería de análisis de datos orientados a la biología *bioinfokit* [86] fue usada para realizar cálculos previos necesarios para la prueba de Shapiro-Wilk.

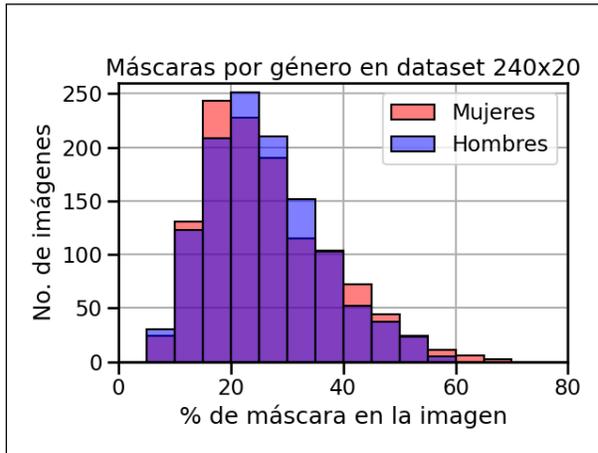
Capítulo 4

Resultados

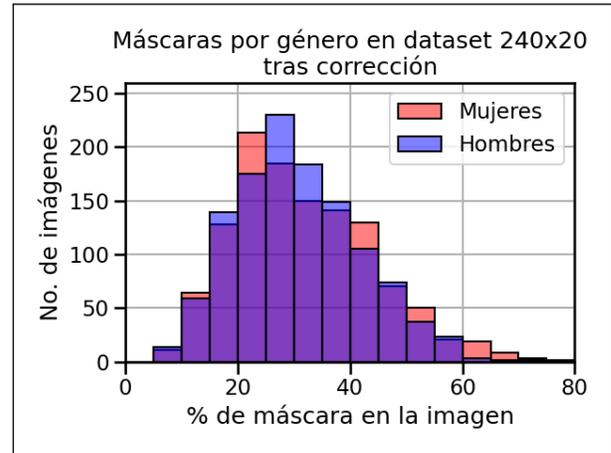
En este capítulo, se muestran los resultados obtenidos tras aplicar la metodología descrita en el capítulo anterior. En primer lugar, se describe el efecto de generar pares de máscaras en la distribución de las máscaras de la base de datos, y cómo éste varía al cambiar el umbral de penalización. En segundo lugar, se estudia el efecto de corregir las máscaras de oclusión en la distribución de las máscaras. Luego, se muestran los resultados de clasificación de género, y cómo los dos factores ya mencionados (pares y correcciones) inciden en la clasificación. Finalmente, se realiza una discusión sobre los resultados obtenidos.

4.1. Efectos del pareo de máscaras en la distribución de máscaras por género

En la Figura 4.1 se muestra la distribución original de máscaras, separadas por género. Al ser distinta la distribución de cada género, un clasificador podría usar esta información para clasificar por género, como se explicó en la Sección 3.1.



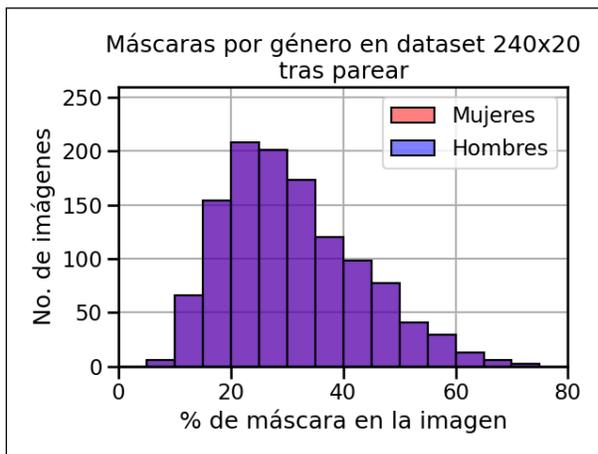
(a) Usando máscaras originales



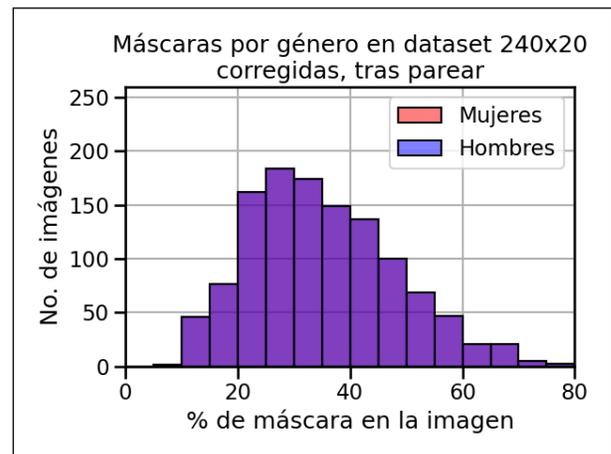
(b) Usando máscaras corregidas

Figura 4.1: Distribución de máscaras por género antes de formar pares. Las columnas se ven moradas cuando ambos géneros se encuentran superpuestos.

Tras formar pares con todas las imágenes del conjunto de entrenamiento, la distribución de máscaras se vuelve idéntica para ambos géneros. Esto se ve reflejado en la Figura 4.2, donde los histogramas se encuentran completamente superpuestos. Nótese que en comparación con la Figura 4.1, hay más imágenes con un mayor porcentaje de máscara. Esto se debe a que al formar pares, la cantidad de máscara sólo puede aumentar, no disminuir.



(a) Usando máscaras originales



(b) Usando máscaras corregidas

Figura 4.2: Distribución de máscaras por género tras formar pares. Las columnas se ven moradas cuando ambos géneros se encuentran superpuestos.

4.1.1. Distribución de pares a distintos umbrales de penalización

Como se describió en la Sección 3.1.1, se debe penalizar los pares con mayor crecimiento. Para definir el umbral adecuado, se debe analizar dos elementos: el crecimiento promedio de los pares formados, y el número de pares con “alto crecimiento”. La definición de pareo

con alto crecimiento es arbitraria, y se define como aquel pareo con crecimiento mayor al 10 %.

Se estudiaron umbrales entre el 0 % y el 20 % de crecimiento, a intervalos de 0.5 %. Para cada umbral se generaron los pares correspondientes, y se tomó nota del crecimiento de cada par; en aquellos pares donde el crecimiento fue penalizado, se consideró el crecimiento, no el penalizado. Teniendo los crecimientos de todos los pares formados, se calculó el promedio y se contabilizaron aquellos con alto crecimiento.

Los resultados de este experimento se muestran en la Figura 4.3. En esta figura se puede observar que a medida que el umbral decrece, el crecimiento promedio aumenta hasta cierto máximo entre los umbrales 5 % y 7.5 %. Pasado este punto, el crecimiento promedio decae, eventualmente volviendo al mismo promedio del umbral más alto.

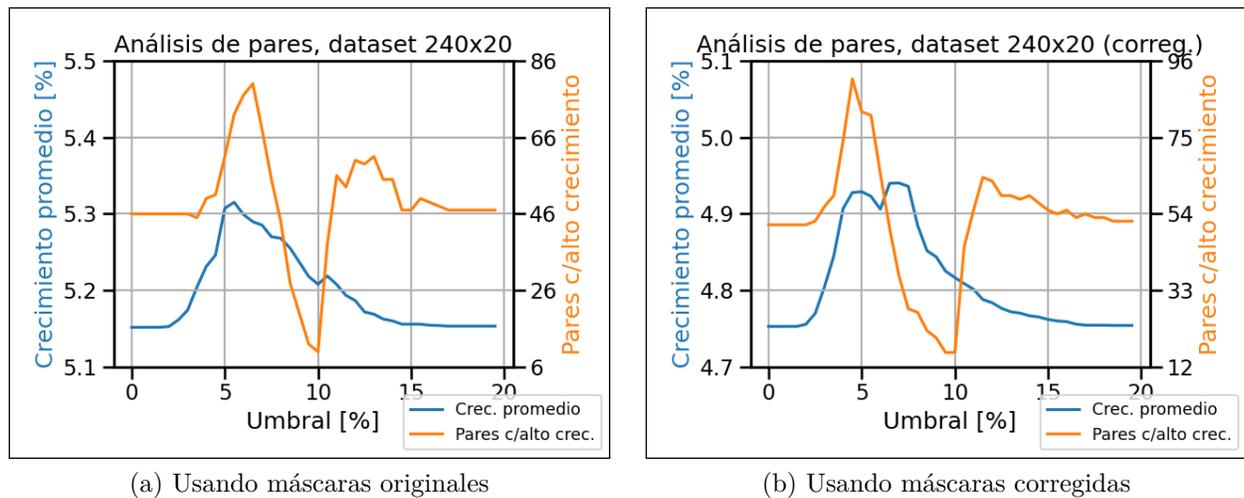


Figura 4.3: Crecimiento promedio y número de pares con alto crecimiento, para distintos umbrales de penalización.

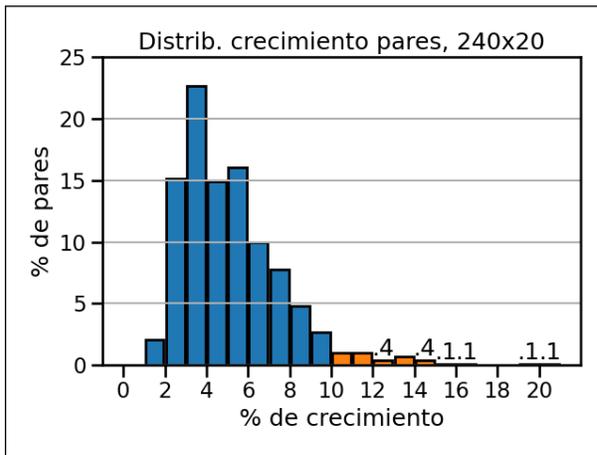
Para explicar este comportamiento, se debe recordar que el algoritmo que forma los pares entrega la solución que minimiza la suma de crecimientos. Esto es equivalente a minimizar el promedio de los crecimientos. A medida que se añaden restricciones –i.e. penalizando pares por sobre cierto valor–, las soluciones se alejan de este mínimo. Llegado cierto punto, la mayoría de los pares potenciales han sido penalizados, lo que hará que el algoritmo forme todos los pares no penalizados posibles, y luego continúe con los no penalizados. Al llegar al umbral de 0 %, todos los posibles pares (a excepción de que hubieran máscaras idénticas en iris de distinto género) han sido penalizados, por lo que el algoritmo se comporta de forma idéntica como si ningún pareo potencial estuviera penalizado.

Por otro lado, se ve un número de pares con alto crecimiento mínimo en el umbral 10 %. Alejándonos de este umbral, el número de pares con alto crecimiento aumenta abruptamente en aproximadamente 50 a 70 pares, para luego descender a valores similares en ambos extremos, aproximadamente 40 pares por sobre el mínimo.

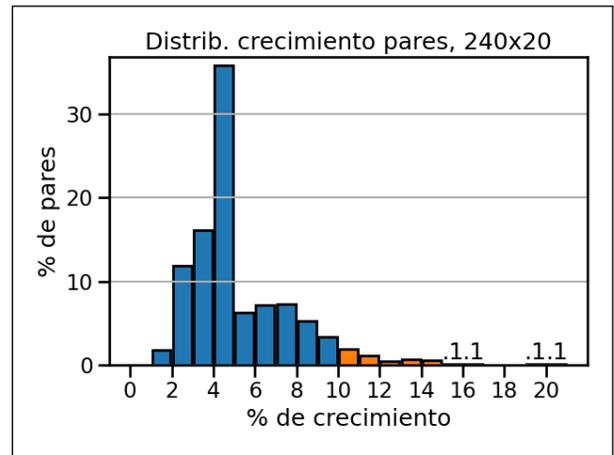
Una posible explicación para este comportamiento está en la definición elegida para alto crecimiento. Si se define alto crecimiento como aquellos pares con crecimiento estricto-

tamente mayor al $x\%$, el umbral que penaliza pares con crecimiento estrictamente mayor al $x\%$ será el umbral menos restrictivo de entre aquellos que penalizan los pares con alto crecimiento.

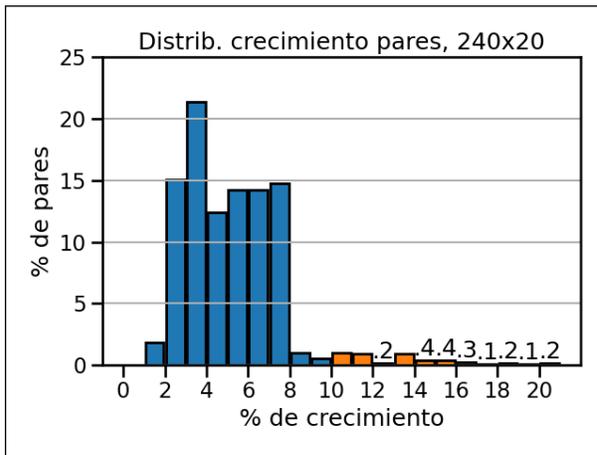
Otra posible explicación para este comportamiento se encuentra en la distribución de pares en base a su crecimiento. En la Figura 4.4 se observa esta distribución para distintos umbrales de penalización. En todos los gráficos se ve que la mayoría de los pares tienen crecimiento menor o igual al 10% . Con el umbral fijado al 10% , el grupo minoritario de pares con crecimiento mayor al 10% queda penalizado en su totalidad, y es posible reducirlo a un mínimo. Los umbrales fijados al 12.5% y al 15% eliminan por completo los pares que están sobre el 12.5% de crecimiento, pero ya que los umbrales no son lo suficientemente estrictos, tienen más pares con crecimiento mayor al 10% que cuando el umbral es fijado al 10% .



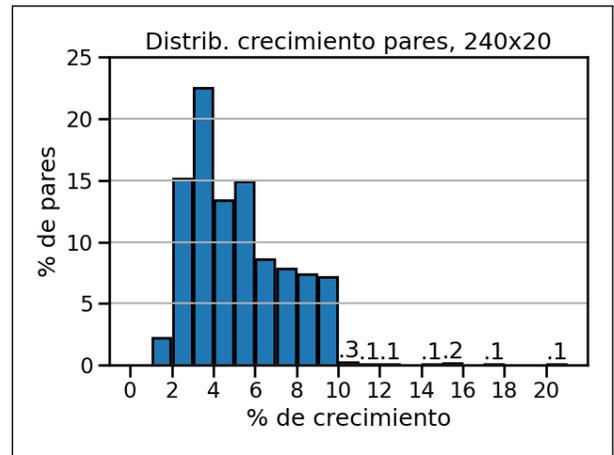
(a) Umbral = 0 %



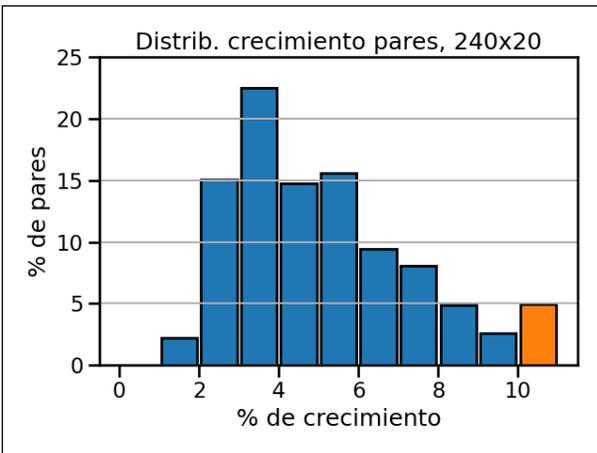
(b) Umbral = 5 %



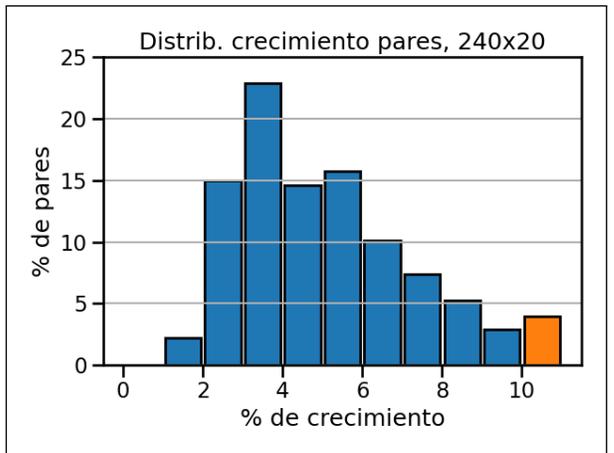
(c) Umbral = 7.5 %



(d) Umbral = 10 %



(e) Umbral = 12.5 %



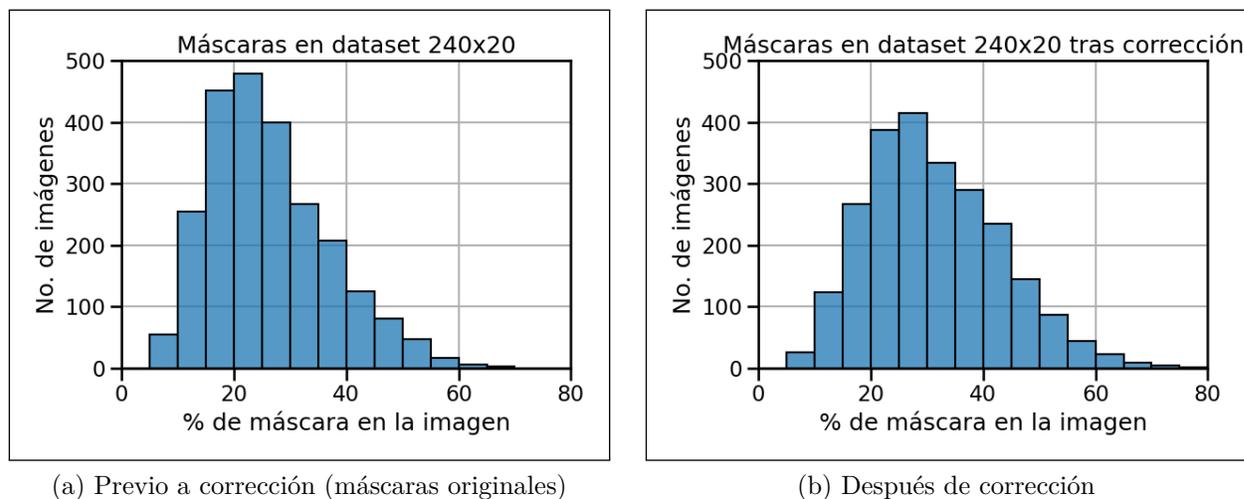
(f) Umbral = 15 %

Figura 4.4: Distribución de pares en base a su crecimiento para distintos umbrales de penalización, usando la resolución 240×20 . Las barras de color anaranjado corresponden a los pares con crecimiento mayor al 10 %. Se agregó una etiqueta a los valores menores a 0.5 % para facilitar su visualización.

4.2. Corrección de máscaras de oclusión

En esta sección se muestra el efecto que tuvo la corrección de las máscaras en la distribución de éstas en la base de datos. Los gráficos mostrados corresponden a las bases de datos GFI, con imágenes normalizadas de 240×20 píxeles, los que son similares a los gráficos usando imágenes normalizadas de 240×40 píxeles.

En la Figura 4.5 se muestra la distribución de máscaras antes y después de la corrección. Ya que la corrección de máscaras se hizo de forma aditiva, la cantidad de máscara en las imágenes sólo se mantuvo o aumentó, lo que se ve reflejado en los histogramas.



(a) Previo a corrección (máscaras originales)

(b) Después de corrección

Figura 4.5: Distribución de máscaras antes y después de corrección.

4.3. Resultados de clasificación

En esta sección se muestran los resultados obtenidos al clasificar género en base a iris, siguiendo los experimentos descritos en la Sección 3.4.4. En primer lugar se obtienen resultados base, los que se usarán como punto de comparación. Luego, se analiza cómo varían los resultados al aplicar pares a distintos umbrales de penalización, y en base a esto se fijará un umbral para usar en el resto de los experimentos. Tras esto, se realizan experimentos eliminando los pares con mayor crecimiento para ver si el desempeño del algoritmo de pareo incide en los resultados. Finalmente, se realizan experimentos completos de clasificación usando pares y máscaras corregidas. La discusión sobre los resultados se realizará en la sección siguiente.

El *accuracy* corresponde al porcentaje de muestras clasificadas correctamente, respecto del total de muestras evaluadas. En cada caso, los valores de *accuracy* mostrados corresponden al promedio y desviación estándar obtenidos de 30 repeticiones aleatorias, según se describió en la Sección 3.4.2.

4.3.1. Experimento 1: Resultados base

Los resultados base corresponden a aquellos obtenidos con la base de datos GFI, usando las máscaras originales (obtenidas directamente usando OSIRIS). Los resultados de los demás experimentos serán comparados con los resultados base para determinar el impacto de las distintas metodologías propuestas.

Los resultados obtenidos por Kuehlkamp y Bowyer en [8] también serán incluidos como punto de comparación. Éstos resultados permitirán comparar los resultados de la metodología utilizada.

En la Tabla 4.1 se muestran los resultados base obtenidos usando VGG+LSVM y VGG-full. En la Tabla 4.2 se muestran los resultados obtenidos en [8] usando imágenes de iris normalizadas tomadas de la base de datos GFI, usando todos los sujetos.

Tabla 4.1: Resultados base usando VGG+LSVM y VGG-full.

Tipo de prueba	Resolución	Accuracy [%]
VGG+LSVM	240 × 20	63.97 ± 1.64
	240 × 40	65.43 ± 1.93
VGG-full	240 × 20	62.13 ± 2.37
	240 × 40	63.34 ± 2.64

Tabla 4.2: Resultados obtenidos por Kuehlkamp y Bowyer en [8].

Tipo de prueba	Accuracy [%]
VGG+LSVM	60.0 ± 1.8
VGG-full	60.1 ± 3.0

Al comparar ambas tablas, se obtiene que los resultados base de este trabajo son en promedio 2.03% a 5.43% más exactos que los de [8], con desviaciones estándar cercanas (< 1% de diferencia).

4.3.2. Experimento 2: Efectos de penalizar pares con alto crecimiento

Para estudiar el impacto de variar el umbral de penalización de pares, se realizaron pruebas de clasificación de género usando máscaras pareadas obtenidas a distintos umbrales de penalización. Los umbrales usados van desde el 3% al 15%. A diferencia de otros experimentos, se usaron 5 particiones en vez de 30 debido al tiempo disponible para realizar el cómputo. Además, se realizó sólo la prueba VGG+LSVM.

Los resultados obtenidos al variar el umbral de penalización se muestran en la Figura 4.6. Si se considera la desviación estándar de los resultados, se ve que no existe relación

entre el umbral elegido y los resultados obtenidos. Esto se confirma al aplicar la prueba ANOVA en los resultados. Usando esta prueba para comparar el *accuracy* a diferentes umbrales se obtiene un p-value de 0.99, por lo que se concluye que no hay correlación entre el umbral y los resultados.

Con estos resultados, sumado a lo observado en la Sección 4.1.1, se fijó el umbral de penalización al 10 % para todos los experimentos posteriores. Este umbral logra reducir el número de pares con alto crecimiento, sin aumentar en gran medida el crecimiento promedio de los pares.

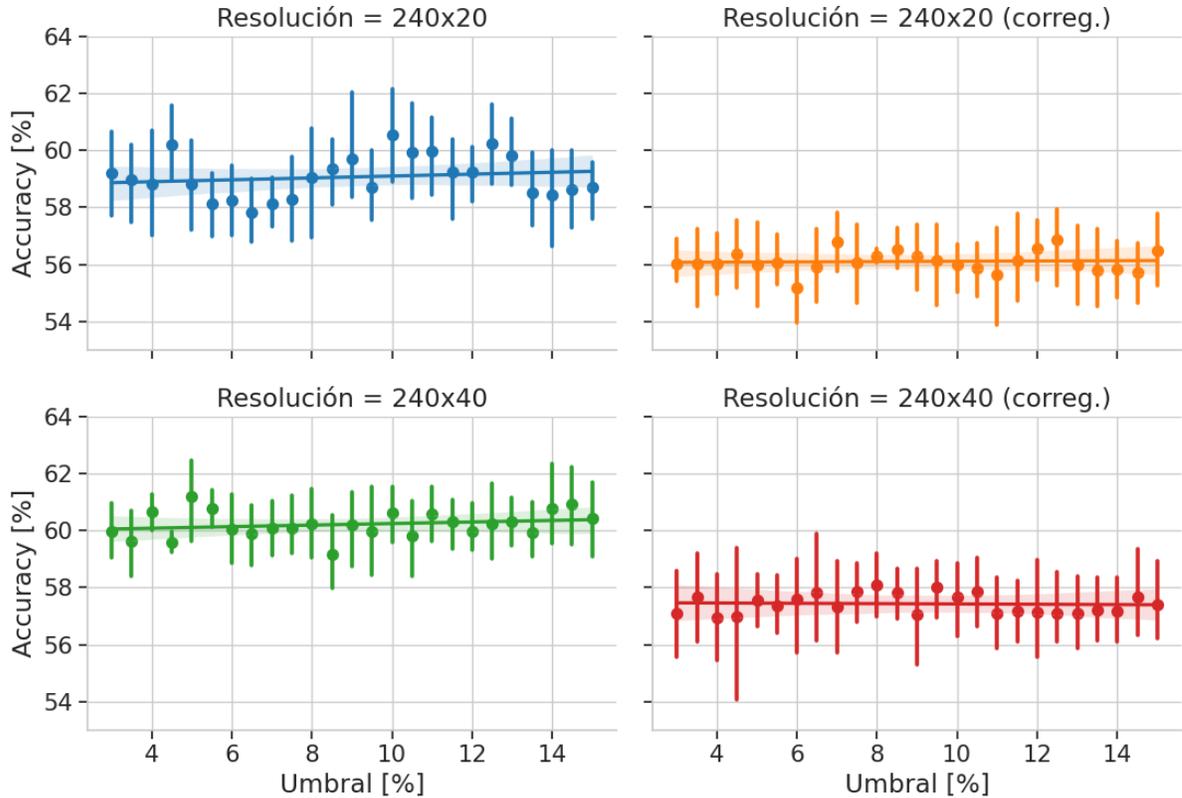


Figura 4.6: Resultados de clasificación obtenidos usando diferentes umbrales de penalización, a distinta resolución. Las resoluciones con “(correg.)” corresponden a uso de máscaras corregidas, y las que no tienen esta palabra corresponden a máscaras originales. La línea recta que atraviesa los resultados es una regresión lineal del promedio de los resultados por umbral.

4.3.3. Experimento 3: Efectos de eliminar pares con alto crecimiento

Se estudiará si los pares con alto crecimiento —y por consiguiente, alta pérdida de información por las máscaras pareadas— afectan los resultados de clasificación. Para esto, se generaron pares de máscaras, y posteriormente se eliminaron los pares con crecimiento mayor a cierto umbral. Este umbral varía entre el 7 % y el 10 %, a intervalos de 1 %. Con

estas imágenes, se realizaron pruebas de clasificación.

Los resultados obtenidos en este experimento se muestran en la Tabla 4.3. Estos resultados muestran que, para una misma resolución y prueba, las diferencias en *accuracy* son mínimas ($< 1\%$ en promedio). Esto se verifica al realizar la prueba de ANOVA. Al estudiar el impacto del umbral de eliminación en el *accuracy* usando ANOVA, se obtiene un p-value alto —cercano a 1.0 en tres de los casos—, por lo que se concluye que eliminar los pares con alto crecimiento no tiene efecto en el *accuracy* al clasificar por género. Esto quiere decir que el que se formen pares con alto crecimiento no influye en los resultados de clasificación.

Tabla 4.3: Resultados usando máscaras pareadas y eliminando pares sobre cierto crecimiento para las pruebas VGG+LSVM y VGG-full. En negrita se indica el *accuracy* máxima para cada resolución y prueba.

Tipo de prueba	Resolución	Umbral	Accuracy [%]	P-value
VGG+LSVM	240×20	10 %	58.66 ± 2.16	0.97
		9 %	58.83 ± 2.19	
		8 %	58.72 ± 2.35	
		7 %	58.90 ± 2.09	
	240×40	10 %	60.48 ± 2.03	0.98
		9 %	60.64 ± 1.98	
		8 %	60.56 ± 2.06	
		7 %	60.41 ± 1.97	
VGG-full	240×20	10 %	58.38 ± 2.59	0.78
		9 %	57.64 ± 2.93	
		8 %	57.78 ± 3.32	
		7 %	57.74 ± 3.37	
	240×40	10 %	59.61 ± 2.91	0.99
		9 %	59.54 ± 2.86	
		8 %	59.72 ± 3.07	
		7 %	59.54 ± 2.83	

4.3.4. Experimento 4: Efectos de usar máscaras pareadas en clasificación

A continuación se mostrarán los resultados de usar máscaras pareadas al clasificar género en base a iris. Para formar los pares, se usó la metodología descrita en la Sección 3.1. El pareo fue realizado sólo en los iris de los conjuntos de entrenamiento, ya que se requiere conocimiento *a priori* del género de los iris. Los resultados de este experimento se muestran en la Tabla 4.4. Estos resultados muestran que, en promedio, se perdió un 4.52 % de *accuracy* al usar máscaras pareadas.

Tabla 4.4: Resultados usando máscaras pareadas para las pruebas VGG+LSVM y VGG-full. En negrita se indican el *accuracy* máximo, la mayor diferencia *versus* los resultados base, y los p-values por debajo de 0.05.

Tipo de prueba	Resolución	Accuracy [%]	Diferencia vs base [%]	P-value
VGG+LSVM	240 × 20	58.72 ± 2.02	-5.25	6.21 × 10⁻¹⁶
	240 × 40	60.62 ± 2.13	-4.81	6.72 × 10⁻¹³
VGG-full	240 × 20	58.36 ± 2.84	-3.77	6.58 × 10⁻⁷
	240 × 40	59.11 ± 3.10	-4.23	4.30 × 10⁻⁷

Estos resultados fueron evaluados usando la prueba ANOVA. Al compararlos con los resultados base y estudiar el impacto de usar o no máscaras pareadas en el *accuracy*, se obtienen p-values muy inferiores a 0.05. Con esto se puede afirmar que los resultados obtenidos son significativos, y que el uso de máscaras pareadas sí influye en los resultados.

4.3.5. Experimento 5: Efectos de usar máscaras corregidas en clasificación

En esta sección se muestran los resultados tras usar máscaras corregidas manualmente al clasificar género en base a iris. La corrección de las máscaras se realizó siguiendo la metodología descrita en la Sección 3.2. A diferencia de las máscaras pareadas, esta corrección se aplicó tanto al conjunto de entrenamiento como al de prueba.

Los resultados obtenidos se muestran en la Tabla 4.5. Como indica la tabla, la corrección de máscaras redujo en promedio un 1.6% el *accuracy* al clasificar. Al observar los p-values obtenidos usando la prueba ANOVA, los resultados son significativos (menores a 0.05) en tres de las cuatro pruebas. Hay que destacar que las correcciones fueron hechas por un operador humano, por lo que éstas podrían tener sesgos o incluso errores.

Tabla 4.5: Resultados usando máscaras corregidas para las pruebas VGG+LSVM y VGG-full. En negrita se indican el *accuracy* máximo, la mayor diferencia *versus* los resultados base, y los p-values menores a 0.05.

Tipo de prueba	Resolución	Accuracy [%]	Diferencia vs base [%]	P-value
VGG+LSVM	240 × 20	62.79 ± 1.67	-1.18	7.38 × 10⁻³
	240 × 40	62.01 ± 2.12	-3.42	1.68 × 10⁻⁸
VGG-full	240 × 20	61.69 ± 2.72	-0.44	5.06 × 10 ⁻¹
	240 × 40	61.89 ± 2.95	-1.45	4.81 × 10⁻²

4.3.6. Experimento 6: Efectos de usar máscaras pareadas y corregidas en clasificación

El usar máscaras corregidas podría haber añadido información dentro de la forma de las máscaras, tal y como se mencionó en la Sección 3.1. Para eliminar este efecto, se aplicó el pareo de máscaras a las máscaras corregidas. Los resultados de este experimento se muestran en la Tabla 4.6.

Tabla 4.6: Resultados usando máscaras corregidas y pareadas para las pruebas VGG+LSVM y VGG-full. En negrita se indican el *accuracy* máximo, la mayor diferencia *versus* los resultados sin parear, la mayor diferencia *versus* los resultados sin corregir, y la mayor diferencia *versus* los resultados base.

Tipo de prueba	Resolución	Accuracy [%]	Dif. vs sin pares [%]	Dif. vs sin correc. [%]	Dif. vs base [%]
VGG+LSVM	240 × 20	56.42 ± 1.77	-6.37	-2.30	-7.55
	240 × 40	56.96 ± 1.69	-5.05	-3.66	-8.47
VGG-full	240 × 20	56.39 ± 3.36	-5.30	-1.97	-5.74
	240 × 40	55.99 ± 2.82	-5.90	-3.12	-7.35

Al juntar los efectos de corregir y de parear las máscaras, se perdió en promedio un 7.28% de *accuracy*. Esa diferencia es mayor a la suma de aplicar los pares y las correcciones por separado. La diferencia entre estos resultados y los obtenidos usando sólo corrección de máscaras (Tabla 4.5) fue de -5.66% en promedio. La diferencia entre estos resultados y los obtenidos usando sólo máscaras pareadas (Tabla 4.4) fue de -2.76% en promedio. Para ambos factores, el efecto en el *accuracy* se vio acentuado con respecto a la aplicación de dichos factores por separado.

Las pruebas con y sin máscaras pareadas y máscaras corregidas se resumen en la Tabla 4.7. El efecto en la clasificación, tanto de haber corregido las máscaras como de haber formado pares, se estudió usando una prueba ANOVA bidireccional. Los resultados de esta prueba se muestran en la Tabla 4.8. En esta prueba se usaron los resultados de dos pruebas, por lo que se tiene el doble de datos que en las pruebas ANOVA anteriores. En todas las pruebas, el impacto de las máscaras pareadas y corregidas fue significativo (p-value menor a 0.05). Por otro lado, la interacción entre los dos factores no alcanza a ser significativa. Con estos resultados se corrobora que las máscaras juegan un papel importante en la clasificación de genero.

4.3.7. Experimento 7: Resultados usando la base de datos ND-CrossSensor-Iris-2013

Para verificar que el comportamiento observado usando máscaras pareadas y la base de datos GFI es generalizable a otras imágenes, se repitió el experimento de parear máscaras

Tabla 4.7: Resumen de resultados usando máscaras corregidas y pareadas para las pruebas VGG+LSVM y VGG-full, con la base de datos GFI. En negrita se indica el *accuracy* máximo por prueba y resolución.

Prueba	Resolución	Factores		Accuracy [%]
		Máscaras pareadas	Máscaras corregidas	
VGG+LSVM	240 × 20	NO	NO	63.97 ± 1.64
			SÍ	62.79 ± 1.67
		SÍ	NO	58.72 ± 2.02
			SÍ	56.42 ± 1.77
	240 × 40	NO	NO	65.43 ± 1.93
			SÍ	62.01 ± 2.12
		SÍ	NO	60.62 ± 2.13
			SÍ	56.96 ± 1.69
VGG-full	240 × 20	NO	NO	62.13 ± 2.37
			SÍ	61.69 ± 2.72
		SÍ	NO	58.36 ± 2.84
			SÍ	56.39 ± 3.36
	240 × 40	NO	NO	63.34 ± 2.64
			SÍ	61.89 ± 2.95
		SÍ	NO	59.11 ± 3.10
			SÍ	55.99 ± 2.82

en la base de datos ND-CrossSensor-Iris-2013. Los resultados base para esta base de datos se muestran en la Tabla 4.9.

Tabla 4.9: Resultados usando la base de datos ND-CrossSensor-Iris-2013 para las pruebas VGG+LSVM y VGG-full. En negrita se indica el *accuracy* máximo.

Tipo de prueba	Resolución	Accuracy [%]
VGG+LSVM	240 × 20	61.44 ± 2.71
	240 × 40	61.10 ± 2.54
VGG-full	240 × 20	59.90 ± 2.68
	240 × 20	60.13 ± 3.05

Los resultados usando máscaras pareadas en la base de datos ND-CrossSensor-Iris-2013 se muestran en la Tabla 4.10. Esta tabla muestra que, en promedio, las máscaras pareadas reducen entre un 2.95 % y un 4.00 % el *accuracy* al clasificar género. Este efecto es comparable, pero menor, al observado en la base de datos GFI. Esto también se ve reflejado en los p-values obtenidos con la prueba ANOVA. En todos los casos, el impacto de las más-

Tabla 4.8: P-values obtenidos con prueba ANOVA bidireccional para cada prueba, resolución y factor. En negrita se indican los p-values menores a 0.05.

Prueba	Resolución	Factor	P-value
VGG+LSVM	240 × 20	Máscaras pareadas	3.85 × 10⁻³⁵
		Máscaras corregidas	4.22 × 10⁻⁷
		Interacción entre factores	8.97 × 10 ⁻²
	240 × 40	Máscaras pareadas	6.26 × 10⁻²⁶
		Máscaras corregidas	6.41 × 10⁻¹⁷
		Interacción entre factores	7.54 × 10 ⁻¹
VGG-full	240 × 20	Máscaras pareadas	2.16 × 10⁻¹⁴
		Máscaras corregidas	2.20 × 10⁻²
		Interacción entre factores	1.44 × 10 ⁻¹
	240 × 40	Máscaras pareadas	1.75 × 10⁻¹⁶
		Máscaras corregidas	2.90 × 10⁻⁵
		Interacción entre factores	1.16 × 10 ⁻¹

caras pareadas fue significativo, pero los p-values fueron mayores a lo observado en la base de datos GFI. Estos resultados permitirían plantear que la metodología es aplicable a otra base de datos de iris.

4.4. Discusión de resultados

Usando pares de máscaras se igualaron las máscaras de ambos géneros, lo que eliminó la posible correlación entre el género y la forma de las máscaras. Al eliminar esta correlación, el *accuracy* en la clasificación de género disminuyó en un 4.52 % en promedio en las base de datos GFI y 3.54 % en la base de datos ND-CrossSensor-Iris-2013, diferencia que resultó ser significativa en todas las pruebas.

Tabla 4.10: Resultados usando mascararas pareadas y la base de datos ND-CrossSensor-Iris-2013 para las pruebas VGG+LSVM y VGG-full. En negrita se indican el *accuracy* máximo, la mayor diferencia *versus* los resultados base, y los p-values menores a 0.05.

Tipo de prueba	Resolución	Accuracy [%]	Diferencia vs base [%]	P-value
VGG+LSVM	240 × 20	57.44 ± 2.82	-4.00	6.06 × 10⁻⁷
	240 × 40	57.25 ± 2.60	-3.85	2.73 × 10⁻⁷
VGG-full	240 × 20	56.54 ± 2.28	-3.36	2.46 × 10⁻⁶
	240 × 20	57.18 ± 2.95	-2.95	3.46 × 10⁻⁴

La corrección manual de máscaras eliminó los elementos que no eran parte del iris y permitió cuantificar su impacto. El uso de máscaras corregidas resultó en una reducción promedio del 1.6%, efecto que resultó ser significativo en la mitad de las pruebas. Cabe recordar que las máscaras fueron corregidas por un humano, lo que podría haber introducido sesgos o errores a las máscaras. Si se quisiera mitigar este factor, habría que cambiar la técnica de segmentación. Por ejemplo, se podría usar la metodología descrita en trabajos más recientes [27–30] para reducir los errores de segmentación.

Al juntar los dos factores, pares y corrección, su efecto se vio acentuado. Cuando las máscaras estaban corregidas, usar pares de máscaras disminuyó en un 5.66% el *accuracy* en la clasificación de género; este efecto fue significativo en todas las pruebas. Asimismo, cuando las máscaras estaban pareadas, corregir las máscaras disminuyó en un 2.76% el *accuracy* en la clasificación de género; este efecto fue significativo en todas las pruebas. **En total, las máscaras contribuyeron en un 7.28% en promedio al *accuracy* de la clasificación de género en base a iris.**

A pesar de haber eliminado la información de género contenida en las máscaras e información en el iris durante el pareo y las correcciones, el *accuracy* de clasificación de género en base a iris se mantuvo arriba del 55%. Esto quiere decir que sí hay información de género presente en la textura del iris. En caso contrario, el *accuracy* de clasificación de género habría rondado el 50%. Sería de interés el estudiar la información de género en la textura del iris para determinar su origen, ya que es posible que aún queden sesgos no eliminados.

En términos de *accuracy*, los resultados obtenidos son consistentes con los reportados en [8] en la base de datos GFI (Tabla 4.2). El *accuracy* que reportaron fueron de 60.0% en la prueba VGG+LSVM y de 60.1% en la prueba VGG-full. Los resultados obtenidos en esta tesis sin usar máscaras pareadas fueron mejores (por 3.67%), mientras que los resultados con máscaras pareadas fueron peores (por 0.85%). Esto significa que a pesar de tener resultados base más altos que en [8], la reducción en desempeño al usar máscaras pareadas fue significativa.

Por otro lado, los resultados de esta tesis no son consistentes con los trabajos anteriores que reportaron tasas de clasificación superiores al 70% al usar iris normalizados [6, 7, 17]. No obstante, ninguno de estos trabajos considera el sesgo que las máscaras podrían introducir.

La diferencia en los resultados entre esta tesis y trabajos como [6, 7, 17] podría deberse a distintos factores. En [8, 20] se mencionan algunos factores, como el uso de particiones de entrenamiento y prueba no separadas por sujeto, la presencia de información que no corresponde al iris, y la utilización de sólo una partición de prueba. Sin embargo, no se cuenta con suficiente información para determinar con certeza la causa de esta diferencia.

La información que es añadida por las máscaras puede ser usada para clasificar género pues se correlaciona con el género por diversos factores, tales como el maquillaje y las pestañas. Por lo tanto, el remover esta información al clasificar género en base a iris normalizados debería realizarse si lo que se desea es evaluar la clasificación de género basado sólo en la información del iris. Además, las máscaras dependen de factores que no están relacionados con el género, como el algoritmo de segmentación usado y las condiciones durante la adquisición de la imagen.

El contenido de información de género en el iris que se pierde al aplicar las máscaras pareadas no ha sido cuantificado. El que el *accuracy* no disminuyera a pesar de eliminar los peores pares parece sugerir que el efecto de aplicar máscaras pareadas, el cual elimina la correlación entre las máscaras y el género, es mayor al efecto de perder información al parear máscaras.

Capítulo 5

Conclusiones y trabajo futuro

En este trabajo se cumplieron los objetivos de definir una metodología para remover información de género que podría estar presente en las máscaras de oclusión, y de medir el impacto de éstas. La metodología consistió en formar pares de iris de distinto género, formar máscaras pareadas, y aplicar estas máscaras a los iris. También se corrigieron manualmente las máscaras, eliminando elementos que no eran parte del iris. Al usar esta metodología en la clasificación de género en base a iris, se demostró que las máscaras contribuyen en promedio en un 7.28 % al *accuracy* de la clasificación.

En base a estos resultados, se verifican las hipótesis planteadas. Se demostró que las pestañas, párpados y otros elementos que no son parte del iris y que son visibles en las imágenes normalizadas de iris, contienen información que un clasificador puede usar para predecir género. Este efecto resultó ser significativo en algunos casos. Además, se verificó que las máscaras contienen información de género, y que al igualar las máscaras de ambos géneros, las tasas de clasificación de género disminuyen significativamente.

Dado lo anterior, si lo que se desea es cuantificar la información de género presente en la textura del iris, los métodos usados para clasificar iris normalizados deberían usar la metodología propuesta para eliminar la información de género presente en las máscaras de oclusión. Esta metodología también acota la cantidad información de iris que se pierde al aplicar las máscaras pareadas gracias a que se aplica optimización combinatoria.

La metodología usada en este trabajo podría ser mejorada de las siguientes formas. En primer lugar, se podría usar una base de datos más numerosa y con etiquetas de maquillaje como la usada en [8]. Esto permitiría estudiar los efectos del maquillaje en la generación de máscaras y en la clasificación de género. Además, no se ha estudiado el impacto del maquillaje en sujetos masculinos en la clasificación de género. Generar una base de datos de estas características abriría el estudio de *gender-spoofing* (utilización de técnicas para lograr una clasificación específica) en la clasificación de género en base a iris, tanto normalizado como periocular. Otro factor a mejorar es el algoritmo de segmentación y normalización usado. En la actualidad existen métodos más modernos basados en *deep-learning*. El uso de estos métodos podría volver prescindible la corrección manual de máscaras dependiendo de su desempeño en segmentación, y los efectos de formar pares de máscaras en este caso podrían

ser diferentes.

Aún quedan interrogantes respecto de la robustez de las metodologías de clasificación de género en base a iris normalizado. Al eliminar la información presente en las máscaras, la tasa de clasificación siguió siendo levemente superior al 50%, por lo que hay información de género presente en el iris. Como trabajo futuro, se podría determinar en qué medida el formar pares elimina información del iris, en contraste con la información de las máscaras, para corroborar si el aporte en información viene principalmente de la máscara. También se podría estudiar otras potenciales fuentes de sesgo en los resultados, por ejemplo, ecualizando la intensidad promedio de los píxeles de las imágenes cada género. Finalmente, se podría estudiar el impacto de trasladar horizontalmente el iris normalizado. Esto equivaldría a rotar el iris, situación que puede darse durante la adquisición de la imagen. Si los algoritmos de clasificación de género son robustos, éstos deberían funcionar aún si se introduce rotación. Por ejemplo, la clasificación según género de algún iris no debería cambiar si el iris es rotado. Si la clasificación cambia al rotar el iris, significaría que el algoritmo de clasificación de género no es robusto.

Bibliografía

- [1] K. W. Bowyer, K. Hollingsworth, and P. J. Flynn, “Image understanding for iris biometrics: A survey,” *Computer Vision and Image Understanding*, vol. 110, no. 2, pp. 281–307, 2008.
- [2] J. Daugman, “How Iris Recognition Works,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 1, pp. 21–30, 2004.
- [3] J. Daugman, “Iris recognition at airports and border-crossings,” in *Encyclopedia of Biometrics* (S. Z. Li and A. Jain, eds.), pp. 819–825, Boston, MA: Springer US, 2009.
- [4] Y. H. Li and M. Savvides, “An automatic iris occlusion estimation method based on high-dimensional density estimation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 4, pp. 784–796, 2013.
- [5] V. Thomas, N. V. Chawla, K. W. Bowyer, and P. J. Flynn, “Learning to predict gender from iris images,” in *2007 1st IEEE Int. Conf. Biometrics: Theory, Applications, and Systems*, (Crystal City, VA, USA), pp. 1–5, sep 2007.
- [6] J. E. Tapia, C. A. Perez, and K. W. Bowyer, “Gender classification from the same iris code used for recognition,” *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 8, pp. 1760–1770, 2016.
- [7] J. Tapia and C. A. Perez, “Gender Classification from NIR Images by Using Quadrature Encoding Filters of the Most Relevant Features,” *IEEE Access*, vol. 7, no. c, pp. 29114–29127, 2019.
- [8] A. Kuehlkamp and K. W. Bowyer, “Predicting gender from iris texture may be harder than it seems,” in *2019 IEEE Winter Conf. Appl. of Comput. Vision (WACV)*, (Waikoloa Village, HI, USA), pp. 904–912, jan 2019.
- [9] D. Reid, S. Samangoei, C. Chen, M. Nixon, and A. Ross, “Chapter 13 - soft biometrics for surveillance: An overview,” in *Handbook of Statistics* (C. Rao and V. Govindaraju, eds.), vol. 31 of *Handbook of Statistics*, pp. 327–352, Elsevier, 2013.
- [10] A. Dantcheva, C. Velardo, A. D’Angelo, and J.-L. Dugelay, “Bag of soft biometrics for person identification: New trends and challenges,” *Multimedia Tools and Applications*, vol. 51, pp. 739–777, 2011.
- [11] M. Edwards, D. Cha, S. Krithika, M. Johnson, and E. J. Parra, “Analysis of iris surface features in populations of diverse ancestry,” *Royal Society Open Science*, vol. 3, no. 1, p. 150424, 2016.
- [12] M. Larsson and N. L. Pedersen, “Genetic correlations among texture characteristics in

- the human iris,” *Molecular vision*, vol. 10, pp. 821–31, 2004.
- [13] J. A. Sanchis-Gimeno, D. Sanchez-Zuriaga, and F. Martinez-Soriano, “White-to-white corneal diameter, pupil diameter, central corneal thickness and thinnest corneal thickness values of emmetropic subjects,” *Surgical and Radiologic Anatomy*, vol. 34, no. 2, pp. 167–170, 2012.
- [14] H. Wagner, B. A. Fink, and K. Zadnik, “Sex- and gender-based differences in healthy and diseased eyes,” *Optometry - Journal of the American Optometric Association*, vol. 79, no. 11, pp. 636–652, 2008.
- [15] T. Suzuki, S. M. Richards, S. Liu, R. V. Jensen, and D. A. Sullivan, “Influence of sex on gene expression in human corneal epithelial cells,” *Molecular vision*, vol. 15, pp. 2554–69, 2009.
- [16] A. Bansal, R. Agarwal, and R. Sharma, “SVM Based Gender Classification Using Iris Images,” in *2012 4th Int. Conf. Comput. Intell. and Commun. Netw.*, (Mathura, India), pp. 425–429, nov 2012.
- [17] J. Tapia and C. Aravena, “Gender classification from nir iris images using deep learning,” *Deep Learning for Biometrics: A Survey. ACM Comput. Surv.*, vol. 51, pp. 219–239, 2017.
- [18] J. Tapia, C. A. Perez, and K. Bowyer, “Gender classification from iris images using fusion of uniform local binary patterns,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 8926, pp. 751–763, 2015.
- [19] D. Bobeldyk and A. Ross, “Iris or periocular? Exploring sex prediction from near infrared ocular images,” *Lecture Notes in Informatics (LNI), Proceedings - Series of the Gesellschaft fur Informatik (GI)*, vol. P-260, 2016.
- [20] A. Kuehlkamp, B. Becker, and K. W. Bowyer, “Gender-from-iris or gender-from-mascara?,” in *2017 IEEE Winter Conf. Appl. of Comput. Vision (WACV)*, (Santa Rosa, CA, USA), pp. 1151–1159, mar 2017.
- [21] N. Othman, S. Guillaume, B. Dorizzi, and S. Garcia-Salicetti, “A biometric reference system for iris OSIRIS version 4.1,” *BioSecure project, Telecom Sud Paris*, 2013.
- [22] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” in *3rd Int. Conf. Learn. Representations (ICLR 2015)* (Y. Bengio and Y. LeCun, eds.), (San Diego, CA, USA), may 2015.
- [23] C. Cortes and V. Vapnik, “Support-vector networks,” *Machine Learning*, vol. 20, pp. 273–297, sep 1995.
- [24] A. K. Jain, A. Ross, and S. Prabhakar, “An Introduction to Biometric Recognition,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 1, pp. 4–20, 2004.
- [25] M. S. Nixon, P. L. Correia, K. Nasrollahi, T. B. Moeslund, A. Hadid, and M. Tistarelli, “On soft biometrics,” *Pattern Recognition Letters*, vol. 68, pp. 218–230, 2015.
- [26] A. Dantcheva, P. Elia, and A. Ross, “What Else Does Your Biometric Data Reveal? A Survey on Soft Biometrics,” *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 3, pp. 441–467, 2016.
- [27] M. Trokielewicz and A. Czajka, “Data-driven segmentation of post-mortem iris images,”

- in *2018 Int. Workshop on Biometrics and Forensics (IWBF)*, (Sassari, Italy), pp. 1–7, jun 2018.
- [28] A. Valenzuela, C. Arellano, and J. E. Tapia, “Towards an efficient segmentation algorithm for near-infrared eyes images,” *IEEE Access*, vol. 8, pp. 171598–171607, 2020.
- [29] E. Jalilian, M. Karakaya, and A. Uhl, “End-to-end off-angle iris recognition using CNN based iris segmentation,” in *BIOSIG 2020 – Proc. 19th Int. Conf. Biometrics Special Interest Group* (A. Brömme, C. Busch, A. Dantcheva, K. B. Raja, C. Rathgeb, and A. Uhl, eds.), vol. P-306, (online), pp. 117–128, sep 2020.
- [30] A. Kuehlkamp, A. Boyd, A. Czajka, K. W. Bowyer, P. J. Flynn, D. Chute, and E. Benjamin, “Interpretable deep learning-based forensic iris segmentation and recognition,” in *IEEE/CVF Winter Conf. Appl. of Comput. Vision (WACV) Workshops*, (Waikoloa, HI, USA), pp. 359–368, jan 2022.
- [31] N. Othman, B. Dorizzi, and S. Garcia-Salicetti, “OSIRIS: An open source iris recognition software,” *Pattern Recognition Letters*, vol. 82, pp. 124–131, 2016.
- [32] G. Sutra, S. Garcia-Salicetti, and B. Dorizzi, “The Viterbi algorithm at different resolutions for enhanced iris segmentation,” in *5th IAPR Int. Conf. Biometrics (ICB 2012)* (A. K. Jain, A. Ross, S. Prabhakar, and J. Kim, eds.), (New Delhi, India), pp. 310–316, mar 2012.
- [33] R. Connaughton, A. Sgroi, K. Bowyer, and P. J. Flynn, “A multialgorithm analysis of three iris biometric sensors,” *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 3, pp. 919–931, 2012.
- [34] G. D. Forney, “The viterbi algorithm,” *Proceedings of the IEEE*, vol. 61, no. 3, pp. 268–278, 1973.
- [35] R. Gross and V. Brajovic, “An image preprocessing algorithm for illumination invariant face recognition,” in *Audio-and Video-Based Biometric Person Auth., 4th Int. Conf. (AVBPA 2003)* (J. Kittler and M. S. Nixon, eds.), vol. 2688 of *Lecture Notes in Computer Science*, (Guildford, UK), pp. 10–18, jun 2003.
- [36] A. D. Clark, S. A. Kulp, I. H. Herron, and A. A. Ross, “A Theoretical Model for Describing Iris Dynamics,” in *Handbook of Iris Recognition*, pp. 129–150, London: Springer London, 2013.
- [37] I. Tomeo-Reyes, A. Ross, A. D. Clark, and V. Chandran, “A biomechanical approach to iris normalization,” in *Int. Conf. Biometrics (ICB 2015)*, (Phuket, Thailand), pp. 9–16, may 2015.
- [38] D. Bobeldyk and A. Ross, “Predicting gender and race from near infrared iris and periocular images,” *CoRR*, vol. abs/1805.01912, 2018.
- [39] M. Singh, S. Nagpal, M. Vatsa, R. Singh, A. Noore, and A. Majumdar, “Gender and ethnicity classification of iris images using deep class-encoder,” in *2017 IEEE Int. Joint Conf. Biometrics (IJCB)*, (Denver, CO, USA), pp. 666–673, oct 2017.
- [40] J. E. Tapia and C. C. Aravena, “Gender classification from periocular NIR images using fusion of cnns models,” in *2018 IEEE 4th Int. Conf. Identity, Secur., and Behav. Anal. (ISBA)*, (Singapore), pp. 1–6, jan 2018.
- [41] J. Tapia, C. Rathgeb, and C. Busch, “Sex-Prediction from Periocular Images Across

- Multiple Sensors and Spectra,” *Proceedings - 14th International Conference on Signal Image Technology and Internet Based Systems, SITIS 2018*, pp. 529–535, 2019.
- [42] I. Viedma, J. Tapia, A. Iturriaga, and C. Busch, “Relevant features for gender classification in NIR periocular images,” *IET Biometrics*, vol. 8, no. 5, pp. 340–350, 2019.
- [43] L. Breiman, “Bagging predictors,” *Machine Learning*, vol. 24, pp. 123–140, aug 1996.
- [44] S. Lagree and K. W. Bowyer, “Predicting ethnicity and gender from iris texture,” in *2011 IEEE Int. Conf. Technol. for Homeland Secur. (HST)*, (Waltham, MA, USA), pp. 440–445, nov 2011.
- [45] K. I. Laws, “Rapid Texture Identification,” in *Proceedings SPIE* (T. F. Wiener, ed.), vol. D, pp. 376–381, 1980.
- [46] M. C. Fairhurst, M. Erbilek, and M. C. Da Costa-Abreu, “Exploring gender prediction from iris biometrics,” in *BIOSIG 2015 – Proc. 14th Int. Conf. Biometrics Special Interest Group* (A. Brömme, C. Busch, C. Rathgeb, and A. Uhl, eds.), vol. P-245 of *LNI*, (Darmstadt, Germany), pp. 223–230, sep 2015.
- [47] J. E. Tapia and I. Viedma, “Gender classification from multispectral periocular images,” in *2017 IEEE Int. Joint Conf. Biometrics (IJCB)*, (Denver, CO, USA), pp. 805–812, oct 2017.
- [48] A. Rattani, N. Reddy, and R. Derakhshani, “Gender prediction from mobile ocular images: A feasibility study,” in *2017 IEEE Int. Symp. Technol. for Homeland Secur. (HST)*, (Waltham, MA, USA), pp. 1–6, apr 2017.
- [49] A. Rattani, N. Reddy, and R. Derakhshani, “Convolutional neural networks for gender prediction from smartphone-based ocular images,” *IET Biometrics*, vol. 7, no. 5, pp. 423–430, 2018.
- [50] I. Viedma and J. E. Tapia, “Deep gender classification and visualization of near-infra-red periocular-iris images,” in *2018 IEEE Int. Conf. Image Process., Appl. and Syst. (IPAS)*, (Sophia Antipolis, France), pp. 73–78, dec 2018.
- [51] D. Bobeldyk and A. Ross, “Predicting soft biometric attributes from 30 pixels: A case study in NIR ocular images,” *2019 IEEE Winter Conf. Appl. Comput. Vision Workshops (WACVW)*, pp. 116–124, jan 2019.
- [52] A. Manyala, H. Cholakkal, V. Anand, V. Kanhangad, and D. Rajan, “CNN-based gender classification in near-infrared periocular images,” *Pattern Analysis and Applications*, vol. 22, no. 4, pp. 1493–1504, 2019.
- [53] J. E. Tapia and C. Arellano, “Soft-biometrics encoding conditional GAN for synthesis of NIR periocular images,” *Future Generation Computer Systems*, vol. 97, pp. 503–511, 2019.
- [54] J. Tapia and C. Arellano, “Gender Classification from Iris Texture Images Using a New Set of Binary Statistical Image Features,” in *2019 Int. Conf. Biometrics (ICB)*, (Crete, Greece), pp. 1–7, jun 2019.
- [55] J. Tapia, C. Arellano, and I. Viedma, “Sex-classification from Cellphones Periocular Iris Images,” in *Selfie Biometrics: Advances and Challenges*, pp. 227–242, Cham: Springer International Publishing, 2019.
- [56] M. Eskandari and O. Sharifi, “Effect of face and ocular multimodal biometric systems

- on gender classification,” *IET Biometrics*, vol. 8, no. 4, pp. 243–248, 2019.
- [57] M. Fang, N. Damer, F. Kirchbuchner, and A. Kuijper, “Real masks and spoof faces: On the masked face presentation attack detection,” *Pattern Recognition*, vol. 123, p. 108398, 2022.
- [58] M. Ngan, P. Grother, and K. Hanaoka, “Ongoing face recognition vendor test (frvt) part 6b: Face recognition accuracy with face masks using post-covid-19 algorithms,” in *NIST Interagency/Internal Report (NISTIR)*, National Institute of Standards and Technology, Gaithersburg, MD, 2020.
- [59] N. Damer, J. H. Grebe, C. Chen, F. Boutros, F. Kirchbuchner, and A. Kuijper, “The effect of wearing a mask on face recognition performance: an exploratory study,” in *2020 Int. Conf. Biometrics Special Interest Group (BIOSIG)*, (Darmstadt, Germany), pp. 1–6, sep 2020.
- [60] N. Damer, F. Boutros, M. Süßmilch, F. Kirchbuchner, and A. Kuijper, “Extended evaluation of the effect of real and simulated masks on face recognition performance,” *IET biometrics*, vol. 10, no. 5, pp. 548–561, 2021.
- [61] S. Seneviratne, N. Kasthuriarachchi, S. Rasnayaka, D. Hettiachchi, and R. Shariffdeen, “Does a face mask protect my privacy?: Deep learning to predict protected attributes from masked face images,” in *AI 2021: Advances in Artificial Intelligence* (G. Long, X. Yu, and S. Wang, eds.), (Cham), pp. 91–102, Springer International Publishing, 2022.
- [62] F. Alonso-Fernandez, K. H. Diaz, S. Ramis, F. J. Perales, and J. Bigun, “Soft-Biometrics Estimation In the Era of Facial Masks,” *Lecture Notes in Informatics (LNI), Proceedings - Series of the Gesellschaft fur Informatik (GI)*, vol. P-306, pp. 11–19, 2020.
- [63] K. B. Raja, R. Raghavendra, and C. Busch, “Fused Spectral Features in Kernel Weighted Collaborative Representation for Gender Classification Using Ocular Images,” in *Proc. 3rd Int. Conf. Comput. Vision and Image Processing 2018* (B. B. Chaudhuri, M. Nakagawa, P. Khanna, and S. Kumar, eds.), (Jabalpur, India), pp. 131–143, 2020.
- [64] G. S. Sable and M. R. Rajput, “Radon and Multiwavelet-Based Compact Feature Vector Generation for Gender Identification from Iris,” in *Advances in Intelligent Systems and Computing*, vol. 1025, pp. 29–43, Springer, 2020.
- [65] F. Alonso-Fernandez, K. Hernandez-Diaz, S. Ramis, F. J. Perales, and J. Bigun, “Facial masks and soft-biometrics: Leveraging face recognition CNNs for age and gender prediction on mobile ocular images,” *IET Biometrics*, vol. 10, no. 5, pp. 562–580, 2021.
- [66] N. Vetrekar, R. Ramachandra, K. Raja, and R. Gad, “Gender classification under eye-glass occluded ocular region: An extensive study using multi-spectral imaging,” in *AI and Deep Learning in Biometric Security*, pp. 175–203, CRC Press, 2021.
- [67] A. R. Khan, F. Doosti, M. Karimi, M. Harouni, U. Tariq, S. M. Fati, and S. Ali Bahaj, “Authentication through gender classification from iris images using support vector machine,” *Microscopy Research and Technique*, vol. 84, no. 11, pp. 2666–76, 2021.
- [68] Y. LeCun *et al.*, “Generalization and network design strategies,” *Connectionism in perspective*, vol. 19, no. 143-155, p. 18, 1989.
- [69] N. Kriegeskorte, “Deep Neural Networks: A New Framework for Modeling Biological Vision and Brain Information Processing,” *Annual Review of Vision Science*, vol. 1, no. 1, pp. 417–446, 2015.

- [70] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.
- [71] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, “Learning representations by back-propagating errors,” *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.
- [72] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, *et al.*, “Imagenet large scale visual recognition challenge,” *International journal of computer vision*, vol. 115, no. 3, pp. 211–252, 2015.
- [73] R. Burkard, M. Dell’Amico, and S. Martello, *Assignment Problems*. Philadelphia, PA: Society for Industrial and Applied Mathematics, 2009.
- [74] H. W. Kuhn, “The Hungarian method for the assignment problem,” *Naval Research Logistics Quarterly*, vol. 2, no. 1-2, pp. 83–97, 1955.
- [75] R. Jonker and A. Volgenant, “A shortest augmenting path algorithm for dense and sparse linear assignment problems,” *Computing*, vol. 38, no. 4, pp. 325–340, 1987.
- [76] D. F. Crouse, “On implementing 2D rectangular assignment algorithms,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 52, no. 4, pp. 1679–1696, 2016.
- [77] N. Popescu-Bodorin, L. S. Grigore, V. E. Balas, C. M. Noaica, I. Axenie, J. Popa, C. Munteanu, V. Stroescu, I. Manu, A. Herea, *et al.*, “Cross-sensor iris recognition: LG4000-to-LG2200 comparison,” *arXiv preprint arXiv:1801.01695*, 2013.
- [78] H. D. Macedo and J. N. Oliveira, “Typing linear algebra: A biproduct-oriented approach,” *Science of Computer Programming*, vol. 78, no. 11, pp. 2160–2191, 2013.
- [79] B. Bischl, O. Mersmann, H. Trautmann, and C. Weihs, “Resampling methods for meta-model validation with recommendations for evolutionary computation,” *Evolutionary Computation*, vol. 20, no. 2, pp. 249–275, 2012.
- [80] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, “TensorFlow: Large-scale machine learning on heterogeneous systems,” 2015. Software available from tensorflow.org.
- [81] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “Imagenet: A large-scale hierarchical image database,” in *2009 IEEE Conf. Comput. Vision and Pattern Recognit.*, (Miami, FL, USA), pp. 248–255, jun 2009.
- [82] R. A. Fisher, “Xv.—the correlation between relatives on the supposition of mendelian inheritance,” *Earth and Environmental Science Transactions of the Royal Society of Edinburgh*, vol. 52, no. 2, pp. 399–433, 1919.
- [83] S. S. Shapiro and M. B. Wilk, “An Analysis of Variance Test for Normality (Complete Samples),” *Biometrika*, vol. 52, no. 3/4, p. 591, 1965.
- [84] S. Seabold and J. Perktold, “statsmodels: Econometric and statistical modeling with python,” in *9th Python in Science Conf.*, (Austin, TX, USA), jun 2010. [Online]. Available: https://conference.scipy.org/scipy2010/slides/skipper_seabold_statsmodels.pdf.
- [85] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wil-

son, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, Í. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python,” *Nature Methods*, vol. 17, pp. 261–272, 2020.

- [86] R. Bedre, “reneshbedre/bioinfokit: Bioinformatics data analysis and visualization toolkit,” July 2020. [Online]. Available: <https://doi.org/10.5281/zenodo.3965241>.

Anexos

Anexo A. Listado de Acrónimos

- **VGG**: Red convolucional desarrollada por el *Visual Geometry Group*. En particular, en este trabajo se refiere a la VGGNet-16.
- **SVM**: Máquina de vectores de soporte, *Support Vector Machine* en inglés.
- **LSVM**: Máquina de vectores de soporte lineal, *Linear Support Vector Machine* en inglés.
- **OSIRIS**: Software de código abierto para iris, *Open Source for IRIS* en inglés.
- **LBP**: Patrones binarios locales, *Local Binary Patterns* en inglés.
- **ULBP**: Patrones binarios locales uniformes, *Uniform Local Binary Patterns* en inglés.
- **MLP**: Perceptrón multi-capas, *Multi-Layer Perceptron* en inglés.
- **KNN**: K -vecinos más cercanos, *K-Nearest Neighbors* en inglés.
- **BSIF**: Característica de imagen estadística binarizada, *Binarized Statistical Image Feature* en inglés.
- **NIR**: (Espectro) infrarrojo cercano, *Near-Infrared* en inglés.
- **HOG**: Histogramas de gradientes orientados, *Histogram of Oriented Gradients* en inglés.
- **CNN**: Red neuronal convolucional, *Convolutional Neural Network* en inglés.
- **CMIM**: Maximización de información condicional mutua, *Conditional Mutual Information Maximization* en inglés.
- **GFI**: Base de datos de género para iris, *Gender For Iris database* en inglés.
- **ReLU**: Unidad lineal rectificadora, *Rectified Linear Unit* en inglés.
- **M_G**: Máximo crecimiento de máscara, *maximum mask growth* en inglés.
- **ANOVA**: Análisis de varianza, *Analysis Of Variance* en inglés.
- **FC**: Capa completamente conectada, *Fully Connected layer* en inglés.
- **CV**: Validación cruzada, *Cross-Validation* en inglés.

Anexo B. Resultados en imágenes de iris perioculares

Se probó clasificar género en base a imágenes de iris perioculares para contrastar con los resultados obtenidos en imágenes de iris normalizados. Para ambas pruebas, las imágenes fueron escaladas a 224×224 píxeles. Los resultados se muestran en la Tabla B.1. Éstos muestran un *accuracy* significativamente mayor al obtenido usando imágenes normalizadas; en promedio, los resultados obtenidos usando imágenes perioculares son un 19.11 % mayores a los obtenidos usando imágenes de iris normalizados. Además, estos resultados son 1.83 % y 13.5 % mayores a los obtenidos por Kuehlkamp y Bowyer en [8], en la prueba VGG+LSVM y VGG-full respectivamente.

Tabla B.1: Resultados usando imágenes perioculares de la base de datos GFI para las pruebas VGG+LSVM y VGG-full. La diferencia *versus* iris normalizado se calculó en base a los mejores resultados promedio obtenidos en esa prueba. En negrita se indica el mayor *accuracy*, la mayor diferencia *versus* iris normalizado, y los p-values menores a 0.05.

Tipo de prueba	Resolución original (escalada)	Accuracy [%]	Diferencia vs iris normalizado	P-value
VGG+LSVM	480×640 (224×224)	82.33 ± 1.62	+16.90	7.18×10^{-42}
VGG-full	480×640 (224×224)	84.66 ± 1.92	+21.32	2.99×10^{-41}