



**UNIVERSIDAD CATÓLICA DEL NORTE  
FACULTAD DE INGENIERÍA Y CIENCIAS GEOLÓGICAS  
DEPARTAMENTO DE INGENIERÍA DE SISTEMAS Y  
COMPUTACIÓN**

**CLASIFICACIÓN DEMOGRÁFICA POR MEDIO DEL ANÁLISIS  
DE ATRIBUTOS FACIALES SOBRE IMÁGENES EN  
ENTORNOS SIN RESTRICCIONES**

Tesis para optar al grado de Magíster en Ingeniería Informática.

**EDUARDO ANDRÉS AGUILAR TORRES**

Profesor Guía: Doctor (c) Juan Demóstenes Bekios Calfa

Antofagasta, Chile.

2015.

A mi señora, por su apoyo y amor incondicional, y a nuestro hijo que viene en camino.

## AGRADECIMIENTOS

En primer lugar, quiero agradecer a mi familia, en especial a mi señora y mi madre, quienes me dieron su apoyo incondicional durante todo el proceso que conlleva el desarrollo del trabajo de tesis, comprendiendo la poca disponibilidad de tiempo y soportando el mal estado de ánimo que tuve en algunos momentos.

También, agradezco a los académicos de la Universidad Católica del Norte, quienes han sido un pilar fundamental en mi formación como profesional. En particular, doy las gracias a mi profesor guía, Juan Bekios, por la motivación que me ha dado al transmitirme sus experiencias relacionadas con el campo de visión artificial, y por su buena disposición para atender las consultas que surgieron en el transcurso de las investigaciones.

Además, doy las gracias al comité de Magíster en Ingeniería Informática, por haberme otorgado una beca que consistió en la liberación del pago del 100% del arancel de matrícula, y de este modo hicieron posible mi ingreso al programa de magíster.

Finalmente, quiero agradecer a la Pontificia Universidad Católica de Chile por haber facilitado el acceso a sus *cluster*, permitiendo con esto agilizar la obtención de resultados de los diversos experimentos realizados.

# ÍNDICE

<b>ÍNDICE DE TABLAS .....</b>	<b>VI</b>
<b>ÍNDICE DE FIGURAS .....</b>	<b>VIII</b>
<b>RESUMEN.....</b>	<b>IX</b>
<b>I. INTRODUCCIÓN .....</b>	<b>10</b>
1.1 PLANTEAMIENTO DEL PROBLEMA .....	11
1.2 RELEVANCIA DEL PROBLEMA .....	13
1.3 PREGUNTAS DE INVESTIGACIÓN .....	14
1.4 OBJETIVOS .....	15
1.4.1 <i>Objetivo General</i> .....	15
1.4.2 <i>Objetivos Específicos</i> .....	15
<b>II. MARCO TEÓRICO .....</b>	<b>16</b>
2.1 RECONOCIMIENTO DE PATRONES.....	16
2.1.1 <i>Adquisición de datos</i> .....	18
2.1.2 <i>Pre-procesamiento</i> .....	18
2.1.3 <i>Extracción de características</i> .....	18
2.1.4 <i>Clasificación</i> .....	19
2.2 ALGORITMOS DE EXTRACCIÓN DE CARACTERÍSTICAS .....	21
2.2.1 <i>Descriptores Globales</i> .....	21
2.2.2 <i>Descriptores Locales</i> .....	23
2.3 CLASIFICADOR LINEAL.....	29
2.3.1 <i>LDA</i> .....	29
2.3.2 <i>PCA+LDA</i> .....	31
2.4 CONJUNTO DE DATOS .....	32
2.4.1 <i>Condiciones Ideales</i> .....	32
2.4.2 <i>Condiciones Reales</i> .....	38
2.5 CLASIFICACIÓN DE GÉNERO .....	40
2.5.1 <i>Condiciones Ideales</i> .....	42
2.5.2 <i>Condiciones Reales</i> .....	48

2.6	HIPÓTESIS DE INVESTIGACIÓN .....	52
<b>III.</b>	<b>METODOLOGÍA .....</b>	<b>53</b>
3.1	MÉTODO DE INVESTIGACIÓN Y TIPO DE DISEÑO.....	53
3.2	UNIDAD DE ANÁLISIS Y EL MÉTODO DE MUESTREO.....	54
3.3	VALIDEZ Y CONFIABILIDAD DE LOS INSTRUMENTOS DE MEDICIÓN .....	54
3.4	PROCEDIMIENTOS.....	55
3.4.1	<i>Selección de la muestra</i> .....	56
3.4.2	<i>Preparación de los datos</i> .....	57
3.4.3	<i>Extracción de características</i> .....	60
3.4.4	<i>Clasificación</i> .....	64
3.4.5	<i>Análisis de los datos</i> .....	66
<b>IV.</b>	<b>RESULTADOS .....</b>	<b>67</b>
4.1	ANÁLISIS DEL PODER DISCRIMINANTE DE LOS MODELOS PROPUESTOS .....	67
4.2	CORROBORAR EL PODER DISCRIMINANTE MINIMIZANDO EL SESGO .....	71
4.3	EVALUAR EL PODER DISCRIMINANTE SOBRE CARAS EN ENTORNOS SIN CONTROLAR .....	75
4.4	PUBLICACIONES.....	82
<b>V.</b>	<b>CONCLUSIONES.....</b>	<b>83</b>
5.1	TRABAJO FUTURO.....	86
	<b>BIBLIOGRAFÍA .....</b>	<b>88</b>

## ÍNDICE DE TABLAS

TABLA II.1 RENDIMIENTO OBTENIDO EN LA CLASIFICACIÓN DE GÉNERO .....	41
TABLA III.1. FRECUENCIA DE IMÁGENES DE CARAS DE HOMBRES Y MUJERES DISTRIBUIDAS POR POSE. ....	58
TABLA III.2 DISTRIBUCIÓN DE IMÁGENES POR GÉNERO EN CADA POSE. ....	59
TABLA IV.1 TASA DE ACIERTO GLOBAL OBTENIDA EN LA DETECCIÓN DEL GÉNERO PARA CADA MODELO PROPUESTO. ....	68
TABLA IV.2 TASA DE ACIERTO DISTRIBUIDA POR POSE Y MODELO DE CLASIFICACIÓN. ....	70
TABLA IV.3 TASA DE ACIERTO DISTRIBUIDA POR POSE, GÉNERO Y MODELO DE CLASIFICACIÓN. ....	71
TABLA IV.4 TASA DE ACIERTO GLOBAL OBTENIDA POR CADA MODELO DE CLASIFICACIÓN UTILIZANDO LA ESTRATEGIA LOPO. ....	72
TABLA IV.5 TASA DE ACIERTO DISTRIBUIDA POR POSE Y MODELO DE CLASIFICACIÓN (LOPO). ....	73
TABLA IV.6 TASA DE ACIERTO DISTRIBUIDA POR POSE, GÉNERO Y MODELO DE CLASIFICACIÓN. (LOPO) .....	74
TABLA IV.7 TASA DE ACIERTO USANDO LAS ESTRATEGIAS <i>CROSS-VALIDATION</i> Y LOPO .....	74
TABLA IV.8 DISTRIBUCIÓN DE IMÁGENES POR GÉNERO EN CADA POSE DEL CONJUNTO DE PRUEBA. ....	75
TABLA IV.9 TASA DE ACIERTO OBTENIDA CON MODELOS BASADOS EN CARACTERÍSTICAS RESPECTO A LA CANTIDAD DE CLÚSTER UTILIZADOS. ....	76
TABLA IV.10 TASA DE ACIERTO OBTENIDA CON EL MODELO DE CLASIFICACIÓN BASADO EN SIFT EN CADA <i>FOLD</i> DE LA VALIDACIÓN CRUZADA. ....	77
TABLA IV.11 TASA DE ACIERTO OBTENIDA CON EL MODELO DE CLASIFICACIÓN BASADO EN SIFT EN CADA <i>FOLD</i> DE LA VALIDACIÓN CRUZADA. ....	77
TABLA IV.12 TASA DE ACIERTO OBTENIDA CON MODELOS BASADOS EN CARACTERÍSTICAS RESPECTO A LA CANTIDAD DE CLÚSTER UTILIZADOS. ....	78
TABLA IV.13 TASA DE ACIERTO OBTENIDA CON EL MODELO DE CLASIFICACIÓN BASADO EN SIFT EN CADA <i>FOLD</i> DE LA VALIDACIÓN CRUZADA. ....	78
TABLA IV.14 TASA DE ACIERTO OBTENIDA CON EL MODELO DE CLASIFICACIÓN BASADO EN SURF EN CADA <i>FOLD</i> DE LA VALIDACIÓN CRUZADA. ....	79
TABLA IV.15 TASA DE ACIERTO OBTENIDA CON EL MODELO DE CLASIFICACIÓN BASADO EN APARIENCIA EN CADA <i>FOLD</i> DE LA VALIDACIÓN CRUZADA. ....	79
TABLA IV.16 TASA DE ACIERTO OBTENIDA CON EL MODELO DE CLASIFICACIÓN BASADO EN APARIENCIA EN CADA <i>FOLD</i> DE LA VALIDACIÓN CRUZADA AL UTILIZAR ATRIBUTOS <i>POWERSET</i> . ....	80
TABLA IV.17 TASA DE ACIERTO GLOBAL DE LOS MODELOS DE CLASIFICACIÓN EVALUADOS ESTRATIFICADOS POR POSE. ....	80
TABLA IV.18 TASA DE ACIERTO GLOBAL DE LOS MODELOS DE CLASIFICACIÓN EVALUADOS ESTRATIFICADOS POR POSE UTILIZANDO ATRIBUTO <i>POWERSET</i> . ....	80

TABLA IV.19 TASA DE ACIERTO GLOBAL CON LOS MODELOS DE CLASIFICACIÓN EVALUADOS AL CLASIFICAR GÉNERO DE LA MANERA TRADICIONAL Y AL UTILIZAR ATRIBUTOS <i>POWERSET</i> .....	81
---	----

## ÍNDICE DE FIGURAS

FIGURA II.1 SISTEMA DE RECONOCIMIENTO DE PATRONES. (AKSOY, 2012) .....	17
FIGURA II.2 GRÁFICO QUE DESPLIEGA LA DISTRIBUCIÓN DE LAS CLASES “+” Y “0”, Y LA RECTA QUE SEPARA ESTAS CLASES. (THEODORIDIS ET AL., 2008) .....	20
FIGURA II.3 ILUSTRACIÓN DEL DESCRIPTOR SIFT.....	25
FIGURA II.4 CONJUNTO DE IMÁGENES OBTENIDAS DE LA BASE DE DATOS FERET.....	33
FIGURA II.5 CONJUNTO DE IMÁGENES DE LA BASE DE DATOS PAL.....	34
FIGURA II.6 CONJUNTO DE IMÁGENES DE LA BASE DE DATOS MORPH.....	35
FIGURA II.7 ETIQUETAS Y DISTRIBUCIÓN APROXIMADA DE LAS CÁMARAS DENTRO DE LA SALA.....	36
FIGURA II.8 CONJUNTO DE IMÁGENES MULTI-PIE.....	37
FIGURA II.9 CONJUNTO DE IMÁGENES DE LA BASE DE DATOS LFW. ....	38
FIGURA II.10 CONJUNTO DE IMÁGENES DE LA BASE DE DATOS GROUPS.....	39
FIGURA III.1 PROCESO UTILIZADO PARA LA ESTIMACIÓN DE GÉNERO. ....	56
FIGURA III.2 DETECCIÓN DE ROSTRO POR MEDIO DE VIOLA & JONES. ....	58
FIGURA III.3. MODELO DE CLASIFICACIÓN PROPUESTO BASADO EN LA CONSTRUCCIÓN DE VECTOR DE FRECUENCIAS. ....	61
FIGURA III.4. PUNTOS DE INTERÉS DETECTADOS CON SIFT SOBRE IMÁGENES DE UN MISMO SUJETO CON VARIACIÓN DE POSE. ....	62
FIGURA III.5. PUNTOS DE INTERÉS DETECTADOS CON SURF SOBRE IMÁGENES DE UN MISMO SUJETO CON VARIACIÓN DE POSE. ....	63
FIGURA III.6. PUNTOS DE INTERÉS DETECTADOS CON ORB SOBRE IMÁGENES DE UN MISMO SUJETO CON VARIACIÓN DE POSE. ....	63
FIGURA III.7. REPRESENTACIÓN VISUAL DE LA ESTRATEGIA VALIDACIÓN CRUZADA 5 FOLD. ....	64
FIGURA IV.1 GRÁFICO DE BARRAS DEL PORCENTAJE DE TASA DE ACIERTO EN LA DETECCIÓN DEL GÉNERO PARA CADA MODELO PROPUESTO. ....	68
FIGURA IV.2 GRÁFICO DE DISPERSIÓN DE TASA DE ACIERTO VS POSE. ....	69

## RESUMEN

La clasificación demográfica, y en particular el reconocimiento de género, es un tema muy estudiado por investigadores debido a su importancia en diversas aplicaciones, tales como, marketing dirigido, sistemas de seguridad, entre otras. El trabajo de tesis realizado contempla la construcción y evaluación de tres procedimientos para llevar a cabo la clasificación, explorando modelos basados en apariencia y basados en característica, además de aprovechar la información contenida en las distintas clases asociadas a cada instancia del conjunto de entrenamiento. Para el primer procedimiento se realiza la clasificación de género usando vectores de frecuencia basados en descriptores locales SIFT o SURF. En este procedimiento se utilizó para el entrenamiento y la validación la estrategia *cross-validation* sobre la base de datos Multi-PIE. Para el segundo, basado en el anterior, se incorpora el descriptor local ORB, y cambia la estrategia de entrenamiento y validación a *Leave One Person Out* (LOPO). Finalmente, el tercero incorpora la base de datos GROUPS, y utiliza un enfoque de múltiples clases en el entrenamiento del clasificador. Los resultados obtenidos evidencian que los modelos basados en características propuestos poseen suficiente información discriminante y tienen un rendimiento similar a los modelos basado en apariencia. Además mantienen un rendimiento uniforme en la clasificación de género sobre imágenes con variaciones de pose. En conclusión, no se obtienen mejoras significativas en cuanto a la tasa de acierto evidenciada en la literatura, sin embargo se provee de un procedimiento base que podría mejorar la clasificación al incorporar una fase de filtrado previo a la construcción de los vectores de características.

**Palabras clave:** Reconocimiento de género, algoritmos basados en apariencia, algoritmos basados en características, clasificador lineal, descriptor local.

## I. INTRODUCCIÓN

La clasificación demográfica automática de rostros corresponde al proceso que involucra identificar, y por consiguiente, asignar características a individuos detectados en una imagen, tales como: el género, la pose, la edad, la etnia, la expresión, si tiene gafas, entre otras. Un sistema automático de clasificación demográfica debe al menos ser capaz de reconocer el género (hombre, mujer), categorizar la edad (niño, joven, adulto, de mediana edad, anciano) y clasificar la etnia (Asiático, Caucásico y Africano) al analizar el sujeto detectado (Hadid & Pietikäinen, 2013). Comúnmente estas tareas se realizan individualmente por cada clase y no se considera la información implícita de estos atributos a través de una aproximación más general. En este caso, la tesis se centra en el reconocimiento de género, el cual es uno de los aspectos más relevantes que involucra la clasificación demográfica, y será llevado a cabo por medio de un modelo de clasificación basado en descriptores locales, aprovechando de este modo la robustez que ofrecen para las diversas condiciones que pudiesen estar presentes en las imágenes adquiridas en entornos sin restricciones.

El reconocimiento de género por medio del análisis de atributos faciales ha sido un tema estudiado por investigadores desde hace más de 20 años, siendo SEXNET el primer algoritmo desarrollado (Golomb et al., 1990). En la literatura, el análisis de atributos faciales se ha realizado mediante enfoques basados en apariencia global y enfoques basados en características locales. Los enfoques basados en apariencia usan el recorte, el redimensionamiento y la iluminación normalizada de la textura de toda la cara como un atributo de clasificación (Mäkinen & Raisamo, 2008) (Bekios-Calfa et al., 2011). Por otro lado, enfoques basados en características conllevan la extracción de un

conjunto de zonas discriminativas de la cara (M.-H. Yang et al., 2002) (Bekios-Calfa et al., 2011).

Para llevar a cabo la clasificación del género, se debe construir un modelo de clasificación por medio de las características obtenidas del rostro. Se opta por construir este modelo mediante el clasificador lineal PCA (*Principal Component Analysis*) + LDA (*Linear Discriminant Analysis*) debido a que se ha demostrado en la literatura que tiene un bajo costo computacional, eficiente en tiempo de respuesta y un buen rendimiento en términos de tasa de acierto, similares respecto a otros algoritmos de clasificación.

A continuación se describe el planteamiento del problema, la relevancia del problema, la hipótesis de investigación, el objetivo general y los objetivos específicos.

## **1.1 Planteamiento del problema**

A pesar de los buenos resultados obtenidos en los últimos años, el problema de la clasificación de género no se encuentra del todo resuelto. Altas tasas de acierto se han obtenidos sobre imágenes de caras tomadas en ambientes controlados, sin embargo, aún no se ha podido replicar este buen funcionamiento sobre imágenes capturadas en entornos sin restricción. Un algoritmo con alta tasa de acierto sobre entornos controlados no implica que mantendrá su buen desempeño en entornos de la vida real. Esto se puede apreciar fácilmente al analizar los resultados experimentales obtenidos en la publicación de Bekios-Calfa et al. (2014), en la cual se observa que se produce una baja de la tasa de acierto en la clasificación, utilizando la técnica de discriminante lineal, al identificar el género sobre imágenes en entornos sin restricción. En Bekios-Calfa et al. (2011), se observa al entrenar un clasificador

de género con la base de datos FERET una tasa de acierto de un 93.33%, está disminuye bruscamente al utilizar la base de datos GROUPS logrando una tasa de acierto de un 80.53% (Bekios-Calfa et al., 2014). Las diferencias principales entre base de datos de laboratorio (FERET) y base de datos con imágenes en entornos sin restricción (GROUPS), son que las base de datos de laboratorio se caracterizan por tener imágenes de rostro con buena iluminación, correcta postura, baja variación de gestos y/o fondo homogéneo. En entornos sin restricciones difícilmente se van a cumplir estas condiciones, ya que al tomar la imagen en tiempo real, de una o más personas, éstas pueden estar: en movimiento, demostrando sus emociones mediante gestos en el rostro, sujetas a oclusiones, fondos no homogéneos, cambios de iluminación y/o pose.

Por otro lado, a partir de la creación de las bases de datos LFW (Huang et al., 2007), y 2 años después GROUPS (Gallagher & Chen, 2009), se comenzaron a realizar más investigaciones para el reconocimiento de género sobre imágenes de caras en entornos sin restricción, debido a que son bases de datos públicas y por ende es posible realizar comparaciones del rendimiento obtenido de los distintos modelos de clasificación. Éstas investigaciones presentan variadas técnicas, tales como aplicar enfoques de clasificación utilizando múltiples variables clases como el género, edad y postura (Bekios-Calfa et al., 2014), o con el uso de descriptores locales, LBP y Gabor jets, previo a la clasificación (Dago-Casas et al., 2011). Si bien se han realizado investigaciones, el buen desempeño que se ha logrado para conjuntos de datos capturados en ambientes controlados, sugiere que aún queda bastante trabajo por hacer para mejorar la robustez del reconocimiento de género en entornos sin restricción. (Ng et al., 2012).

Finalmente, cabe destacar que la mayoría de las investigaciones han utilizado por separado los enfoques globales y locales (Mirza et al., 2013). Sin embargo, la percepción del humano para reconocer y categorizar un género involucra la observación del rostro tanto en zonas específicas como global (Mirza et al., 2013).

Considerando lo descrito previamente, el desafío actual se centra en realizar una mayor investigación para la clasificación de género en condiciones reales, con la finalidad de obtener un método de clasificación que involucra aspectos de pre-procesamiento de imagen, extracción de características locales y/o globales, y el uso de la información a priori de múltiples variables clases.

## **1.2 Relevancia del Problema**

En general, un sistema exitoso de reconocimiento de género tiene un gran impacto en un sin número de aplicaciones tales como: marketing dirigido, autenticación biométrica, sistemas de seguridad, interacción persona-computador, control de material virtual, entre otras. (Ng et al., 2012)

En términos de seguridad, el reconocimiento de género permite acelerar la recuperación de imágenes y su posterior indexación sobre un sujeto específico. Debido a que podemos filtrar las imágenes por atributos faciales (género, edad, raza, etc.) en la base de datos, disminuyendo el tamaño de la muestra para la búsqueda requerida. (NEC, 2014)

El control parental es otro aspecto útil, ya que si se detecta el género y a su vez la edad de la persona que está al frente del computador o televisor, en el caso que sea menor de edad, se puede tomar una serie de acciones tales

como evitar contenido erótico y de violencia. Por otro lado se pueden mejorar las prestaciones de búsqueda en Internet, mostrar comerciales, programas de televisión y películas adaptadas a las características de cada persona.

Marketing dirigido es un campo que se podría explotar de mejor manera con un clasificador de género, en la que por medio de la detección del género se podría desplegar publicidad orientada al público que está observando, en este caso diferenciar entre un hombre o una mujer. En Japón ya se está usando esta tecnología, aplicada a máquinas de bebidas calientes y heladas, en las que se recomienda que producto comprar dependiendo del género y edad de la persona. (AX3, 2013)

Por último, otro tipo de aplicación es la interacción de la persona con el computador, en la que el computador podría cambiar de diseño y emitir una serie de eventos dependiendo de la persona que tenga en frente.

### **1.3 Preguntas de Investigación**

- ¿Son suficientemente discriminativos los vectores de características contruidos a partir de los descriptores locales SURF, SIFT y ORB, para el diseño de un clasificador de género (hombre, mujer) aplicado sobre imágenes adquiridas en entornos sin restricciones?
- ¿Los algoritmos de clasificación de género basados en características permiten mejorar la tasa de acierto en la clasificación cuando se considera la información contenida en las clases (género, pose, edad)?

## **1.4 Objetivos**

En esta sección se enuncia el objetivo general y los objetivos específicos derivados para su concreción.

### **1.4.1 Objetivo General**

Generar un método para la clasificación demográfica, específicamente sobre el género, por medio del análisis de atributos faciales sobre imágenes en entornos sin restricciones.

### **1.4.2 Objetivos Específicos**

- Analizar los métodos de clasificación demográfica en cuanto a las estrategias utilizadas y a las tasa de acierto obtenidas.
- Evaluar el uso de descriptores locales y globales para la clasificación de género.
- Evaluar el uso de descriptores locales y globales utilizando la información contenida en las clases, para la clasificación de género.
- Determinar un método para preparar la imagen antes de la extracción de características.
- Determinar un método de extracción de características y algoritmo de clasificación.
- Diseñar un procedimiento para clasificación de género sobre imágenes adquiridas en entornos sin restricciones.
- Contrastar resultados obtenidos con estudios previos.
- Difundir los resultados obtenidos a través de congresos y publicación en revistas indexadas.

## **II. MARCO TEÓRICO**

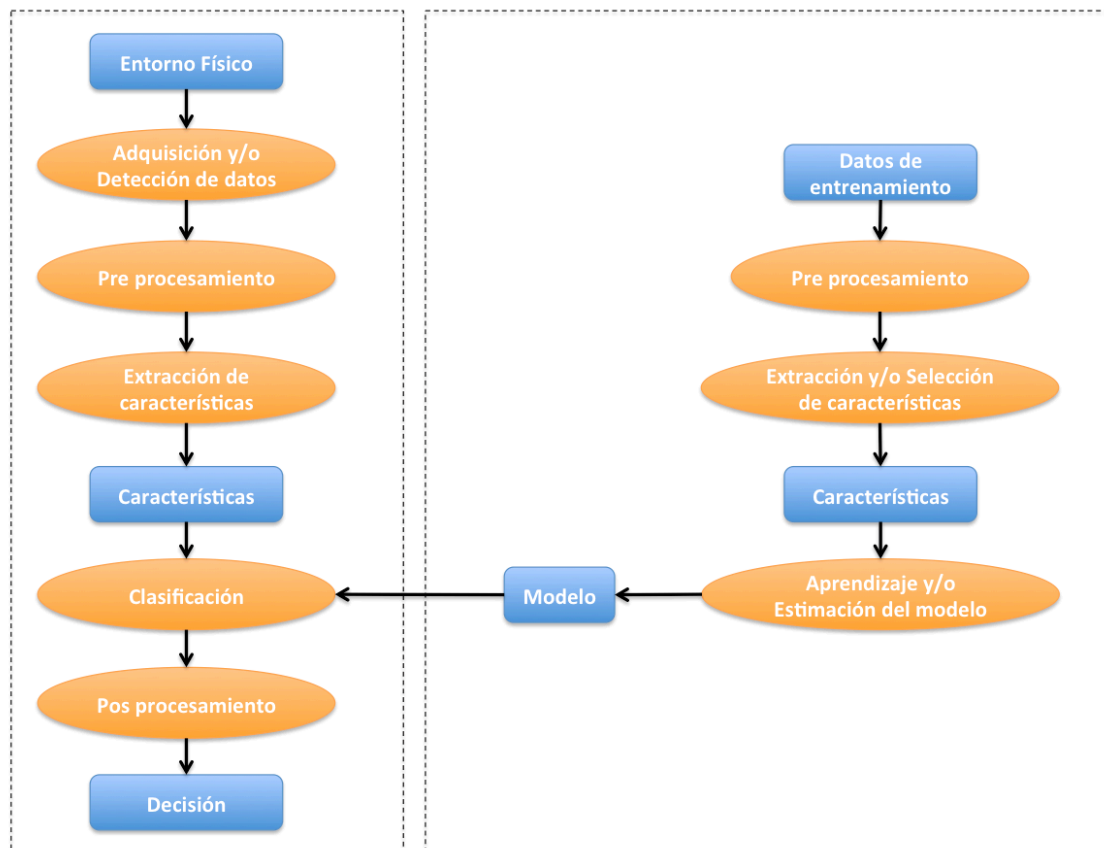
El sustento teórico de la investigación se muestra desde lo general a lo particular. En primera instancia, se da a conocer los elementos básicos del reconocimiento de patrones, especificando los conceptos teóricos que lo conforman. Luego se detallan diversos algoritmos para la extracción de característica, algoritmos para la clasificación lineal, y distintos conjuntos de bases de datos disponibles para los investigadores. Finalmente se hace una revisión de las publicaciones relacionada con la clasificación demográfica.

### **2.1 Reconocimiento de Patrones**

El reconocimiento de patrones es una disciplina científica que tiene como principal objetivo clasificar objetos en un número de categorías o clases. Específicamente, para la clasificación demográfica por medio de atributos faciales, los objetos a clasificar son las imágenes de caras de humanos, refiriéndose a ellos con el término genérico “patrón”.

La identificación de huellas dactilares, la autenticación de firma, la recuperación de texto, y el reconocimiento de gestos y cara, son algunas de las aplicaciones típicas relacionadas con el reconocimiento de patrones. Sobre todo las últimas aplicaciones nombradas han traído bastante interés en la investigación, dado que facilitan la interacción persona-computador, potencian los dispositivos de vigilancia, permiten realizar marketing dirigido y mejoran el rol de la informática en la ofimática, la cual consiste en la personalización automática de entornos. (Theodoridis & Koutroumbas, 2008)

Un sistema de reconocimiento de patrones es mostrado en la Figura II.1 (Aksoy, 2012). En la parte izquierda se presenta el proceso desde la adquisición de los datos en el entorno hasta la decisión; y en la parte derecha el proceso de entrenamiento con los datos *a priori* hasta la generación del modelo que se utiliza para la clasificación. En las próximas secciones se explican los procesos que se consideran relevantes para realizar la investigación, tales como: la adquisición de datos, el pre-procesamiento, la extracción de características y la clasificación.



**Figura II.1** Sistema de reconocimiento de patrones. (Aksoy, 2012)

### **2.1.1 Adquisición de datos**

La adquisición de datos involucra el proceso necesario para recoger las observaciones a clasificar. Los datos pueden ser adquiridos mediante sensores o muestreo en el caso de clasificar datos estadísticos. Específicamente para el reconocimiento de imágenes de caras de humanos, el sensor a utilizar es una cámara de video, cámara fotográfica, webcam, entre otros.

### **2.1.2 Pre-procesamiento**

Luego de adquirir los datos, se pueden utilizar técnicas de pre-procesamiento para disminuir la sensibilidad del clasificador. Entre las técnicas de pre-procesamiento aplicadas a imágenes de caras se encuentran: la ecualización de histograma, para remover el ruido y/o para suavizar los cambios de iluminación (Mäkinen et al., 2008)(Bekios-Calfa et al., 2011), recortar las imágenes para remover características externas a la cara (Baluja & Rowley, 2006)(Bekios-Calfa et al., 2011), alinear las caras utilizando transformaciones geométricas (Moghaddam & Yang, 2002)(Dago-Casas et al., 2011), entre otras.

### **2.1.3 Extracción de características**

La extracción de característica es el proceso de transformar la información observada (adquirida), en valores numéricos y/o simbólicos para ser utilizados en el proceso de clasificación. Para ello es necesario identificar las medidas adecuada, que permita separar y distinguir cada una de las

clases, por ejemplo, el valor medio y desviación estándar de las intensidades de cada región de interés.

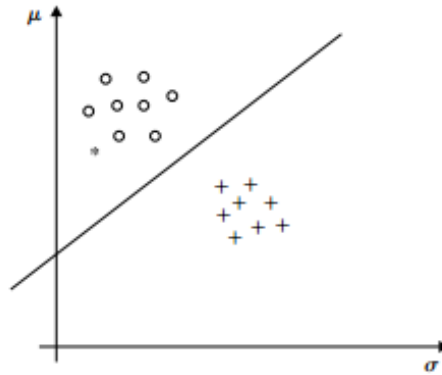
Las medidas utilizadas para la clasificación son conocidas como “características”. En general si para la clasificación se usan  $n$  características  $x_i$ , donde  $i = 1, 2, \dots, n$ , el vector de características se define formalmente:

$$x = [x_1, x_2, \dots, x_n]^T \quad (1)$$

Cada uno de los vectores de características identifica de forma exclusiva un único patrón (objeto). Debido a esto las características y los vectores de características serán tratados como variables y vectores aleatorios respectivamente. Eso es evidente, ya que las mediciones resultantes de diferentes patrones exhiben una variación aleatoria (Theodoridis et al., 2008). Esto se debe principalmente a las distintas características de cada patrón, por ejemplo en imágenes de caras de humanos, la dispersión de las clases está sujeta a una serie de factores tales como diferencia de rasgos de los individuos, condiciones en entorno en que es tomada la imagen, entre otras cosas.

#### **2.1.4 Clasificación**

La clasificación es el proceso que, basado en las características extraídas (vector de características), permite asignar una categoría o clase a un patrón. Por ejemplo, como se muestra en la Figura II.2, si se tienen dos clases “+” y “o”, y el vector de características contiene la desviación estándar y la media. Si las clases son distantes, es posible mediante una línea recta poder distinguir entre ellas. Si un vector de característica “\*” corresponde a un patrón desconocido, se clasifica con la clase correspondiente a la categoría de la región en que se encuentra. (Theodoridis et al., 2008).



**Figura II.2** Gráfico que despliega la distribución de las clases “+” y “o”, y la recta que separa estas clases. (Theodoridis et al., 2008)

A continuación se explican los tres tipos de clasificación que existen: clasificación supervisada, no supervisada y parcialmente supervisada.

#### **2.1.4.1 Clasificación supervisada**

El aprendizaje de clasificación se denomina supervisado, cuando el diseño del clasificador se realiza con los datos de entrenamiento con su categoría *a priori* conocida. (Theodoridis et al., 2008)

#### **2.1.4.2 Clasificación no supervisada**

La clasificación se denomina no supervisada, cuando las categorías de los datos de entrenamiento no se conocen o no están disponibles. El objetivo es ir agrupando las clases por medio de las similitudes de los vectores de característica. (Theodoridis et al., 2008)

### **2.1.4.3 Clasificación parcialmente supervisada**

El aprendizaje se denomina parcialmente supervisado, cuando los datos de entrenamiento que se tienen son similares a la clasificación supervisada pero además considera patrones de entrenamiento cuya verdadera clase se desconoce. Este enfoque es bastante útil cuando el diseño del clasificador tiene acceso a una pequeña cantidad de datos etiquetados. En este caso el agrupamiento de los patrones desconocidos está restringido a las clases existentes. (Theodoridis et al., 2008)

## **2.2 Algoritmos de extracción de características**

Para la construcción del modelo de clasificación, un proceso fundamental es la extracción de característica, como se muestra al lado derecho de la Figura II.1. La extracción se puede realizar sobre las imágenes de caras con un enfoque holístico, considerando todos los píxeles de la imágenes, o con enfoque más específico, considerando zonas del rostro definidas (ojos, labios, nariz, etc.), o por sectores del rostro tratados independientemente. Para ambos enfoques existen varios algoritmos para la extracción de características denominados descriptores globales y descriptores locales.

### **2.2.1 Descriptores Globales**

Los descriptores globales utilizan la imagen completa para su descripción, permitiendo generalizar el objeto que se desea reconocer en un simple vector. Debido a esto, su uso en técnicas de clasificación es sencillo y a su vez el uso de recursos computacionales requerido es menor que en técnicas de clasificación con descriptores locales. Sin embargo, los

descriptores globales al utilizar la imagen como un todo para la clasificación, son sensibles a las oclusiones y/o a fondos no homogéneos. Ejemplos de estos algoritmos son: PCA, CCA, SOM e ICA. (Sun et al., 2002)(Buchala et al., 2005)(Bekios-Calfa et al., 2011)

A continuación se detalla el descriptor global PCA.

### 2.2.1.1 PCA

*Principal Components Analysis* (PCA), es una técnica que permite reducir las dimensiones del vector de característica conservando la mayor cantidad de información posible. Esto se logra mediante una proyección que maximiza la varianza y minimiza el error cuadrático medio de reconstrucción al mismo tiempo.

Para el reconocimiento de imágenes de caras de humanos se menciona en Martinez & Kak (2001) que dada una representación vectorial de  $t$ -dimensional, PCA puede ser utilizado para encontrar un sub-espacio cuyos vectores base corresponden a las direcciones de máxima varianza en el espacio original. Sea  $W$  la transformación lineal que mapea el espacio  $t$ -dimensional original en un sub-espacio de  $f$  dimensiones donde normalmente  $f \ll t$ . Los nuevos vectores de características  $y_i \in \mathbb{R}^f$  se definen por  $y_i = W^T x_i, i = 1, \dots, N$ . Las columnas de  $W$  son los valores propios  $e_i$  obtenidos mediante la resolución de la descomposición valor y vector propio  $\lambda_i e_i = Q e_i$ , donde  $Q = XX^T$  es la matriz de covarianza y  $\lambda_i$  el valor propio asociado al vector propio  $e_i$ . Antes de obtener los vectores propios de  $Q$ :

- Los vectores se normalizan de tal manera que  $\|x_i\| = 1$  para hacer que el sistema invariante a la intensidad de la fuente de iluminación.
- Se resta la media de todas las imágenes de todos los vectores normalizados para asegurar que el vector propio con el valor propio más alto representa la dimensión en el espacio característico en el que la varianza de los vectores es máxima sobre la correlación de estos.

### **2.2.2 Descriptores Locales**

Los descriptores locales se basan en la extracción de características de múltiples regiones de una imagen. En términos generales el proceso que se realiza es el análisis de vecindad de los píxeles de una región definida, con lo que se obtiene un conjunto de vectores que describen cada punto de interés de la imagen. Al describir cada punto de interés de la imagen, permiten tener un clasificador robusto a oclusiones y a diversos ruidos que pudieran estar presentes.

A continuación se detallan los descriptores locales que se utilizan en la etapa experimental. Estos son SIFT, SURF y ORB.

#### **2.2.2.1 SIFT**

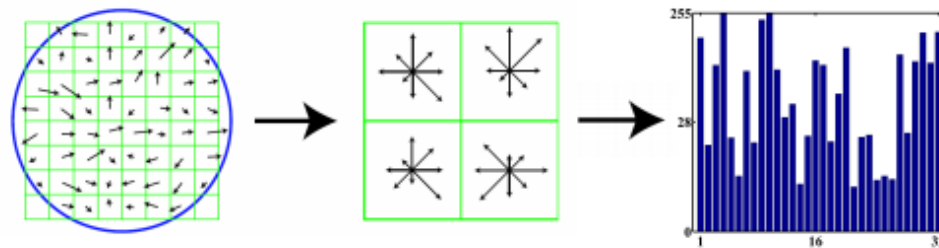
*Scale Invariant Feature Transform* (SIFT) es un detector y descriptor diseñado por Lowe (2004). El algoritmo permite transformar la imagen a una representación compuesta de puntos de interés invariantes a la escala de la imagen y a la rotación. Además provee de descriptores robustos para emparejamiento, admitiendo substanciales rangos de distorsión afín, cambios de perspectiva, adición de ruido y cambios de iluminación.

A continuación se muestran en términos generales los pasos que realiza el algoritmo para generar el conjunto de características asociado a la imagen:

- Detección extrema del espacio escalar: El primer paso consiste en buscar puntos de interés sobre todas las escalas y posiciones de la imagen representada en diferentes escalas. Esto se implementa eficientemente mediante el uso de una función de diferencia gaussiana para identificar potenciales puntos que son invariables a escala y orientación.
- Localización de puntos de interés: En cada punto detectado, un modelo detallado es entrenado para determinar aquellos puntos que se mantienen invariantes en cuanto a cambios de escala y localización. Para ello se estudia cada pixel y se realiza una comparación con los pixeles vecinos. Los puntos de interés son seleccionados en base a las medidas de su estabilidad.
- Asignación de orientación: Una o más orientaciones son asignadas para cada punto de interés de acuerdo a las direcciones del gradiente y a la zona que rodea dicho punto.
- Descriptor de puntos de interés: Consiste en calcular un descriptor para cada punto de interés de imagen local. Los gradientes de la imagen son medidos para seleccionar la escala y región alrededor del punto. Estos son transformados en una representación que permite ser invariable a significativos cambios de iluminación y perspectiva.

Para cada descriptor obtenido con SIFT, todos los gradientes ponderados se normalizan a la orientación principal de la región circular. La región circular alrededor del punto de interés es dividida en sub-regiones de  $4 \times 4$  pixeles sin superponerse y se calculan los histogramas de las orientaciones del gradiente

dentro de estas regiones. El suavizado del histograma se realiza con el fin de evitar cambios repentinos de la orientación, además el tamaño del contenedor se reduce a 8 con el fin de limitar la dimensión del descriptor. Esto se traduce en un vector de características de 128 dimensiones ( $4 \times 4 \times 8$ ) para cada punto de interés. En la Figura II.3 se ilustra el procedimiento descrito para una ventana de  $2 \times 2$  en vez de  $4 \times 4$ , por lo que el histograma obtenido tiene solo 32 dimensiones.



**Figura II.3** Ilustración del descriptor SIFT.

### 2.2.2.2 SURF

*Speeded Up Robust Features* (SURF) es un detector y descriptor desarrollado por Bay et al. (2006) como una versión acelerada de SIFT. Al igual que SIFT permite, a partir de una imagen, construir una representación compuesta de puntos de interés invariantes a la escala y rotación, pero no es tan bueno con imágenes que presentan cambios de perspectiva y cambios de iluminación.

El algoritmo de SURF puede ser descrito en 3 etapas:

- Detección de puntos de interés: El primer paso consiste en detectar puntos de interés por medio de un *box filter* que al igual que en el caso de SIFT con la diferencia gaussiana, aproxima al filtro *Laplacian*

of *Gaussian* para identificar potenciales puntos que son invariables a los cambios de escala. Debido al uso de *box filter* y las imágenes integrales, no se tiene que aplicar iterativamente el mismo filtro a la salida de una capa previamente filtrada, sino que puede aplicar tales filtros de cualquier tamaño, exactamente a la misma velocidad, directamente en la imagen original. Por lo tanto, el espacio escalar se analiza por el aumento del tamaño del filtro en lugar de reducir la imagen original.

- Asignación de la orientación: Para que los puntos de interés sean invariantes a la rotación, primero se calcula las respuestas de Haar-wavelet en las direcciones  $x$  e  $y$ , en un radio de  $6s$  alrededor del punto. El parámetro  $s$  es la escala con que el punto de interés fue detectado. Luego se calcula y pondera las respuestas wavelet con una función Gaussiana centrada en el punto de interés, la respuesta se representa como vector. La orientación dominante se determina mediante la suma de todas las respuestas dentro de una ventana deslizante que cubre la orientación de un ángulo de  $60^\circ$ . Se suman las respuestas horizontales y verticales dentro de la ventana, formando nuevos vectores, donde el más largo de estos se le extrae su orientación para el punto de interés.
- Extracción del descriptor: Para la extracción de los descriptores, el primer paso consiste en la construcción de una ventana centrada en el punto de interés, y orientada a lo largo de la dirección seleccionada por la suma de todas las respuestas horizontales y verticales dentro de la ventana. La región se divide regularmente en pequeñas sub-regiones de  $4 \times 4$ . Para cada sub-región las respuestas wavelet horizontales y verticales son tomadas para construir el vector de característica. Este vector queda representado con 64 dimensiones pero puede ser extendido a 128 dimensiones, esto último permite

obtener características más distintivas pero se disminuye la rapidez de la extracción.

### **2.2.2.3 ORB**

*Oriented FAST and Rotated BRIEF (ORB)* es un detector y descriptor, invariante a la rotación y resistente al ruido, desarrollado por Rublee et al. (2011). ORB surge como una alternativa eficiente a SIFT y SURF ya que dentro de sus principales características se destaca su bajo costo de computación, mantiene el buen rendimiento y, a diferencia de SIFT y SURF, no se debe pagar por su uso.

ORB se compone por la fusión de un detector de puntos de interés y un descriptor, denominados FAST (Rosten & Drummond, 2006) y BRIEF (Calonder et al., 2010) respectivamente. A estos algoritmos se les incorporan una serie de modificaciones para mejorar el rendimiento y lograr un algoritmo invariante a la rotación y robusto a los cambio de escala.

FAST se utiliza para encontrar los puntos de interés, de los cuales por medio de la medida de esquina de Harris se seleccionan los  $N$  mejores. Además, debido a que FAST no produce características multi-escala, se emplea una pirámide de escala de la imagen para generar característica para cada nivel de la pirámide. Otra carencia del detector FAST es que no computa la orientación, sin embargo esta es calculada mediante la técnica del centroide. Se utiliza el centroide de la intensidad como medida de la orientación, el cual asume que la intensidad de la esquina es desplazada desde su centro, y este vector puede ser usado para obtener la orientación.

En (Rosin, 1999) se define el momento de un *patch* como:

$$m_{pq} = \sum_{x,y} x^p y^q I(x,y), \quad (2)$$

donde  $I(x,y)$  es la función para obtener la intensidad del pixel en la posición  $x,y$  de la imagen  $I$ . Con estos momentos se puede buscar el centroide:

$$C = \left( \frac{m_{10}}{m_{00}}, \frac{m_{01}}{m_{00}} \right), \quad (3)$$

entonces se puede construir un vector desde el centro de la esquina al centroide. La orientación de este *patch* queda definido como:

$$\theta = \text{atan2}(m_{01}, m_{10}), \quad (4)$$

donde  $\text{atan2}$  es la versión del arco tangente que permite devolver el cuadrante correspondiente del ángulo calculado.

Respecto al descriptor, BRIEF tiene como limitante que no es invariante a la rotación. Por lo tanto para incorporar la invariancia, ORB utiliza una versión dirigida de BRIEF de acuerdo a la orientación de los puntos de interés. Por cada conjunto de característica de  $n$  pruebas binarias en la localización  $(x_i, y_i)$ , se define la matriz  $S$  de  $2 \times n$ , que contiene las coordenadas de estos pixeles.

$$S = \begin{pmatrix} x_1, \dots, x_n \\ y_1, \dots, y_n \end{pmatrix} \quad (5)$$

Entonces se utiliza la orientación del *patch*  $\theta$  y la matriz de rotación correspondiente  $R_\theta$  para construir la versión dirigida  $S_\theta$  de  $S$ .

$$S_\theta = R_\theta S, \quad (6)$$

Finalmente, el descriptor BRIEF dirigido se define como:

$$g_n(p, \theta) := f_n(p) | (x_i, y_i) \in S_\theta, \quad (7)$$

donde  $f_n(p)$  corresponde al descriptor BRIEF.

## 2.3 Clasificador Lineal

A continuación se describen dos métodos de clasificación lineal denominados LDA y PCA+LDA.

### 2.3.1 LDA

*Linear Discriminant Analysis* (LDA), aunque genéricamente se refiere a las técnicas que producen funciones discriminantes que son lineales en las variables de entrada, también se utiliza en un sentido específico a referirse a la técnica FDA (*Fisher Discriminant Analysis*) en la que se busca una transformación que, en cierto sentido, maximiza la separabilidad entre clase ( $S_B$ ) y reduce al mínimo la variabilidad dentro de la clase ( $S_W$ ).

Según Webb (2002), el algoritmo tiene las siguientes características:

- Una transformación se produce a un espacio de dimensión a lo más  $C - 1$ , donde  $C$  es el número de clases.
- La transformación es de libre distribución, en la que no se hace ninguna hipótesis respecto a la normalidad de los datos.
- Los ejes del sistema de coordenadas transformado puede ser ordenado en términos de "importancia para la discriminación". Los más importantes pueden ser utilizados para obtener una representación gráfica de los datos mediante el trazado de los datos en este sistema de coordenadas.
- La discriminación puede ser realizada en este espacio reducido dimensiones utilizando cualquier clasificador conveniente.

- Análisis discriminante lineal se puede utilizar como un post-procesador para clasificadores no lineales más complejos.

Además, se menciona que existen varias maneras de generalizar el criterio de Fisher ( $J_F$ ) para el caso múltiples clases. La optimización de estos criterios da transformaciones que reducen al discriminante lineal de Fisher en el caso de dos clases y que, en cierto sentido, maximizar la dispersión entre clase y reducir al mínimo la dispersión dentro de la clase. Se presenta un enfoque como sigue:

Considerando el criterio

$$J_F(a) = \frac{a^t S_B a}{a^t S_W a} \quad (8)$$

donde las estimaciones basadas en muestras de  $S_B$  y  $S_W$  están dadas por

$$S_B = \sum_{i=1}^C \frac{n_i}{n} (m_i - m)(m_i - m)^T \quad (9)$$

y

$$S_W = \sum_{i=1}^C \frac{n_i}{n} \Sigma_i \quad (10)$$

donde  $m_i$  y  $\Sigma_i$ ,  $i = 1, \dots, C$ , son las medias de las muestras y matrices de covarianza de cada clase (con muestras de  $n_i$ ) y  $m$  es la media de la muestra. Esto conduce a la ecuación de vector propio simétrica generalizada que maximizan el vector de características  $a_i$  está dada por:

$$S_B A = S_W A D \quad (11)$$

donde  $A$  es la matriz cuyas columnas son el  $a_i$  y  $D$  es la matriz diagonal de valores propios.

### 2.3.2 PCA+LDA

Según Bekios-Calfa et al. (2011), PCA+LDA permite solucionar los problemas de clasificación lineal de imágenes en donde es común tener un cantidad menor de muestra (imágenes) que de vectores de características. En estos casos, la matriz de dispersión dentro de la clase  $S_W$  es singular y por ende la proyección de la matriz  $A$  de LDA no puede ser calculada dado a que la matriz de covarianza del conjunto completo de la muestra es  $S_m = S_B + S_W$ . Una solución alternativa está dada por el uso de  $S_m$  en lugar de  $S_W$  en la ecuación de valor propio anterior (Fukunaga, 1990). En este caso, la realización de análisis de componentes principales (PCA) reteniendo sólo los vectores propios asociados a los valores propios distintos de cero y luego realizar LDA en el sub-espacio transformado PCA, es equivalente a la realización de LDA en el sub-espacio original de J. Yang & Yang (2003). Por lo tanto, cuando se descartan los vectores propios asociados a los valores propios distintos de cero, PCA + LDA no es estrictamente equivalente a la contrapartida global de LDA.

Cuando se tienen una mayor cantidad de muestra que características, el clasificador lineal LDA logra un rendimiento similar a PCA+LDA, en caso contrario PCA+LDA logra una mejor tasa de acierto. Esto se debe a la componente PCA que hace posible seleccionar solo los auto-valores que tengan una mayor variabilidad y descartar el resto de tal manera que se contengan los componentes con mayor información discriminante.

## **2.4 Conjunto de Datos**

El conjunto de datos es de vital importancia para poder construir el modelo de clasificación, por medio del entrenamiento y validación del clasificador, y por consiguiente realizar el proceso de clasificación demográfica. Para el reconocimiento de género se pueden distinguir dos grupos de bases de datos: en condiciones ideales, también denominadas en condiciones controladas o semi-controladas; y bases en condiciones reales, también conocidas como sin restricción, no controladas o *in the wild*.

### **2.4.1 Condiciones Ideales**

Las imágenes contenidas en las bases de datos en condiciones ideales son tomadas con leves cambios de iluminación, baja variabilidad de gestos, buen contraste y sin presencia de oclusiones. A continuación se detallan algunas de las bases de datos utilizadas en investigaciones relacionadas con el reconocimiento de género.

#### **2.4.1.1 Color FERET**

La base de datos FERET, desarrollada por el programa "*Facial Recognition Technology*", es un recurso público disponible como un estándar común para experimentos de reconocimiento de rostro (Phillips et al., 1998)(Phillips et al., 2000). Las imágenes de FERET fueron tomadas con la cámara *color Kodak Ultra* en condiciones semi-controladas, ya que se coleccionaron en 15 sesiones entre 1993 y 1996, lo que implica un cierto grado de variabilidad en las condiciones de adquisición, como los cambios de iluminación y pose.

En el año 2003 se libera una nueva versión de ésta base de datos, denominada Color FERET, la cual contiene múltiples imágenes de 994 individuos en varios ángulos, 591 hombres y 403 mujeres, específicamente 11338 imágenes de caras humanas con una resolución de 512x769 píxeles. Las imágenes de Color FERET son un sub-conjunto de la mayoría de las imágenes de la versión anterior, pero en color de alta resolución. Una pequeña muestra de las imágenes contenidas en esta base de datos se pueden ver en la Figura II.4



**Figura II.4** Conjunto de imágenes obtenidas de la base de datos FERET.

#### **2.4.1.2 PAL**

La base de datos de cara PAL construida en *Productive Aging Laboratory* de la Universidad de Texas en Dallas (Minear & Park, 2004), consiste imágenes frontales de 576 individuos, 219 hombres y 357 mujeres, en que sus edades varían entre 18 y 93 años. Por cada participante, se recupera la información asociada a su género, edad y etnia o raza. Las imágenes se etiquetan con el género, el grupo de edad (18-29, 30-49, 50-69, y 70-93) y la raza (Africano-Americano, Caucásico u otros).

Las imágenes fueron tomadas con una cámara digital *Kodak DC 3400 zoom 2.0 megapixel* utilizando flash y una resolución de 1760x1168 píxeles. Se les solicitó a los participantes asumir una expresión neutra y para mirar directamente a la cámara. Las imágenes adquiridas fueron editadas para una mayor uniformidad debido a que fueron tomadas en diferentes lugares. Finalmente se redimensionan a 640x480 píxeles. Algunas de las imágenes contenidas en esta base de datos se pueden apreciar en la Figura II.5.



**Figura II.5** Conjunto de imágenes de la base de datos PAL.

### 2.4.1.3 MORPH II

MORPH (*Craniofacial Longitudinal Morphological Face Database*), es la base de datos de imágenes de caras más grande a disposición del público (Ricanek & Tesafaye, 2006). Se disponen tres series de imágenes para el público, dos de uso comercial Álbum 1 y Álbum 2, y un subconjunto de imágenes del Álbum 2 disponible para investigaciones académicas que lo denominaré como MORPH II. Las tres series de imágenes incluyen un metadato con información de la raza, genero, fecha de nacimiento y fecha de adquisición de la imagen.

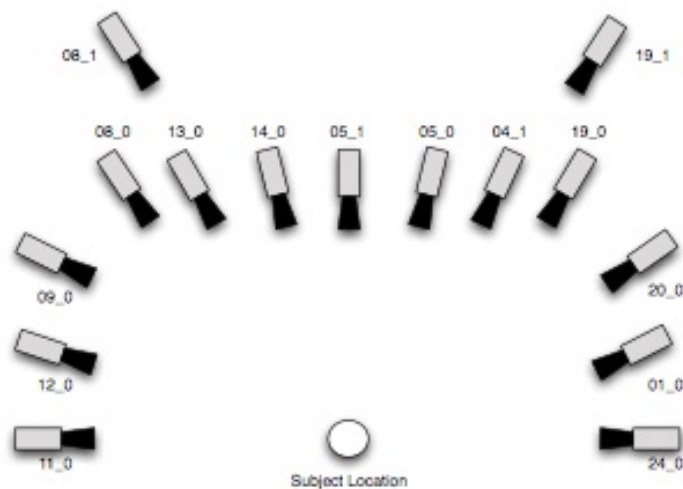
Específicamente, la base de datos MORPH II se compone de 55134 imágenes en color de más de 13000 personas, con diferentes resoluciones de 200x240 o bien 400x480 pixeles, tomadas desde el año 2003 hasta el 2007. El rango de edad varía desde 16 a 77 años, presentando la mayor cantidad de imágenes desde los 20 a los 50 años. El promedio de imágenes por persona son 4 y el tiempo medio entre cada foto es de 164 días. MORPH II tiene desbalanceado el género, donde 46645 imágenes corresponden a hombres y 8489 a mujeres. También tiene desbalanceada la raza, donde el 80% corresponde a africanos y el resto a europeos, asiáticos, hispanos u otros. En la Figura II.6 se muestran algunas imágenes de la base de datos MORPH.



**Figura II.6** Conjunto de imágenes de la base de datos MORPH.

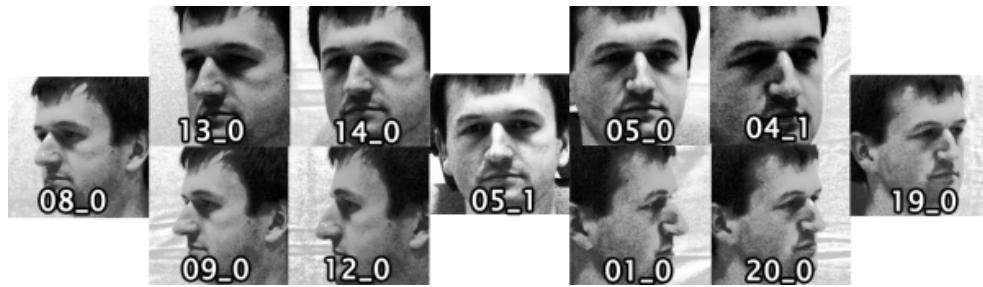
#### 2.4.1.4 MULTI-PIE

La base de datos Multi-PIE (Gross et al., 2008), creada por investigadores de Carnegie Mellon University, contiene imágenes de caras adquiridas en diferentes condiciones de iluminación, pose y expresiones faciales. En total almacena 755370 imágenes, que fueron tomadas de 337 individuos, 235 hombres y 102 mujeres, en cuatro sesiones en un período de 6 meses. Cabe destacar que los individuos son de diferentes etnias (60% europeos-americanos, 35% asiáticos, 3% africanos y 2% otros) y su edad promedio es de 27,9 años. Específicamente las imágenes con variación de pose e iluminación se capturaron con 15 cámaras en forma simultánea aplicando 19 cambios de iluminación, 13 colocadas a la altura de la cabeza del sujeto, con una variación de 15% grados entre ellas, y 2 sobre el sujeto tal como se muestra en la Figura II.7. La mayoría de las cámaras fueron producidas por Sony (11 de 15) y el resto por Panasonic (posiciones: 11\_0, 08\_1, 19\_1 y 24\_0).



**Figura II.7** Etiquetas y distribución aproximada de las cámaras dentro de la sala.

Un ejemplo de imágenes contenidas en la base de datos con variaciones de pose, iluminación normal de la sala y expresión neutral, adquiridas por las cámaras Sony en la primera sesión se puede apreciar en la Figura II.8.



**Figura II.8** Conjunto de imágenes MULTI-PIE.

## 2.4.2 Condiciones Reales

Bases de datos en condiciones reales contienen imágenes tomadas en entornos sin restricciones, donde se encuentran presentes cambios de iluminación, oclusiones, pose aleatoria de las personas, cambios de perspectivas, expresiones, entre otras.

### 2.4.2.1 LFW

La base de datos de caras “*Labeled Faces in the Wild*”, consiste de 13233 imágenes de caras de 5749 sujetos (4263 hombres, 1486 mujeres), con una resolución de 250x250 píxeles, obtenidas desde la web (Huang et al., 2007). De estos, 1680 personas tienen más de dos imágenes y el resto sólo una imagen en la base de datos. Cabe destacar que la mayoría de las imágenes están en color, pero tienen unas pocas en tonalidades grises. La distribución de la edad se encuentra enfocada a adultos de mediana edad. Además las imágenes no se encuentran balanceadas en género, siendo 10256 hombres y 2977 mujeres. En la Figura II.9 se muestran algunas de las imágenes de ésta base de datos.



**Figura II.9** Conjunto de imágenes de la base de datos LFW.

### 2.4.2.2 GROUPS

La base de datos GROUPS consiste de 28231 imágenes de caras extraídas desde 5800 imágenes de grupos de personas, etiquetadas con el género y categoría de edad (0-2, 3-7, 8-12, 13-19, 20-36, 37-65, y 66+) (Gallagher et al., 2009). Muchas de las imágenes de las caras son de baja resolución. La media tiene solo 18.5 píxeles de distancia entre el centro de los ojos y el 25% de las imágenes bajo 12.5 píxeles. Ésta base de datos contiene imágenes capturadas en entornos sin restricciones, adquiridas desde *Flickr image*, que involucra diferentes condiciones de iluminación, expresiones faciales, poses de caras y amplia distribución demográfica. La Figura II.10 contiene una muestra de las imágenes de la base de datos.



**Figura II.10** Conjunto de imágenes de la base de datos GROUPS.

## 2.5 Clasificación de Género

La literatura relacionada con este tema se puede dividir en dos grandes grupos: clasificación de género en condiciones ideales, y clasificación de género en condiciones reales. El primer grupo se refiere a todas las publicaciones que utilizan una muestra de imágenes de caras obtenidas en condiciones controladas o semi-controladas para la construcción del modelo de clasificación. En cambio el segundo, se relaciona con el uso de una muestra de imágenes de caras obtenidas en condiciones sin controlar, es decir imágenes en entornos sin restricciones.

En la Tabla II.1 se muestra un resumen de las publicaciones directamente relacionadas con la clasificación de género revisadas, ordenadas desde el año 1990 hasta el año 2014, destacando el tipo de algoritmo de extracción de característica, el clasificador y el tipo de base de datos utilizada, además del rendimiento obtenido. Cabe señalar que enfoques basados en apariencia fueron ampliamente utilizados hasta el año 2005 logrando una tasa de acierto de alrededor de un 96.6% en el mejor caso sobre imágenes en condiciones controladas. Sin embargo a partir del año 2005 a la fecha, la tendencia de las investigaciones se ha inclinado a usar enfoque basados en características obteniendo buenos resultados al evaluarlos con imágenes en entornos controlados o sin controlar, la tasa de acierto obtenida en el mejor de los casos es de un 99,1% y 89.79% respectivamente. Específicamente el rendimiento de un 89.79% sobre imágenes en entornos sin controlar (GROUPS) se logra por medio del uso de 12 clasificadores, por lo tanto el costo computacional hace que esta propuesta no sea una solución eficiente para clasificar imágenes en tiempo real. De los resultados obtenidos en las publicaciones, se destacan con rojo aquellos que tienen el mejor rendimiento, en cuanto a la tasa de acierto, en la clasificación de género sobre imágenes

tomadas en entornos controlados (FERET) y en entornos sin restricciones (LFW y GROUPS).

**Tabla II.1** Rendimiento obtenido en la clasificación de género

Año	Enfoque		Base de datos		Clasificador	Tasa de acierto	Autor
	Apariencia	Característica	Ideal	Real			
1990	Pixel	-	Privada	-	Red neuronal	91.9%	Golomb et al
2002	Pixel	-	FERET	-	SVM	96.6%	Moghaddam et al
	-	Haar-like	-	WEB	AdaBoost	78.0%	Shakhnarovich et al
	PCA	-	Privada	-	SVM,LDA,NN, B	95.3%,91%	Sun et al
2005	PCA	PCA ( O, B)	FERET	-	SVM	92.3%	Buchala et al
2006	-	Comp. Pixel	FERET	-	AdaBoost	94.4%	Baluja et al
2008	Haar-like	AAM	FERET, A, R, WEB	-	SVM	92.4%	Xu et al
	-	Haar Feature	FERET	-	LUT AdaBoost	93.3%	Makinen et al
	-	LBP	-	WEB	SVM	67.3%	
2010	-	Color, Borde	-	WEB	AdaBoost	87.7%	Dong Chen et al
2011	-	LGBP-LDA	<b>FERET</b>	-	SVMAC	<b>99.1%</b>	Zheng et al
	Pixel	-	FERET	-	PCA+LDA	93.6%	Bekios-Calfa et al
	-	Gabor	-	LFW	PCA+SVM, PCA+LDA	94.01%,93.41%	Dagos Casas et al
	-	Gabor	-	GROUPS	PCA+SVM, PCA+LDA	86.1%,86.58%	
2012	-	(F,O,N,B,BA)+(P,R)	FERET	-	SVM	95.10%	Li et al
	-	SWLD	<b>FERET</b>	-	Chi-cuadrado	<b>99.08%</b>	Ullah et al
	-	LBP + AdaBoost	-	<b>LFW</b>	SVM	<b>94.81%</b>	Shan
	-	LBP	-	LFW	SVM	90.60%	Ramón-Balmaceda et al
	-	LBP	-	GROUPS	SVM	82.65%	
2013	-	LBP	-	LFW	PCA+ Joint Bayesian	93.18%	Duang Chen et al
2013	PCA+DCT	LBP+DCT	FERET	-	KNN	98.16%	Mirza et al
2013	-	LBP,HOG (R,H,C)	-	<b>GROUPS</b>	SVM (11)+ Naive Bayes	<b>89.79%</b>	Castrillon Santana et al
2014	Pixel	-	-	GROUPS	PCA+LDA	80.53%	Bekios-Calfa et al

A continuación se describen las investigaciones mostradas en la Tabla II.1, colocando especial énfasis a los métodos utilizados para clasificar, el rendimiento obtenido y otros aspectos de interés para el trabajo a realizar.

### 2.5.1 Condiciones Ideales

Las investigaciones sobre la clasificación de género se inician en el año 1990. El primer algoritmo para clasificar género desde el rostro fue desarrollado por Golomb et al. (1990) y se denomina SEXNET. SEXNET se basa en redes neuronales multi-capa con *backpropagation*. El algoritmo utiliza dos redes neuronales, una red es usada para comprimir las imágenes y la segunda red para discriminar entre sexos. Como se utilizan imágenes de  $30 \times 30$ , para la primera red se utilizan 900 neuronas de entrada, 40 neuronas en la capa oculta y 900 neuronas en la capa de salida. Para la segunda red se toma como entrada las 40 neuronas de la capa oculta de la red anterior, 40 neuronas en la capa oculta y 1 neurona de salida. La neurona de salida entrega un 1 en caso de ser hombre y un 0 en caso de ser mujer. El error reportado es de 8.1%.

Una de las mejores tasas de acierto en la clasificación de género, sobre imágenes de caras de la base de datos FERET, fue obtenido por Moghaddam et al. (2002), que a diferencia de los estudios anteriores, se utiliza el algoritmo de clasificación SVM con RBF Kernel Gaussiano, con lo cual se logra una tasa de acierto de un 96.6%. El clasificador SVM obtuvo un rendimiento de alrededor de un 10% superior al clasificador FDA. También compararon el rendimiento de la clasificación sobre imágenes con dos resoluciones,  $(21 \times 21)$  y  $(84 \times 84)$ , en donde obtuvieron una diferencia de 1% a favor de las imágenes de mayor tamaño.

La importancia de seleccionar un buen conjunto de características previo a la clasificación fue reportado por Sun et al. (2002). Además demuestran que al remover características que no aportan información importante, para determinar el género, se reduce el error en la clasificación. Para la selección

de características, en primera instancia utilizaron PCA para representar cada imagen como un vector de características en un espacio de baja dimensionalidad, y luego emplearon algoritmos genéticos para obtener sub-conjunto de características que contenga la información más relevante de este espacio dimensional. En la experimentación, se comparan 4 clasificadores: Bayesiano, LDA, NN y SVM. Previo al entrenamiento, se ecualizó el histograma de las imágenes para suavizar las diferentes condiciones de iluminación. La muestra consiste en 400 imágenes de una base de datos privada, donde 300 son utilizadas para el entrenamiento, 50 para la validación y 50 para las pruebas. Los resultados obtenidos reafirman la importancia de la selección de un sub-conjunto de características, reduciendo el error significativamente en todos los clasificadores: de un 22.5% a un 13.3% usando clasificador bayesiano, de un 17.7% a un 11.3% usando red neuronal (NN), 14.2% a un 9% usando LDA y finalmente de un 8.9% a un 4.7% usando SVM.

El rol que tienen la información global y local en el reconocimiento de género y que su uso en conjunto puede mejorar la tasa de acierto en la clasificación fue mostrado en Buchala et al. (2005). La idea surge dado que en estudios previos se utilizó la cara completa para clasificar, dando la misma importancia a cada pixel. En la experimentación se ocuparon en conjunto técnicas representación global y basada en características, para el enfoque global se utilizó la imagen reducida de toda la cara (64x64) y para el enfoque local se utilizaron dos regiones (32x64) que contienen los ojos y la boca respectivamente. A estas tres imágenes se le realizó una reducción de dimensión (PCA, CCA, SOM), con la finalidad de eliminar redundancia y datos irrelevantes, y luego se clasificó con SVM. En la experimentación, se utilizaron las bases de datos FERET, AR y Biold, en cada una se extrajeron 400 caras con la misma cantidad de hombres y mujeres. Se obtuvo una tasa de acierto

de 92.25% utilizando la fusión de características locales y globales respecto a un 86.5% obtenido usando solo características globales.

Un novedoso método basado en AdaBoost sobre imágenes de caras en escalas grises fue presentado en Baluja et al. (2006). AdaBoost combina múltiples clasificadores débiles, un poco mejor que aleatorio, para generar un clasificador con mejor tasa de acierto. El método implementado es sencillo, basado en comparación de píxeles, y bastante rápido. La tasa de acierto obtenida es sobre un 93%, lo cual se asemeja o mejora las tasas de acierto obtenidas con SVM, pero con un rendimiento 50 veces más rápido. La base de datos utilizada es la de FERET usando solo las caras frontales.

Una mejora en el proceso de clasificación fue presentado en Xu et al. (2008), quienes se basan en un enfoque híbrido mediante la fusión de características globales y locales. Para extraer características globales se utilizó AdaBoost y para la extracción de característica local se usó un modelo de apariencia activo (AAM). Para el entrenamiento y validación, se usan un conjunto de imágenes extraídas de las bases de datos FERET, AR y en menor cantidad de la WEB. La tasa de acierto obtenida es de 92.38% con el enfoque híbrido, 82.97% con el enfoque global, por medio de 143 clasificadores débiles, y 88.55% con el enfoque local utilizando 10 características. El enfoque híbrido muestra efectividad y robustez sobre imágenes de rostro con cambios de expresión, iluminación y leves variaciones de pose.

Un análisis del rendimiento de diversos métodos de clasificación de género como SVM, Red Neuronal, AdaBoost con diferentes métodos de alineación fue hecho por Mäkinen et al. (2008). Los experimentos fueron realizados sobre la base de datos FERET y WWW. En el caso de FERET se utilizaron 1196 imágenes de caras, en donde se removieron las personas

duplicadas, quedando finalmente 900 imágenes frontales de igual cantidad de hombres y mujeres. En el caso de WWW se utilizaron 4720 imágenes frontales de caras con los géneros balanceados. En FERET la mejor tasa de acierto se obtuvo con el algoritmo LUT AdaBoost, la cual fue de 93.33% sobre imágenes normalizadas, es decir, rotadas y alineadas. Los mismos algoritmos fueron probados en la base de datos WWW, logrando tan solo un 67.25% de tasa de acierto con el algoritmo LBP+SVM en el mejor de los casos. Dentro de las conclusiones de los autores se menciona que es estadísticamente significativa la diferencia en la tasa de acierto de los clasificadores al pre-procesar (normalizar) las imágenes previo a la clasificación y además que el usar imágenes con o sin pelo no representa una diferencia significativa en la clasificación.

La mejor tasa de acierto en la clasificación de género documentada fue obtenida en Zheng & Lu (2011), quienes al evaluar el clasificador sobre imágenes de caras frontales de la base de datos FERET obtienen una tasa de acierto del 99.1% y sobre imágenes con una variación de pose de hasta 30° (*yaw*, *pitch*), obtenidas de la base de datos CAS-PERAL, alrededor del 100%, pero para esto último fue necesario un detector de pose. El algoritmo de clasificación utilizado es SVMAC, denominado máquina de soporte vectorial con confianza automática. El desarrollo experimental involucró en la evaluación de tres métodos de clasificación (KNN, SVM y SVMAC) aplicados a cinco algoritmos de extracción de características (Gray, Gabor, LBP, MLBP y LGBP) sobre 11 subconjuntos creados, 9 de la base de datos CAS-PERAL separando por pose, y las dos restantes corresponden a FERET y BCMI. SVMAC en todas las pruebas experimentales fue superior a KNN y SVM tradicional.

Una revisión del estado del arte de técnicas lineales y sus relaciones para la clasificación de género fue presentado por Bekios-Calfa et al. (2011). En la fase experimental se evalúan cuatro clasificadores lineales, LDA, PCA+LDA, PCA-M+LDA, ICA+LDA; y se comparan los clasificadores que mejores resultados han entregado, SVM+RBF y AdaBoost de Baluja et al. (2006). Las bases de datos de imágenes de caras utilizadas son UCN (privada), Color FERET, PAL y Gray FERET. Al entrenar y validar los clasificadores, se aprecia que LDA requiere de una gran cantidad de datos para obtener resultados competitivos, esto queda claro al analizar la diferencia de tasa de acierto que obtuvo el clasificador entre las bases de datos FERET y UCN, en que la primera con menos de 1000 imágenes se obtuvo tan solo un 77.68% de éxito, a diferencia de la segunda que con alrededor de 10000 imágenes se obtuvo una tasa de acierto de 92.65%. En general los demás clasificadores lineales obtuvieron resultados similares a SVM+RBF y AdaBoost. Al probar los clasificadores utilizando un enfoque de cruce de base de datos, es decir, utilizar una base de datos para entrenar y la otra para validar, se produce una disminución en la clasificación lo que evidencia que existe una relación entre diferencias demográficas de las variables tales como el género, edad, etnia u otros. Por lo tanto, cuando un clasificador es entrenado con una base de datos con escasa variabilidad demográfica, y luego, testeado con imágenes más generales, la tasa de clasificación disminuye significativamente.

Un *framework* para clasificar género que involucra el uso de características faciales e información externa como el pelo y la ropa fue desarrollado por Li et al. (2012). Respecto a las características faciales, consideraron cinco componentes del rostro: la frente, los ojos, la nariz, la boca y la barbilla. Para cada característica se entrena un único clasificador, y sus salidas son combinadas por varias estrategias. La tasa de acierto obtenida es

de un 95.1% sobre FERET al usar en conjunto las cinco componentes de la cara más las 2 componentes externas. Dentro de los problemas que señalan los autores, se encuentra el proceso que realizan para la extracción de característica, el cual se puede ver afectado por fondos no homogéneos, específicamente para el pelo. Además el *framework* propuesto demanda mucho tiempo en clasificar por lo que no puede ser implementado en imágenes de tiempo real.

Un simple método que obtiene similar tasa de acierto que el algoritmo propuesto por Zheng et al. (2011) basado en SVMAC fue propuesto por Ullah et al. (2012). Éste simple método consiste en un descriptor de textura denominado SWLD, el cual concatena cada vector de características obtenido al aplicar el descriptor WLD sobre una serie de bloques que dividen la imagen. Algunas de las propiedades de este descriptor es que es invariante a la iluminación y robusto al ruido. Luego de una serie de experimentos, se determinan los parámetros adecuados para SWLD que implique una mejor tasa de acierto sobre la clasificación de género en imágenes frontales de FERET. La clasificación se determina mediante la distancia mínima obtenida por una función Chi-Cuadrado, obteniendo una tasa de acierto de un 99.08%.

Un método híbrido que consiste en fusionar la extracción de características globales y locales obtenidas de las imágenes de caras, basados en la premisa que la percepción humana de la cara involucra una observación global y detallada de la cara para categorizar una persona, fue desarrollado por Mirza et al. (2013). Enfoques PCA y DCT son utilizados para extraer características globales y un LBP aumentado con 2d-DCT es utilizado para encontrar características locales. El entrenamiento y validación del modelo de clasificación se hizo con imágenes de caras frontales de la base de

datos FERET normalizadas y cortadas a 60x48 pixeles. La tasa de acierto obtenida es de 98.16%.

### **2.5.2 Condiciones Reales**

Investigaciones de reconocimiento de género sobre imágenes en entornos no controlados surgen en el año 2002, pero su mayor auge inicia luego de la creación de la base de datos LFW y GROUPS en los años 2007 y 2009 respectivamente. Al ser estas bases de datos públicas, permiten comparar y evaluar el rendimiento de clasificadores propuestos por los investigadores. Con estas bases de datos se genera un nuevo desafío para los investigadores, ya que, a diferencia de las anteriores, contienen imágenes tomadas de la web sin restricciones.

Los primeros en clasificar demográficamente imágenes obtenidas de *World Wide Web* (WWW) fueron Shakhnarovich et al. (2002), quienes lograron una tasa de acierto de alrededor del 78%, lo cual es muy inferior al clasificador propuesto por Moghaddam et al. (2002) el mismo año sobre FERET. Su trabajo se enfoca en la creación de un sistema integrado, robusto, que involucra la detección de rostro y análisis demográfico en tiempo real. Para la detección de la cara se utilizó el algoritmo de Viola & Jones (2001), la extracción de características con Haar-like y la clasificación mediante AdaBoost. El clasificador demográfico propuesto, sobre imágenes sin alinear ni rectificar, distingue género e identifica si la persona es asiático o no. Si bien la tasa de acierto obtenida es baja, al evaluar el algoritmo de Moghaddam et al. (2002) sobre este conjunto imágenes el rendimiento decrece bruscamente a un 77% de acierto y es 1000 veces más lento.

Un sistema de reconocimiento de género en tiempo real sobre imágenes en condiciones no controladas fue propuesto por D.-Y. Chen & Lin (2010). Su sistema se basa en un conjunto de características híbridas en el que se computa el color y las características de los bordes de las imágenes. La clasificación se realiza mediante un enfoque basado en AdaBoost. El clasificador se prueba con 465 caras extraídas de 100 imágenes recolectadas del motor de búsqueda de google y se obtiene una tasa de acierto de un 87.65%.

Un protocolo estándar para la clasificación de género sobre la base de datos GROUPS fue establecido en Dago-Casas et al. (2011). Además realizan una serie de experimentos para evaluar cómo se comportan diversos algoritmos de descripción global y local en experimentos en condiciones realistas. Se hace una revisión y comparación de algoritmos Pixels, Gabor y LBP, para la extracción de características, y para la clasificación PCA+SVM y PCA+LDA. Se utilizan las bases de datos LFW y GROUPS para entrenar y validar el modelo de clasificación. En ambas se usa la estrategia “*cross-validation*”, que consiste en entrenar y validar sobre la misma base de datos, y la estrategia “*cross-database*”, que consiste en entrenar con una base de datos y validar con otra. En los experimentos realizados con “*cross-validation*” se obtuvo en el mejor caso, usando el descriptor local Gabor y clasificando con PCA+SVM, un 94.01% sobre LFW y un 86.1% sobre GROUPS. Cabe señalar que al clasificar con PCA+LDA se obtienen resultados similares, y también al usar los demás algoritmos de extracción de características se obtiene un buen rendimiento, con la excepción del basado en apariencia. Por medio de la estrategia “*cross-database*”, se obtiene mejor tasa de acierto al utilizar GROUPS en el entrenamiento y al validar con LFW que al revés, por lo tanto se deduce que las condiciones demográficas y el balanceo de género que hay en GROUPS permiten una mejor generalización. Finalmente, proponen para la

clasificación de género en condiciones sin restricciones el uso de la base de datos GROUPS.

Un método de clasificación de género, sobre la base de datos LFW, basado en el uso de LBP para extraer las características del rostro, y AdaBoost para seleccionar las características más discriminativas fue desarrollado en Shan (2012), quienes lograron un 94.81% de tasa de acierto mediante el uso del clasificador SVM. En la experimentación verificaron que la mayoría de la información discriminativa para el género es contenida en los patrones uniformes de LBP. Otro aspecto que evidenciaron es que debido al desbalance de género contenido en las imágenes de caras de la base de datos LFW, alrededor de 4 veces más hombres que mujeres, el clasificador tiende a entregar una mejor tasa de acierto en la clase dominante.

El rendimiento de la clasificación de género utilizando base de datos grandes (>5000 imágenes) fue evaluado en Ramón-Balmaseda et al. (2012). Además verifican que la base de datos que proporciona la mejor tasa de acierto, también obtiene una buena tasa de acierto al evaluar con las demás bases de datos. Se utilizan tres bases de datos MORPH, GROUPS y LFW. La primera contiene fotos tomadas en condiciones ideales y las dos últimas en condiciones reales. Para la extracción de características se utiliza LBP y para la clasificación SVM. El mejor rendimiento se obtiene sobre MORPH dando una tasa de acierto de 97.64% al usar los patrones uniformes de LBP. Cuando se evalúa la clasificación sobre una base de datos distinta a la del entrenamiento la tasa de acierto decrece en un 20%, por lo que señalan que queda bastante por hacer para que se logre clasificar género obteniendo una buena tasa de acierto independiente de las condiciones de las imágenes.

El desempeño de los clasificadores sobre una función de alta dimensión fue estudiado en D. Chen et al. (2013), quienes demostraron que este requisito es fundamental para obtener un alto rendimiento. Dado a que la alta dimensión involucra mayores costos computacionales y de almacenamiento, proponen un método de proyección dispersa que permiten reducir los costos asociados a la alta dimensión sin sacrificar la calidad de tasa de acierto. El método propuesto se denomina “regresión rotación dispersa”. Al utilizar LBP para la extracción de características y PCA+Joint Bayesian para clasificar, obtiene una tasa de acierto de 93.18% sobre LFW sobre alta dimensión. Con el método propuesto se obtiene similar a la tasa de acierto, pero con un ahorro de recursos computacionales significativo.

La extracción de características de la cara junto a su contexto local (cabeza y hombros) para mejorar la tasa de acierto del reconocimiento fue estudiado por Castrillón-Santana, Modesto Lorenzo-Navarro & Ramón-Balmaseda (2013). Su trabajo se centra en la base de datos de imágenes GROUPS, recomendada por Dago-Casas et al. (2011) por tener un número equilibrado de hombres y mujeres, además de una gran variabilidad demográfica, de pose y condiciones, permitiendo así poder construir un modelo de clasificación que generalice mejor en entornos sin restricciones. Se estudian diferentes descriptores, resoluciones y clasificadores. El algoritmo que proponen utiliza combinaciones de descriptores LBP uniforme y HOG, tanto para cara como para hombros y/o cabeza, formando un total 11 clasificadores que utilizan SVM en la etapa de entrenamiento. Los resultados de cada uno de estos clasificadores son pasados a otro clasificador (SVM, Naive Bayes, C45, Nearest Neighbor y Bagging), el cual se encarga de entregar la decisión final. El mejor resultado se obtiene con Naive Bayes sobre la base de datos GROUPS con una tasa de acierto de 89.79%.

La caída en la tasa de acierto de los algoritmos de clasificación de género al ser sometidos a un ambiente en condiciones no controladas fue analizado en Bekios-Calfa et al. (2014). En su análisis plantean que la baja de rendimiento se produce debido a la dependencia de atributos demográficos faciales que no han sido considerados en el entrenamiento del clasificador, y se validan la dependencia de atributos de género, pose y edad. Luego de validar la dependencia, realizan un experimento utilizando un método denominado “*Attributes Powerset*”, que consiste en realizar un producto cruz entre clases de los atributos demográficos. En el caso de GROUPS, se realiza el método “*Attributes Powerset*” sobre el género y la edad, teniendo como resultado 14 clases (7 clases de edad x 2 de género). El mismo proceso se realiza sobre género y pose, obteniendo una tasa de éxito de 80.53% utilizando un clasificador lineal y múltiples variables clases. Cabe señalar que el enfoque multidimensional aumentó en un 2% la tasa de acierto de clasificador lineal.

## **2.6 Hipótesis de Investigación**

La clasificación de género (hombre, mujer) aplicado sobre imágenes adquiridas en entornos sin restricciones, por medio del clasificador lineal PCA+LDA, sobre el modelo de clasificación construido a partir de los vectores de frecuencias conformados por palabras visuales basadas en características extraídas con el descriptor SIFT, SURF u ORB, es suficientemente discriminativa, es decir, produce una tasa de acierto en la clasificación similar o superior a lo evidenciado en la literatura, cuando se considera la información contenida en las clases de cada instancia.

### **III. METODOLOGÍA**

#### **3.1 Método de investigación y tipo de diseño.**

La presente investigación se basa en el método experimental cuantitativo, donde el tipo de diseño es transaccional con un alcance descriptivo.

Es una investigación cuantitativa debido a que la hipótesis es planteada antes de la recolección y el análisis de los datos, a diferencia del enfoque cualitativo que por lo general la hipótesis es planteada durante el estudio de un fenómeno en particular o al final de éste (Hernández Sampieri, 2010). Además porque los datos son analizados estadísticamente. Específicamente para la validación de la hipótesis, se utiliza la tasa de acierto como indicador estadístico para medir el rendimiento para cada una de las propuestas desarrolladas sobre la clasificación demográfica. La tasa de acierto es aplicada sobre una muestra de imágenes, y posteriormente estos resultados son generalizados para explicar y evaluar, sobre la hipótesis planteada. El enfoque es experimental, donde el fenómeno de estudio corresponde a la clasificación demográfica. En la experimentación, se diseñan una serie de modelos de clasificación, los cuales son evaluados con la finalidad de obtener el más adecuado para imágenes tomadas en entornos sin restricciones. Finalmente, tiene un alcance descriptivo porque se busca situar a un grupo de personas en una o diversas variables, por ejemplo describir el género, edad, raza u otros, de algún individuo por medio de la clasificación demográfica. Cabe señalar que la hipótesis planteada también es descriptiva ya que pronostican un valor respecto a la tasa de acierto de la clasificación.

### **3.2 Unidad de análisis y el método de muestreo.**

La unidad de análisis es la tasa de acierto aplicada a los clasificadores demográficos sobre un conjunto de imágenes de caras humanas. Esta medida se utiliza para validar la hipótesis propuesta, además, de comparar éstas con otros modelos de clasificación encontrados en la literatura sobre entornos controlados y no controlados de adquisición.

Respecto al método de muestreo, este trabajo no contempla la adquisición de la muestra dado que se utilizan bases de datos públicas, de imágenes de caras humanas, ampliamente usadas por los investigadores. Las condiciones en que fueron adquiridas las imágenes de éstas bases de datos se mencionan en la sección 2.4.

### **3.3 Validez y confiabilidad de los instrumentos de medición**

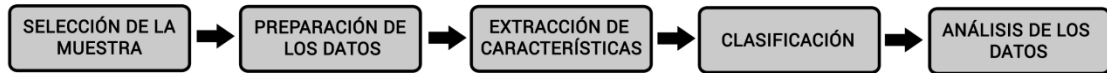
La investigación contempla la clasificación demográficamente de las personas (género, edad, etnia, etc.) utilizando para esto imágenes de su rostro que permitan estimar los atributos faciales contenidos en ella. Por lo tanto, para realizar éste análisis se requiere de una muestra que contenga imágenes de caras humanas. Esta condición se cumple, ya que las bases de datos descritas en la sección 2.4, tienen un conjunto de imágenes suficientes para realizar la etapa de experimentación. Al tratarse de imágenes, la evaluación del instrumento de medición contempla las características de las cámaras y el entorno en que fueron tomadas las imágenes. Sin embargo, esto no es un impedimento, ya que las diversas condiciones en que fueron adquiridas las imágenes de caras en las bases de datos son consideradas como parte del entrenamiento del modelo de clasificación. Finalmente, se puede decir que el instrumento de medición es confiable, debido a que la tasa de acierto obtenida

en la clasificación demográfica sobre la misma base de datos se mantiene, a menos que se cambia el modelo de clasificación; es válido ya que mide lo que se pretende medir, y además porque que las imágenes contenidas éstas bases de datos son ampliamente utilizadas en investigaciones asociadas a la clasificación demográfica.

### **3.4 Procedimientos**

Para el desarrollo experimental de la tesis se utilizó una estrategia experimental incremental, es decir, cada estudio posterior utiliza como base el proceso del estudio anterior. En total se realizan tres estudios experimentales. El primer estudio corresponde a la verificación de la información discriminante contenidas en los vectores de característica construidos a partir de los descriptores de SIFT y SURF. El segundo, que tiene como finalidad corroborar los resultados obtenidos en el primer estudio, se utiliza una estrategia distinta de entrenamiento y validación (LOPO), la cual permite minimizar el sesgo producido al tener varias imágenes de un mismo sujeto en la base de datos. Además, se incorpora un tercer descriptor al estudio denominado ORB. En ambos se ocupa como muestra las imágenes contenidas en la base de datos MULTI-PIE. El tercer, y último, estudio incorpora un enfoque de múltiples clases con la finalidad de evaluar si el método diseñado provee un mejor rendimiento en la clasificación de género sobre imágenes en entornos sin restricciones. Para llevar a cabo la evaluación se utiliza la base de datos GROUPS, y se descarta MULTI-PIE debido a que esta última no contiene imágenes en condiciones no controladas.

El proceso utilizado en cada uno de los estudios mencionados está compuesto principalmente de cinco etapas y que son definidas a continuación, ver Figura III.1:



**Figura III.1** Proceso utilizado para la estimación de género.

### **3.4.1 Selección de la muestra**

De todas las bases de datos de caras públicas revisadas en el marco teórico, se seleccionaron dos: MULTI-PIE y GROUPS. La primera, si bien es cierto es una base de datos construida en condiciones de laboratorio, es ideal para realizar investigaciones que involucren someter los métodos de clasificación sobre un conjunto de imágenes que contengan variaciones de iluminación, pose y/o expresión. La segunda, corresponde a una base de datos que contiene imágenes obtenidas en condiciones sin restricciones.

Para los experimentos realizados con MULTI-PIE, se seleccionan aquellas imágenes pertenecientes a la galería sin iluminación y expresión facial neutral. Se utilizan todos los individuos en sus diferentes poses adquiridas en la primera sesión con cámaras Sony modelo DXC-9000. Se acota la muestra de imágenes que provee la base de datos a este conjunto ya que se decide observar el rendimiento del clasificador al tener que verificar el género sobre imágenes de caras con variaciones de poses, debido a que esto último es una de las múltiples variables que se presentan en entornos reales.

En el caso de GROUPS, se utilizan todas las imágenes caras que tengan un tamaño superior a 60x60 y que no sean niños. No se utiliza toda las imágenes de la base de datos ya que no fue posible llevar a cabo la experimentación debido a los recursos computacionales que involucran el

cálculo de los clúster por el algoritmo k-medias utilizados por los modelos basados en características propuestos. De todas maneras, hay estudios que eliminan los niños y/o imágenes de baja resolución (Bekios-Calfa et al., 2014)(Dago-Casas et al., 2011)(Castrillón-Santana, Modesto Lorenzo-Navarro et al., 2013), y por ende es posible contrastar los resultados obtenidos con esas publicaciones.

### **3.4.2 Preparación de los datos**

La preparación de los datos contempla en primer lugar la detección de una cara en la imagen, luego se recorta y finalmente es pre-procesada para eliminar elementos redundantes no contemplados en éste análisis. Para la detección del rostro se utiliza el algoritmo descrito en Viola & Jones (2004). Los rostros detectados son revisados manualmente para eliminar aquellos que fueron identificado de forma errónea (falsos positivos), es decir, aquellas imágenes consideradas como rostro cuando en realidad no lo eran. En la Figura III.2 se muestra que el algoritmo detecta dos rostros en una imagen, el que está en azul es el rostro correcto y el rojo es el denominado falso positivo. Por otro lado, también se presentan casos en que solo se detectan rostros correctos o falsos positivos.

En el caso de MULTI-PIE, que tienen una resolución de 640x480 pixeles, al aplicar el algoritmo de detección sobre el conjunto de muestra seleccionado, se obtiene en total 2625 imágenes, 1865 hombres y 760 mujeres, con una resolución media de 200x200 pixeles (ver Tabla III.1).



**Figura III.2** Detección de rostro por medio de Viola & Jones.

**Tabla III.1.** Frecuencia de imágenes de caras de hombres y mujeres distribuidas por pose.

Pose	Hombre	Mujer
12_0	164	51
09_0	173	69
08_0	173	75
13_0	166	74
14_0	173	76
05_1	173	76
05_0	173	76
04_1	161	73
19_0	173	73
20_0	172	68
01_0	164	49

En el caso de GROUPS, luego de aplicar el algoritmo de detección y eliminar los falsos positivos, se obtiene un total de 12037 imágenes, de las cuales 9630 son usadas para entrenamiento y las 2407 restantes son para prueba. Cabe mencionar que se mantiene el balance de género presentado en esta base de datos, donde alrededor del 48% corresponden a hombres y el

52% restante a mujeres. En cuanto a la resolución de las imágenes, estas fluctúan desde 60x60 hasta 550x550.

En la Tabla III.2 se muestran el conjunto de imágenes seleccionadas de la base de datos GROUPS estratificadas por género y pose. Cabe destacar que esta base de datos no contiene etiquetada la pose de los individuos, por ende, para poder llevar a cabo los experimentos fue necesario obtenerla, para ello se utiliza el mismo método propuesto en el artículo de Bekios-Calfa et al. (2014), en donde se calcula la pose mediante la clusterización de la posición de los ojos.

**Tabla III.2** Distribución de imágenes por género en cada pose.

Pose	Hombre	Mujer	Total
0	546	245	791
1	1730	1705	3435
2	865	1485	2350
3	1040	990	2030
4	970	980	1950
5	580	901	1481
	<b>5731</b>	<b>6306</b>	<b>12037</b>

Una vez obtenidos los rostros de cada una de las imágenes, éstos son transformados a escala de grises (intensidades), y luego se aplica una ecualización del histograma para mejorar el contraste y normalizar la iluminación.

Finalmente, para modelos basados en apariencia se realizó un pre-procesamiento adicional, que consiste en redimensionar la imagen a un tamaño de 25x25. Para los modelos basados en característica (SIFT, SURF y ORB) se utiliza la imagen del tamaño original en que fue detectado el rostro.

### 3.4.3 Extracción de características

En el caso del modelo basado en apariencia, la construcción del vector de características es bastante simple, en donde cada intensidad que pertenece a la imagen redimensionada (25x25 píxeles) es utilizada para conformar un vector de 625 dimensiones.

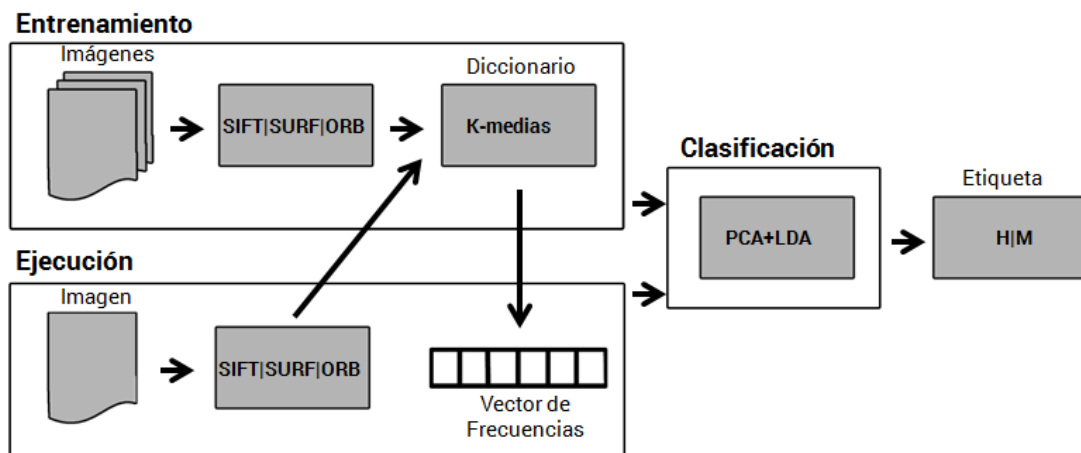
Por otro lado, para modelos basados en características, se utiliza y adapta el modelo de espacio vectorial (Salton et al., 1975) que es utilizado para la representación y búsqueda de documentos. El modelo representa los documentos de texto como vectores definidos como histogramas de las frecuencias de palabras que se encuentran en éstos. Este modelo también es denominado representación *bag-of-word*, donde el orden y la localización de las palabras son ignorados (Jan Erik Solem, 2012).

Para aplicar esta técnica de recuperación información sobre imágenes fue necesario crear un elemento visual similar a las palabras utilizadas en los documentos. Esto se logra utilizando SIFT, SURF y ORB para la detección y descripción de puntos de interés de la imagen. Los descriptores de un conjunto de imágenes son almacenados para construir un diccionario de palabras visuales. El diccionario se construye identificando grupos de descriptores por medio de un algoritmo de *clustering* k-medias. El algoritmo asigna un número de *clusters* o de pertenencia a los nuevos descriptores. Finalmente, el modelo de espacio vectorial se construye a partir del modelo definido por k-medias y que actúa como un diccionario de palabras. El vector resultante tiene tantas dimensiones como *clusters* sean definidos en el modelo k-medias, en el caso de los dos primeros experimentos se consideraron 1000 *clusters*, y para el tercero se evaluó el rendimiento variando la cantidad de *clusters* desde 100 a 4000. Por último, el vector de características se construye a partir de la frecuencia con la

que aparecen las palabras visuales en la imagen a partir de los descriptores obtenidos por los algoritmos (ver Figura III.3).

En resumen, el proceso que se realiza es el siguiente:

1. Se obtienen los descriptores asociados a cada punto detectado en las imágenes.
2. Se toma todos los descriptores y se agrupan en  $n$  clusters por medio del algoritmo de k-medias, donde  $n$  corresponde al número de clusters utilizado en cada experimento.
3. Por cada imagen se construye un vector de característica de  $n$  dimensiones, donde cada dimensión representa la frecuencia de los descriptores que aparecen en la imagen y que corresponde a un cluster en específico asociado a esta dimensión.



**Figura III.3.** Modelo de clasificación propuesto basado en la construcción de vector de frecuencias.

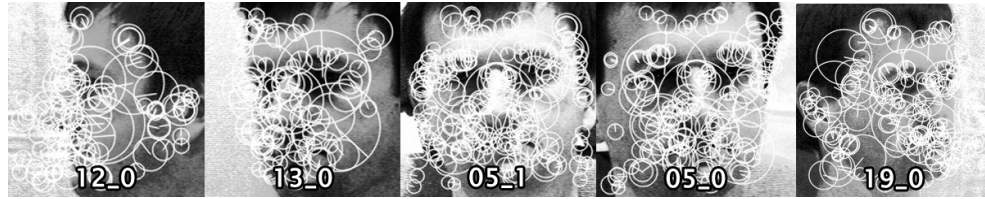
El cálculo de SIFT, SURF y ORB se realiza utilizando la librería de OpenCV, los cuales son configurados inicialmente con los parámetros

recomendados en sus publicaciones (Lowe, 2004)(Bay et al., 2008)(Rublee et al., 2011). SIFT fue diseñado para la detección y reconocimiento de objetos en general, es decir objetos rígidos y con transiciones bruscas entre sus diferentes lados, pudiendo identificar claramente cada segmento que lo conforma. Sin embargo, las caras no son rígidas ni presentan transiciones bruscas, por lo que utilizar SIFT con las configuraciones propuestas no es recomendable (Al-Azzawy & Al-Azzawy, 2012). Por lo tanto, se configura SIFT basado en el enfoque propuesto por Geng & Jiang (2009) denominado Volume-SIFT para eliminar los puntos no confiables, los cuales se identifican analizando la escala de los puntos de interés a diferencia del algoritmo original que lo hace por medio del contraste. En el caso de SURF se utilizan los parámetros recomendados por su autor, variando solo el umbral del Hessiano de manera tal de obtener una cantidad de descriptores similares a los obtenidos con SIFT. En el caso de ORB, se deja la configuración por defecto, limitando a un máximo de detección de 100 *features*.

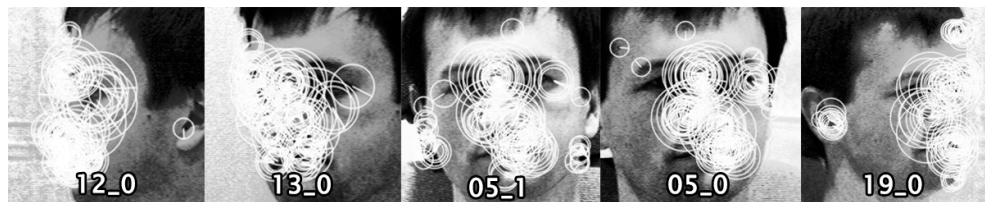
Los puntos de interés (*features*) detectados por SIFT, SURF y ORB, sobre un mismo sujeto de la base de datos MULTI-PIE, se pueden apreciar en la Figura III.4, Figura III.5 y Figura III.6 respectivamente.



**Figura III.4.** Puntos de interés detectados con SIFT sobre imágenes de un mismo sujeto con variación de pose.



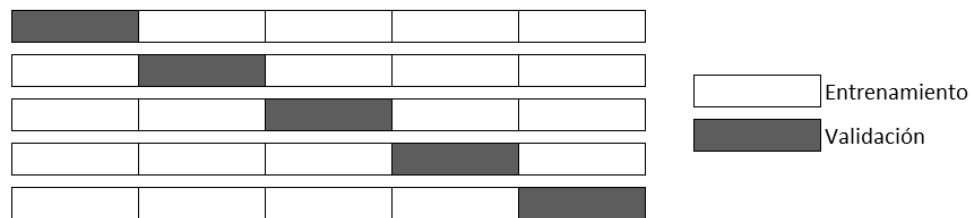
**Figura III.5.** Puntos de interés detectados con SURF sobre imágenes de un mismo sujeto con variación de pose.



**Figura III.6.** Puntos de interés detectados con ORB sobre imágenes de un mismo sujeto con variación de pose.

### 3.4.4 Clasificación

Para el primer estudio, en modelos de clasificación basados en apariencia y basados en características, el entrenamiento se realiza con validación cruzada utilizando 5 *fold* estratificados por género. Es decir, el conjunto de datos se divide en 5 secciones, que mantienen la proporción de hombres y mujeres, en las que una sección es para validar y las cuatro restantes para entrenar. El proceso se realiza 5 veces cambiando la sección de validación. (Ver Figura III.7)



**Figura III.7.** Representación visual de la estrategia validación cruzada 5 *fold*.

En el caso del modelo de apariencia, en cada iteración se toman los vectores de características asociados a las secciones de entrenamientos y se les aplica la reducción de dimensión por medio de PCA + LDA, descrito en Bekios-Calfa et al. (2011). Éstos vectores se utilizan para entrenar al clasificador lineal *Gaussian Naive Bayes*. Luego, al igual que con los vectores de entrenamiento, se reduce la dimensión de los vectores de la sección de prueba, y por consiguiente se le pasan al clasificador para obtener la clase estimada. Por último, se compara la clase estimada con la clase real para obtener la tasa de acierto en la clasificación. El número de componentes utilizados en PCA se determinan empíricamente seleccionando la configuración que permita mejorar la tasa de acierto media obtenida por las iteraciones.

En el caso del modelo basado en características locales el proceso es similar al anterior, con la salvedad de que en cada iteración se tienen que realizar los pasos 2 y 3 mencionados en la sección previa. Es decir por cada iteración se toman los descriptores asociados a las imágenes de entrenamiento con los que se generan los 1000 *clusters*, y luego se construyen los vectores de características asociado a cada imagen. Con relación al entrenamiento y validación del clasificador es idéntico a la descrita previamente.

Para el segundo estudio, el proceso que se realiza es similar para ambos enfoques de clasificación, con la salvedad de que se cambia la estrategia de validación a *Leave One Person Out* (LOPO). Ésta estrategia divide la muestra en una cantidad de secciones igual al número de personas contenidas en la base de datos, cada sección contiene las imágenes de un sólo individuo. En este caso, como se utilizó MULTI-PIE son en total 249 secciones. Por lo tanto, cada iteración corresponde al entrenamiento del clasificador con 248 secciones y la validación con la sección restante.

Para el tercer estudio, el proceso utilizado es similar al primero, en donde se cambia la base de datos objetivo a GROUPS. Por otro lado, se contemplan dos maneras de llevar a cabo la clasificación. Una que consiste en ocupar sólo la etiqueta del género para el entrenamiento y para la validación, la cual también fue utilizada en los dos estudios anteriores. Y la otra, denominada clasificación por medio de atributos *powerset*, que utiliza el producto cruz del género con la pose para entrenar y sólo el género para validar. Finalmente, la última modificación con relación a los estudios anteriores es la cantidad de *clusters*, los cuales varían de 100 a 4000.

### 3.4.5 Análisis de los datos

En la investigación desarrollada se evalúan algoritmos basados en características y apariencia para construir el modelo de clasificación por medio de la tasa de acierto obtenida en la clasificación de género. La tasa de acierto se calcula en base a las imágenes correctamente clasificadas respecto al total de imágenes evaluadas. Para la validación de las caras correctamente clasificadas, se utiliza en el primero y tercer estudio la estrategia denominada “*cross-validation*”, en cambio en el segundo, se utiliza LOPO. Ambas estrategias permite reducir la variabilidad de los resultados, debido a que se realizan múltiples rondas para diferentes particiones de la base de datos, para entrenar y validar el clasificador. La tasa de acierto obtenida en cada ronda se promedia obteniendo la tasa de acierto final. La ecuación para calcular la tasa de acierto del clasificador se define como sigue:

$$Tasa\ de\ acierto = \frac{T_h + T_m}{H + M}, \quad (12)$$

donde  $T_h + T_m$  corresponden a la cantidad de hombre y mujeres clasificados correctamente, y  $H + M$  es el total de individuos.

A partir de los resultados, se generan una serie de tablas estratificadas por algoritmo, género y/o pose, y con éstas es posible analizar eficientemente los resultados y obtener una serie de indicadores que permiten medir la tasa de acierto de la clasificación, analizar la presencia de sesgo en la muestra, identificar la clase dominante, entre otras cosas.

## **IV. RESULTADOS**

En esta sección se presentan los resultados obtenidos por los tres estudios desarrollados en esta tesis. Cada uno de estos se describe en las tres primeras subsecciones respectivamente, y en la cuarta subsección se nombran las publicaciones aceptadas.

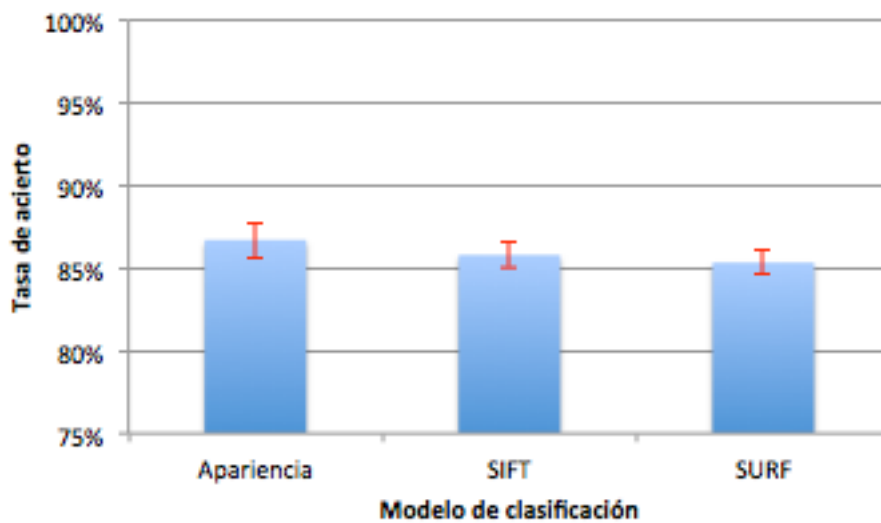
### **4.1 Análisis del poder discriminante de los modelos propuestos**

Los resultados del primer estudio, tabulados y graficados, se describen a continuación.

En la Tabla IV.1 se muestra en la primera columna el enfoque de clasificación utilizado, basado en apariencia y basado en características (SIFT y SURF). En la siguiente columna se muestra la cantidad de componentes principales utilizados para la clasificación, estos se calculan empíricamente como fue mencionado previamente. En el modelo de apariencia se reduce la dimensión del vector de 625 dimensiones a 230. En el caso de SIFT y SURF se reduce la dimensión de 1000 a 320 y 280 dimensiones respectivamente. La última columna muestra la tasa de acierto global en la clasificación junto con la desviación estándar para cada modelo. De los resultados obtenidos, se puede deducir que los tres modelos presentan rendimientos similares, donde el modelo de apariencia es marginalmente superior. La representación visual de esta tabla se presenta en la Figura IV.1, en la que aparece con azul las tasas de acierto media y con rojo los valores en que estas fluctúan, los cuales son determinados por la desviación estándar.

**Tabla IV.1** Tasa de acierto global obtenida en la detección del género para cada modelo propuesto.

	PCA	Tasa de acierto
<b>Apariencia</b>	230	86,70% ± 1,09%
<b>SIFT</b>	320	85,79% ± 0,76%
<b>SURF</b>	280	85,33% ± 0,71%

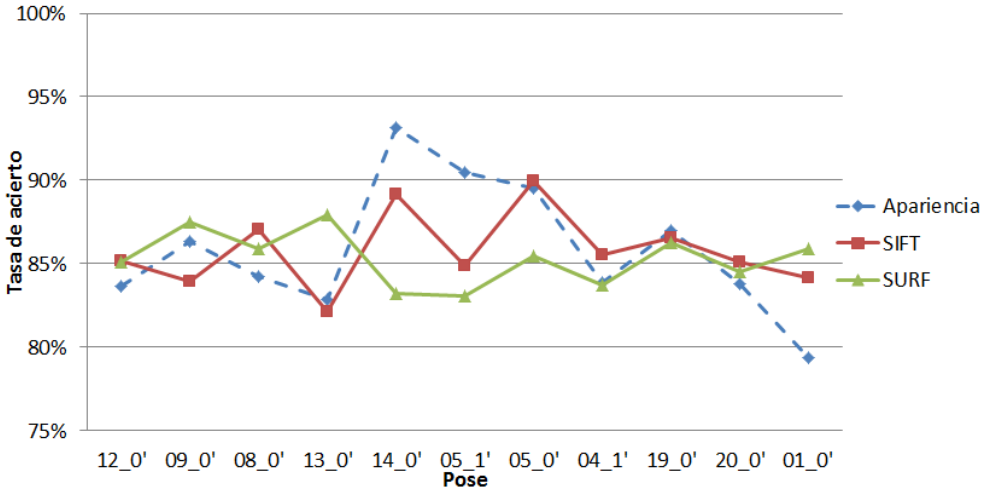


**Figura IV.1** Gráfico de barras del porcentaje de tasa de acierto en la detección del género para cada modelo propuesto.

Para un análisis más profundo de los datos, se estratifican los resultados obtenidos para detectar la influencia de cada una de las aproximaciones en el rendimiento de las salidas con respecto a las variables clases estudiadas (género y pose). En la Tabla IV.2 se muestra el rendimiento de los modelos de clasificación distribuidos por pose de los individuos. La pose 05\_1 corresponde al rostro completamente de frente a la cámara. Desde la pose 05\_1 a la 12\_0 representan la rotación del rostro hacia la izquierda hasta completar un ángulo de alrededor de 180°. De manera similar desde la pose 05\_1 a la 01\_0 pero con una rotación del rostro hacia la derecha. Los resultados obtenidos

confirman el buen rendimiento que tienen los modelos basados en apariencia cuando son clasificadas las caras frontales, obteniendo en el mejor de los casos un 93% de tasa de acierto. De la misma manera, se observa que los modelos basados en características mantienen un comportamiento estable con la rotación del rostro. Para cambios bruscos de la pose de la cara, la aproximación basada en característica obtiene los mejores resultados con una tasa de acierto de un 87,5%.

En la Figura IV.2 se grafican los resultados mostrados en la Tabla IV.2. En ésta se puede observar el comportamiento de cada modelo de clasificación de género en diferentes condiciones de orientación de la cara o pose. Con ello, se distingue fácilmente los modelos de clasificación que son más adecuados para detectar el género del individuo para una pose en específico.



**Figura IV.2** Gráfico de dispersión de tasa de acierto vs pose.

**Tabla IV.2** Tasa de acierto distribuida por pose y modelo de clasificación.

Pose	Apariencia	SIFT	SURF
12_0'	83,62% ± 7,20%	<b>85,15% ± 2,28%</b>	85,09% ± 1,54%
09_0'	86,36% ± 7,08%	83,91% ± 3,13%	<b>87,46% ± 4,30%</b>
08_0'	84,26% ± 4,99%	<b>87,05% ± 7,23%</b>	85,91% ± 5,77%
13_0'	82,84% ± 6,47%	82,11% ± 2,92%	<b>87,9% ± 4,30%</b>
14_0'	<b>93,17% ± 1,82%</b>	89,16% ± 2,72%	83,19% ± 3,55%
05_1'	<b>90,47% ± 5,10%</b>	84,86% ± 3,39%	83,05% ± 6,32%
05_0'	89,55% ± 6,06%	<b>89,96% ± 3,18%</b>	85,49% ± 6,35%
04_1'	83,87% ± 5,93%	<b>85,51% ± 3,21%</b>	83,72% ± 4,84%
19_0'	<b>87,00% ± 2,61%</b>	86,57% ± 4,17%	86,25% ± 3,64%
20_0'	83,77% ± 3,21%	<b>85,07% ± 4,17%</b>	84,53% ± 3,70%
01_0'	79,36% ± 5,38%	84,15% ± 7,12%	<b>85,89% ± 2,48%</b>
General	86,70% ± 1,09%	85,79% ± 0,76%	85,33% ± 0,71%

Además de distribuir los rendimientos por poses, se opta por estratificarlos por género y así identificar de mejor manera como se comportan los modelos tanto para detectar hombres como para mujeres. Los resultados obtenidos se pueden ver en la Tabla IV.3, en ésta se evidencia que el desbalance de género que existe entre el número de hombres y mujeres contenidos en la base de datos, impacta en el rendimiento en la clasificación de género inclinando el modelo de clasificación a estimar mejor sobre la clase dominante.

**Tabla IV.3** Tasa de acierto distribuida por pose, género y modelo de clasificación.

Pose	Apariencia (%)		SIFT (%)		SURF (%)	
	Hombre	Mujer	Hombre	Mujer	Hombre	Mujer
12_0'	90,12 ± 4,96	68,61 ± 9,75	88,45 ± 4,85	74,42 ± 9,06	92,68 ± 2,7	60,75 ± 11,83
09_0'	94,67 ± 4,52	66,48 ± 20,67	93,11 ± 4,38	60,95 ± 5,22	97,66 ± 1,31	61,71 ± 12,47
08_0'	93,09 ± 3,22	71,61 ± 7	95,36 ± 4,90	67,71 ± 16,54	97,13 ± 2,86	60,18 ± 16,56
13_0'	95,17 ± 2,73	56,35 ± 20,3	92,29 ± 6,04	59,55 ± 11,03	94,56 ± 2,56	73,07 ± 8,89
14_0'	93,04 ± 1,7	90,79 ± 7,35	92,5 ± 3,25	81,65 ± 8,57	89,66 ± 5,87	68,54 ± 9,95
05_1'	92,08 ± 6,7	89,36 ± 6,42	89,18 ± 4,76	75,07 ± 8,1	89,53 ± 6,85	68,21 ± 8,25
05_0'	94,25 ± 4,49	80,13 ± 5,1	95,97 ± 2,54	76,11 ± 15,79	91,93 ± 3,68	70,77 ± 16,03
04_1'	97,46 ± 2,69	53,89 ± 16,1	92,53 ± 4,26	70,64 ± 10,29	86,28 ± 7,17	78,42 ± 8,24
19_0'	93,05 ± 4,39	68,54 ± 6,89	98,25 ± 1,6	58,97 ± 9,68	96,6 ± 2,31	61,6 ± 7,89
20_0'	93,62 ± 4,39	67,56 ± 11,12	95,31 ± 4,44	59,67 ± 9,18	95,38 ± 1,55	56,79 ± 12,47
01_0'	91,44 ± 4,04	48,06 ± 10,56	93,3 ± 6,91	52,83 ± 11,11	93,92 ± 3,01	57,89 ± 19,77
General	<b>93,46 ± 1,47</b>	<b>70,13 ± 4,52</b>	<b>93,30 ± 0,63</b>	<b>67,37 ± 3,68</b>	<b>93,24 ± 2,14</b>	<b>65,92 ± 3,20</b>

## 4.2 Corroborar el poder discriminante minimizando el sesgo

En esta sección se muestran los resultados experimentales del segundo estudio, que tiene como finalidad corroborar el poder discriminante, de los vectores de características, evidenciado en la Tabla IV.I. Para ello, se utiliza una estrategia distinta de entrenamiento y validación denominada LOPO, que a diferencia de *cross-validation 5-fold*, permite disminuir el sesgo producido al tener imágenes repetidas de un mismo sujeto en la base de datos.

En la Tabla IV.4, al igual que la Tabla IV.1, se muestra el número de componentes principales utilizados para cada modelo de clasificación y la tasa de acierto obtenida en el mejor de los casos.

**Tabla IV.4** Tasa de acierto global obtenida por cada modelo de clasificación utilizando la estrategia LOPO.

	PCA	Tasa de acierto
<b>Apariencia</b>	40	85,34% ± 26,64%
<b>SIFT</b>	210	84,67% ± 24,08%
<b>SURF</b>	140	85,10% ± 24,30%
<b>ORB</b>	130	78,09% ± 28,86%

En la Tabla IV.5 se muestra los resultados obtenidos distribuidos por pose utilizando la estrategia de entrenamiento y validación LOPO (*Leave One Person Out*). En negrilla se marcan la mejor tasa de acierto obtenida para cada pose. Se mantienen la misma tendencia identificada del estudio anterior, en que los modelos basados en característica, con la excepción de ORB, presentan mejor resultado en variaciones de poses, y los basados en apariencia con las caras frontales. Cabe destacar, que cada tasa de acierto está acompañada de una alta desviación estándar, esto se debe principalmente a que al utilizar la estrategia LOPO el clasificador deberá comprobar en cada iteración sobre todas las variaciones de poses de un mismo individuo, en algunos casos no se clasifica bien al individuo en ninguna de sus poses obteniendo una tasa de acierto del 0%, en otros casos se clasifica bien en todas las variaciones de poses obteniendo el 100%, y también hay caso que se clasifica bien en algunas de las poses.

**Tabla IV.5** Tasa de acierto distribuida por pose y modelo de clasificación (LOPO).

Pose	APP LOPO		SIFT LOPO		SURF LOPO		ORB LOPO	
12_0'	83,26%	± 37,42%	<b>89,30%</b>	<b>± 30,98%</b>	87,44%	± 33,22%	76,74%	± 42,35%
09_0'	<b>84,85%</b>	<b>± 35,78%</b>	83,40%	± 37,28%	83,40%	± 37,28%	78,01%	± 41,51%
08_0'	85,08%	± 35,70%	<b>86,69%</b>	<b>± 34,03%</b>	85,48%	± 35,30%	79,03%	± 40,79%
13_0'	82,92%	± 37,71%	83,33%	± 37,35%	<b>83,75%</b>	<b>± 36,97%</b>	78,75%	± 40,99%
14_0'	<b>89,16%</b>	<b>± 31,15%</b>	86,35%	± 34,41%	81,53%	± 38,89%	82,33%	± 38,22%
05_1'	<b>88,35%</b>	<b>± 32,14%</b>	86,35%	± 35,24%	85,54%	± 35,24%	79,12%	± 40,73%
05_0'	<b>89,56%</b>	<b>± 30,64%</b>	86,35%	± 34,41%	85,94%	± 34,41%	78,31%	± 41,29%
04_1'	82,91%	± 37,73%	83,76%	± 36,96%	<b>84,19%</b>	<b>± 36,56%</b>	78,63%	± 41,08%
19_0'	83,74%	± 36,98%	82,11%	± 38,40%	<b>85,37%</b>	<b>± 35,42%</b>	80,08%	± 40,02%
20_0'	85,83%	± 34,94%	83,33%	± 37,35%	<b>88,75%</b>	<b>± 31,66%</b>	77,92%	± 41,57%
01_0'	82,63%	± 37,98%	83,57%	± 37,14%	<b>86,85%</b>	<b>± 33,87%</b>	74,18%	± 43,87%
<b>General</b>	<b>85,34%</b>	<b>± 26,64%</b>	<b>84,67%</b>	<b>± 24,08%</b>	<b>85,10%</b>	<b>± 24,30%</b>	<b>78,09%</b>	<b>± 28,86%</b>

La Tabla IV.6 muestra los resultados distribuidos por pose y género, manteniendo un buen rendimiento en general hacia la clase dominante. Respecto a la clasificación sobre el género femenino, se evidencian rendimientos en algunas poses inferiores al 50%. Sin embargo, la relación entre hombres y mujeres de la muestra utilizada es de alrededor de un 70% y 30% respectiva, por lo tanto, el rendimiento obtenido es superior que haber clasificado al azar. De todas maneras un rendimiento por debajo del 50% no es un buen resultado pero, como fue mencionado previamente, es mejor que clasificar aleatoriamente. En negrilla se resaltan los rendimientos más altos y más bajos para cada modelo de clasificación. En algunos casos, como en la pose 19\_0 para SIFT y la pose 05\_1 para apariencia, un rendimiento alto en género masculino implica un rendimiento bajo en el género femenino. Esto no se puede generalizar ya que no sucede en todas las variaciones de pose. En general, ambos algoritmos detectaron mejor el género femenino cuando la pose era frontal.

**Tabla IV.6** Tasa de acierto distribuida por pose, género y modelo de clasificación.  
(LOPO)

Pose	Apariencia LOPO (%)		SIFT LOPO (%)		SURF LOPO (%)	
	Hombre	Mujer	Hombre	Mujer	Hombre	Mujer
12_0'	92,68 ± 26,12	52,94 ± 50,41	92,07 ± 27,10	<b>80,39 ± 40,1</b>	97,56 ± 15,47	54,90 ± 50,25
09_0'	<b>98,84 ± 10,72</b>	49,26 ± 49,99	97,11 ± 16,80	48,53 ± 50,35	97,69 ± 15,07	<b>47,06 ± 50,28</b>
08_0'	95,95 ± 19,76	60,00 ± 49,32	<b>97,69 ± 15,07</b>	61,33 ± 49,03	95,95 ± 19,76	61,33 ± 49,03
13_0'	98,19 ± 13,36	48,65 ± 50,32	95,18 ± 21,48	56,76 ± 49,88	95,78 ± 20,16	56,76 ± 49,88
14_0'	94,22 ± 23,40	77,63 ± 41,95	95,95 ± 19,76	64,47 ± 48,18	<b>88,44 ± 32,07</b>	65,79 ± 47,76
05_1'	<b>90,75 ± 29,06</b>	<b>82,89 ± 37,91</b>	<b>89,02 ± 31,36</b>	77,63 ± 41,95	93,64 ± 24,47	67,11 ± 47,30
05_0'	93,64 ± 24,47	80,26 ± 40,07	93,64 ± 24,47	69,74 ± 46,24	95,38 ± 21,06	64,47 ± 48,18
04_1'	98,14 ± 13,56	49,32 ± 50,34	95,03 ± 21,80	58,90 ± 49,54	89,44 ± 30,83	<b>72,60 ± 44,91</b>
19_0'	94,22 ± 23,40	58,90 ± 49,54	<b>97,69 ± 15,07</b>	<b>45,21 ± 50,11</b>	98,27 ± 13,09	54,79 ± 50,11
20_0'	97,67 ± 15,12	55,88 ± 50,02	95,93 ± 19,82	51,47 ± 50,35	<b>100,00 ± 0,00</b>	60,29 ± 49,29
01_0'	93,29 ± 25,09	<b>46,94 ± 50,42</b>	92,68 ± 26,12	53,06 ± 50,42	94,51 ± 22,84	61,22 ± 49,23
General	95,06 ± 11,05	63,21 ± 36,79	94,77 ± 7,98	61,68 ± 31,62	95,16 ± 8,73	62,20 ± 31,83

Por último, en la Tabla IV.7 se muestra el rendimiento general obtenido por los clasificadores utilizando distintas estrategias de entrenamiento y validación. Al utilizar LOPO se evidenció una disminución en la clasificación con relación al estudio anterior de alrededor del 1%, lo cual en términos de clasificación no es significativo. Sin embargo al comparar el rendimiento de la Tabla IV.2 y Tabla IV.4 claramente hay diferencias, donde lo obtenido en la Tabla IV.4 es una medida más realista de la tasa de acierto en la clasificación para cada pose.

**Tabla IV.7** Tasa de acierto usando las estrategias *cross-validation* y LOPO

	Tasa de acierto	
	cross-validation	LOPO
Apariencia	86,70% ± 1,09%	85,34% ± 26,64%
SIFT	85,79% ± 0,76%	84,67% ± 24,08%
SURF	85,33% ± 0,71%	85,10% ± 24,30%

### 4.3 Evaluar el poder discriminante sobre caras en entornos sin controlar

Como fue mencionado en la sección 3.4.1, para realizar este último experimento se selecciona la base de datos con imágenes de caras obtenidas en condiciones sin controlar denominada GROUPS. Para el entrenamiento y validación del clasificador se utilizó la estrategia *cross-validation* 5 fold, con lo cual en cada iteración se divide la muestra en un 80% para entrenamiento y el 20% restante para prueba. El promedio de imágenes del conjunto de prueba distribuidas por pose y el porcentaje de hombres y mujeres respectivo, se muestra en la Tabla IV.8. Además se observa que la muestra se encuentra balanceada, teniendo el género femenino una cantidad de imágenes levemente superior.

**Tabla IV.8** Distribución de imágenes por género en cada pose del conjunto de prueba.

Pose	Hombre	Mujer	Total
0	68,99%	31,01%	158
1	50,36%	49,64%	687
2	36,81%	63,19%	470
3	51,23%	48,77%	406
4	49,74%	50,26%	390
5	39,19%	60,81%	296
	<b>1146</b>	<b>1261</b>	<b>2407</b>

En la Tabla IV.9 se puede ver la tasa de acierto que se obtiene con los modelos basados en características al utilizar una cantidad diferente de *clusters* para la construcción del vector de frecuencia. A simple vista se evidencia que a medida que se aumenta el número de los *clusters* es mejor la tasa de acierto que se obtiene. Sin embargo, la mejora obtenida utilizando 3000 o 4000 clúster es marginal (alrededor de un 0,2%), lo cual sugiere que se ha llegado a la máxima tasa de acierto que provee los modelos propuestos sobre esta base de datos. Cabe resaltar que SIFT es el detector y descriptor con el cual se consiguen mejores resultados. El bajo rendimiento de SURF,

respecto a SIFT, se debe en gran parte a que este algoritmo identifica pocos puntos de interés en imágenes pequeñas, lo cual impide que se puedan obtener mejores características distintivas entre ambos géneros.

**Tabla IV.9** Tasa de acierto obtenida con modelos basados en características respecto a la cantidad de clúster utilizados.

Nº Cluster	ACC			
	SIFT		SURF	
100	67,28%	± 0,47%	65,61%	± 1,25%
500	74,89%	± 1,24%	70,50%	± 0,55%
1000	76,79%	± 2,59%	71,99%	± 0,84%
1500	78,23%	± 0,97%	72,79%	± 0,53%
2000	78,55%	± 0,62%	73,42%	± 0,86%
3000	80,78%	± 0,78%	73,63%	± 0,91%
4000	80,93%	± 0,55%	74,30%	± 0,65%

Debido a que los mejores resultados, en términos de tasa de acierto y tiempo de ejecución, se obtienen en modelos basados en características utilizando 3000 *clusters*, los demás resultados que se muestran consideran esta configuración del modelo de clasificación.

En la Tabla IV.10 se muestran los resultados obtenidos al utilizar el algoritmo SIFT para obtener los puntos característicos y formar un vector de frecuencia de 3000 dimensiones. En la primera columna se muestra el número de *fold* correspondiente al proceso de la validación cruzada, en la segunda el número de componentes principales utilizado para simplificar el tamaño del vector de frecuencia dejando las componente que maximizan la variabilidad entre las clases, y en las siguientes columnas las tasas de acierto estratificadas por género y el promedio de éstas. Dado que la base de datos GROUPS se encuentra balanceada, la tasa tanto en hombres como en mujeres es similar, donde en promedio se obtiene un 80,78% con una desviación estándar de 0,74%. Con relación a la Tabla IV.11, se presentan los mismos datos que en la

tabla anterior pero utilizando el algoritmo SURF. En este caso, la tasa de acierto es menor, obteniendo 73,63% con una desviación estándar de 0,91%, donde se evidencia que la clasificación se inclina al género femenino obteniendo alrededor de 6% mejor clasificación que sobre el género masculino.

**Tabla IV.10** Tasa de acierto obtenida con el modelo de clasificación basado en SIFT en cada *fold* de la validación cruzada.

SIFT (3000)				
Fold	PCA	Hombre	Mujer	Global
1	440	79,86%	82,63%	81,31%
2	1010	81,26%	80,89%	81,06%
3	820	79,76%	81,60%	80,72%
4	630	80,56%	78,57%	79,52%
5	1460	81,69%	80,87%	81,26%
		<b>80,62%</b>	<b>80,91%</b>	<b>80,78%</b>

**Tabla IV.11** Tasa de acierto obtenida con el modelo de clasificación basado en SIFT en cada *fold* de la validación cruzada.

SURF (3000)				
Fold	PCA	Hombre	Mujer	Global
1	1150	69,40%	79,30%	74,58%
2	760	71,32%	75,42%	73,46%
3	660	70,33%	78,43%	74,57%
4	970	71,93%	73,33%	72,66%
5	830	70,88%	74,68%	72,87%
		<b>70,77%</b>	<b>76,23%</b>	<b>73,63%</b>

Las siguientes tablas (de la IV.12 a la IV.4), muestran los resultados obtenidos en la clasificación de género al utilizar los atributos *poweraset*. En la Tabla IV.12, se observa que la máxima tasa de acierto que se obtiene en la clasificación se logra al utilizar 4000 clúster. Sin embargo, al igual que en la Tabla IV.9, la diferencia es marginal respecto a la obtenida con 3000 clúster, por lo cual no se justifica utilizar ese tamaño dado que el costo en recursos

computacionales es bastante mayor. Por otro lado, al clasificar con atributos *powerset* no se obtiene una mejora significativa que al utilizar sólo el género, más bien los resultados son bastante similares y en algunos caso levemente inferiores.

**Tabla IV.12** Tasa de acierto obtenida con modelos basados en características respecto a la cantidad de clúster utilizados.

Nº Cluster	POWERSET			
	SIFT		SURF	
100	67,22%	± 0,36%	65,44%	± 0,86%
500	75,01%	± 1,35%	70,67%	± 0,72%
1000	76,84%	± 0,98%	72,29%	± 0,76%
1500	77,99%	± 0,76%	72,84%	± 0,51%
2000	78,78%	± 0,93%	73,32%	± 0,94%
3000	80,68%	± 0,78%	73,67%	± 0,90%
4000	80,89%	± 0,62%	74,23%	± 0,87%

Las Tablas IV.13 y IV.14, muestran los resultados estratificados por género obtenidos al utilizar 3000 clúster tanto para SIFT como para SURF. Con SIFT se logra un 80,68% con una desviación estándar de 0,78%, en donde se observa una diferencia de un 5% mejor en la clasificación para el género masculino. Respecto a SURF, se logra 73,63% con una desviación estándar de 0,91%, en donde se obtiene similar tasa de acierto en ambos géneros.

**Tabla IV.13** Tasa de acierto obtenida con el modelo de clasificación basado en SIFT en cada *fold* de la validación cruzada.

Fold	PCA	SIFT (POWERSET)		
		Hombre	Mujer	Global
1	590	83,35%	80,10%	81,64%
2	1010	83,70%	77,80%	80,61%
3	820	81,59%	79,38%	80,43%
4	870	83,35%	76,11%	79,56%
5	1460	83,26%	79,21%	81,14%
		<b>83,05%</b>	<b>78,52%</b>	<b>80,68%</b>

**Tabla IV.14** Tasa de acierto obtenida con el modelo de clasificación basado en SURF en cada *fold* de la validación cruzada.

SURF (POWERSET)				
Fold	PCA	Hombre	Mujer	Global
1	1290	72,54%	75,57%	74,13%
2	1010	75,07%	72,32%	73,63%
3	690	73,91%	75,89%	74,95%
4	1000	76,37%	69,68%	72,87%
5	430	74,19%	71,51%	72,79%
		<b>74,42%</b>	<b>73,00%</b>	<b>73,67%</b>

En la Tabla IV.15, se muestran los resultados estratificados por género obtenidos al clasificar utilizando un modelo basado en apariencia. La tasa de acierto global es de un 80,93% con una desviación estándar de 0,99%, en donde se obtiene una mejor clasificación en mujeres, alrededor de un 5% más, que en hombres. Similares resultados se muestran en la Tabla IV.16, en donde se obtiene una mejora global de alrededor de un 0,6% al utilizar atributos *powerset*, manteniendo una mejor clasificación en mujeres

**Tabla IV.15** Tasa de acierto obtenida con el modelo de clasificación basado en apariencia en cada *fold* de la validación cruzada.

Apariencia				
Fold	PCA	Hombre	Mujer	Global
1	250	78,81%	83,90%	81,48%
2	230	76,55%	83,11%	79,98%
3	300	81,33%	82,24%	81,80%
4	280	75,76%	83,33%	79,73%
5	260	79,69%	83,41%	81,64%
		<b>78,43%</b>	<b>83,20%</b>	<b>80,93%</b>

**Tabla IV.16** Tasa de acierto obtenida con el modelo de clasificación basado en apariencia en cada *fold* de la validación cruzada al utilizar atributos *poweraset*.

Apariencia (POWERSET)				
Fold	PCA	Hombre	Mujer	Global
1	470	80,30%	84,46%	82,48%
2	230	77,42%	83,35%	80,52%
3	260	81,68%	82,71%	82,22%
4	300	76,46%	83,89%	80,35%
5	340	79,25%	84,37%	81,93%
		<b>79,02%</b>	<b>83,75%</b>	<b>81,50%</b>

Los resultados obtenidos estratificados por pose, se observan las Tablas IV.17 y IV.18, en las cuales se destacan en negrilla los tasas de acierto más altas en la clasificación de género obtenidas en cada pose. Al utilizar la forma tradicional para clasificar género, el modelo basado en características utilizando el descriptor SIFT, presentan los mejores resultados logrando tener un mejor rendimiento sobre 4 poses. En cambio, al utilizar *poweraset*, el modelo basado en apariencia logra un mejor rendimiento.

**Tabla IV.17** Tasa de acierto global de los modelos de clasificación evaluados estratificados por pose.

Modelo	Pose					
	0	1	2	3	4	5
Apariencia	82,95%	77,71%	<b>86,20%</b>	<b>78,30%</b>	84,68%	77,62%
SIFT (3000)	<b>84,22%</b>	<b>77,88%</b>	82,88%	78,20%	<b>86,43%</b>	<b>78,42%</b>
SURF (3000)	77,09%	70,93%	76,83%	71,41%	77,92%	70,39%

**Tabla IV.18** Tasa de acierto global de los modelos de clasificación evaluados estratificados por pose utilizando atributo *poweraset*.

Modelo	Pose					
	0	1	2	3	4	5
Apariencia	84,73%	<b>78,84%</b>	<b>84,67%</b>	<b>80,02%</b>	84,22%	<b>79,37%</b>
SIFT (3000)	<b>84,74%</b>	78,81%	83,01%	77,75%	<b>85,30%</b>	77,08%
SURF (3000)	75,58%	71,28%	76,49%	72,00%	79,00%	69,05%

Finalmente, la Tabla IV.17 muestra los resultados globales obtenidos por los tres modelos de clasificación utilizados. Se puede apreciar que el modelo basado en apariencia y el modelo basado en características extraídas con SIFT obtienen similares resultados en cuanto a tasa de acierto, en caso contrario lo que sucede con SURF el cual obtienen un poco más de 7% de diferencia con los otros.

**Tabla IV.19** Tasa de acierto global con los modelos de clasificación evaluados al clasificar género de la manera tradicional y al utilizar atributos *powerset*.

Modelo	Tasa de acierto	
	Género	Powerset
Apariencia	80,93% ± 0,99%	81,50% ± 0,99%
SIFT (3000)	80,78% ± 0,74%	80,68% ± 0,78%
SURF (3000)	73,63% ± 0,91%	73,67% ± 0,90%

#### **4.4 Publicaciones**

Cada uno de los estudios realizados en la etapa experimental tiene como finalidad la generación de un artículo con el objetivo de difundir los resultados. A continuación se muestran donde han sido difundidos los resultados de los estudios de manera respectiva, quedando pendiente el artículo del tercer estudio:

1. Artículo aceptado en el V Congreso Internacional de Computación e Informática del Norte de Chile, INFONOR CHILE 2014, denominada “Clasificación de género utilizando vectores de frecuencia basados en descriptores locales”.
2. Artículo aceptado para publicación en el Volumen 24 N° 1, enero - marzo 2016, de Ingeniare. Revista chilena de ingeniería, ISSN 0718-3305 versión en línea.

## V. CONCLUSIONES

En esta tesis se ha demostrado empíricamente que el modelo de clasificación basado en vectores de frecuencia, construidos a partir de características locales, por medio de SIFT y SURF, entrega información discriminante y obtiene resultados similares a los modelos basados en apariencia. No se puede decir lo mismo con relación a ORB para la construcción de características locales, dado que con éste se evidenció un bajo desempeño respecto a la tasa de acierto, logrando alrededor de un 7% menos, sobre la base de datos MULTI-PIE, que los otros tres modelos de clasificación.

Una de las estrategias más utilizadas para la validación de clasificadores es *cross-validation*. Ésta permite medir el rendimiento del clasificador sobre diferentes grupos de muestras, entregando como resultado una evaluación razonable del modelo. En el caso de la estimación de género por medio de imágenes, se estudió que la mejor forma de validar los modelos de clasificación es utilizando individuos diferentes en el conjunto de entrenamiento y de prueba lo que evita cualquier sesgo que se pueda introducir al modelo durante su entrenamiento. En esta tesis se observa que al evaluar el método de clasificación propuesto bajo dos técnicas diferentes de validación (5-Fold y LOPO), basadas en *cross-validation*, se obtiene una disminución del rendimiento al utilizar LOPO. Sin embargo, al evaluar el mismo clasificador utilizando LOPO, se evidencia una baja en la tasa de acierto obtenida sobre el conjunto de datos MULTI-PIE. Si bien es cierto, esta baja en rendimiento es marginal, ésta última aproximación es más significativa ya que siempre clasifica rostros de individuos diferentes. No obstante, LOPO es una estrategia que consume mucho tiempo de cómputo. Para los experimentos realizados se utilizaron 249 individuos lo que implica que se deben entrenar y probar la misma cantidad de clasificadores. Por

lo tanto, se recomienda utilizar estrategia *cross-validation* en el caso de no tener rostros repetidos en la base de datos, en caso contrario es preferible utilizar LOPO siempre y cuando no existan complicaciones con el tiempo que involucra la realización de los experimentos.

Por otro lado, al analizar los resultados obtenidos en el segundo estudio, tabulados en la Tabla IV.5, se corrobora que los modelos de apariencia funcionan bastante bien para clasificar individuos en condiciones controladas, específicamente con caras frontales (14\_0, 05\_1, 05\_0). En el caso de querer clasificar rostros de individuos adquiridos en condiciones sin restricciones, son preferibles los modelos basados en características ya que su comportamiento se mantiene uniforme en las variaciones de poses.

También se observó en los resultados experimentales, sobre la base de datos MULTI-PIE, que se obtuvo una mejor tasa de acierto en hombres que en mujeres. Esto debe principalmente al desbalance entre las imágenes de hombres y mujeres que contiene la muestra utilizada de esta base de datos. Como consecuencia, los clasificadores entrenados tienden a ajustar el modelo a favor de las clases dominantes para obtener una mejor tasa de acierto global. Esto queda en evidencia al revisar los resultados obtenidos sobre la base de datos GROUPS, la cual al tener balanceado el género, la tasa de acierto es similar para ambas clases. Específicamente para MULTI-PIE, se decidió utilizar la misma muestra que en Bekios-Calfa et al. (2014), la cual estaba desbalanceada en género, para poder contrastar los resultados obtenidos con esta publicación. Sin embargo, se recomienda para futuras investigaciones equilibrar la cantidad de hombres y mujeres en la muestra para evitar que la clasificación tienda a ser mejor para la clase dominante.

Con relación a los resultados obtenidos con GROUPS, se logran tasas de acierto similares a la línea base de la clasificación de género sobre esta base de datos, y por ende no se obtiene una mejora significativa.

Finalmente, respecto a la manera en que se realiza la clasificación de género, utilizando un clasificador unidimensional o considerando múltiples variables clases, ambos enfoques probados evidenciaron una tasa de acierto similar en los modelos de clasificación basado en característica. No obstante, se obtiene una leve mejora en la tasa de acierto, alrededor de un 1%, al utilizar el modelo basado en apariencia con el enfoque de múltiples variables clases.

## 5.1 Trabajo futuro.

Para mejorar el rendimiento del método de clasificación de género desarrollado considerar los siguientes aspectos:

- Ajuste de los parámetros: Se propone ampliar la búsqueda de una configuración adecuada de la cantidad de *clusters* a utilizar y de los parámetros de los descriptores locales SIFT y SURF. Para la cantidad de *clusters*, se debe considerar que el ajuste de parámetros mejore la tasa de acierto de la clasificación sin perjudicar considerablemente el tiempo de respuesta del método. Respecto a SIFT y SURF, una posible mejora sería ajustar los parámetros dependiendo del tamaño de la imagen objetiva, ya que al utilizar siempre la misma configuración es posible que en imágenes muy pequeñas no se detecten descriptores o en imágenes muy grandes los descriptores encontrados no sean los que mejor describan.
- Selección de *features*: Luego de la etapa de extracción de *features*, se propone considerar una fase de filtrado y/o selección de *features*, en la cual se obtengan las más representativas para ambos géneros (hombre y mujer) previo a la construcción del vector de frecuencia. La finalidad es disminuir la presencia de ruido que pudiese estar presente en la imagen, tales como: *features* que no correspondan al rostro del individuo, que no sean discriminantes (aparezcan con similar frecuencia en hombres y mujeres), entre otras.
- Incorporación de la información geométrica: Otro aspecto a considerar es la información geométrica de las *features*, o en términos simples la posición en que éstas son encontradas. La posición puede ser incorporada directamente como parte del vector de características, o también puede ser usada para filtrar los *clusters*, dejando aquellos que contengan las *features* concentradas en alguna ubicación del rostro.

Por otro lado, dado el buen desempeño que han obtenido los métodos de clasificación de género que utilizan fusión de clasificadores (Castrillón-Santana, Modesto Lorenzo-Navarro et al., 2013)(Mirza et al., 2013), se propone realizar la fusión sobre los modelos desarrollados y evaluados en ésta tesis, considerando los basados en apariencia para clasificar caras frontales y los basados en características para variaciones de pose.

Finalmente, se propone evaluar la robustez que otorgan los modelos presentados sobre diversas condiciones que pudiesen ser tomadas las imágenes, ya que, al utilizar sólo algunos puntos de interés y no toda la imagen, es probable que se vean afectado levemente a oclusiones o diversos ruidos presentes en la adquisición. La robustez de los clasificadores basados en características locales fue evaluado en Toews & Arbel (2009) , quienes, dentro de sus experimentos, midieron el rendimiento del clasificador sobre imágenes con diferente grado de oclusión y mostraron que el error en la clasificación aumenta suavemente con el aumento del grado de oclusión, demostrando la capacidad de adaptación de éstos clasificadores a las características que faltan. Debido a esto, se cree que los modelos propuestos también deberían presentar un buen comportamiento frente a estas condiciones, lo cual es sumamente relevante si se desea reconocer el género de individuos en entornos sin restricciones.

## BIBLIOGRAFÍA

- Aksoy, S. (2012). Introduction to Pattern Recognition. Retrieved from [http://www.cs.bilkent.edu.tr/~saksoy/courses/cs551/slides/cs551\\_intro.pdf](http://www.cs.bilkent.edu.tr/~saksoy/courses/cs551/slides/cs551_intro.pdf)
- Al-Azzawy, D. S., & Al-Azzawy, S. (2012). Eigenface and SIFT For Gender Classification. *Journal Of Wassit For Science & Medicine*, 5(1), 60–76. Retrieved from <http://www.iasj.net/iasj?func=fulltext&ald=51973>
- AX3. (2013). Touch screen vending machines?! Retrieved January 20, 2014, from <http://ax3battery.com/2013/04/15/touch-screen-vending-machines-2/>
- Baluja, S., & Rowley, H. a. (2006). Boosting Sex Identification Performance. *International Journal of Computer Vision*, 71(1), 111–119. doi:10.1007/s11263-006-8910-9
- Bay, H., Ess, A., Tuytelaars, T., & Van Gool, L. (2008). Speeded-Up Robust Features (SURF). *Computer Vision and Image Understanding*, 110(3), 346–359. doi:10.1016/j.cviu.2007.09.014
- Bay, H., Tuytelaars, T., & Gool, L. (2006). SURF: Speeded Up Robust Features. In *Computer Vision–ECCV 2006* (Vol. 3951, pp. 404–417). doi:10.1007/11744023\_32
- Bekios-Calfa, J., Buenaposada, J. M., & Baumela, L. (2011). Revisiting linear discriminant techniques in gender recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(4), 858–64. doi:10.1109/TPAMI.2010.208
- Bekios-Calfa, J., Buenaposada, J. M., & Baumela, L. (2014). Robust gender recognition by exploiting facial attributes dependencies. *Pattern Recognition Letters*, 36, 228–234. doi:10.1016/j.patrec.2013.04.028
- Buchala, S., Davey, N., J. Frank, R., & Loomes, M. (2005). The role of Global and Feature based information in gender classification of faces: A Comparisons of human performance and computational models. *Neural of System*, 15, 1–8.
- Calonder, M., Lepetit, V., Strecha, C., & Fua, P. (2010). BRIEF: Binary robust independent elementary features. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 6314 LNCS, pp. 778–792). doi:10.1007/978-3-642-15561-1\_56

- Castrillón-Santana, Modesto Lorenzo-Navarro, J., & Ramón-Balmaseda, E. (2013). Improving Gender Classification Accuracy in the Wild. In J. Ruiz-Shulcloper & G. Sanniti di Baja (Eds.), *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications (CIARP 2013., Vol. 8259, pp. 270–277)*. Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-642-41827-3
- Chen, D., Cao, X., Wen, F., & Sun, J. (2013). Blessing of Dimensionality: High-Dimensional Feature and Its Efficient Compression for Face Verification. *2013 IEEE Conference on Computer Vision and Pattern Recognition, 3025–3032*. doi:10.1109/CVPR.2013.389
- Chen, D.-Y., & Lin, K.-Y. (2010). Robust gender recognition for uncontrolled environment of real-life images. *IEEE Transactions on Consumer Electronics, 56(3)*, 1586–1592. doi:10.1109/TCE.2010.5606301
- Dago-Casas, P., Gonzalez-Jimenez, D., & Alba-Castro, J. L. (2011). Single- and cross- database benchmarks for gender classification under unconstrained settings. In *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)* (pp. 2152–2159). IEEE. doi:10.1109/ICCVW.2011.6130514
- Fukunaga, K. (1990). *Introduction to Statistical Pattern Recognition*. Academic Press.
- Gallagher, A. C., & Chen, T. (2009). Understanding images of groups of people. In *2009 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 256–263). IEEE. doi:10.1109/CVPR.2009.5206828
- Geng, C., & Jiang, X. (2009). Face recognition using sift features. In *Proceedings - International Conference on Image Processing, ICIP* (pp. 3313–3316). doi:10.1109/ICIP.2009.5413956
- Golomb, B. A., Lawrence, D. T., & Sejnowski, T. J. (1990). SexNet: A neural network identifies sex from human faces. In *NIPS* (pp. 572–577).
- Gross, R., Matthews, I., Cohn, J., Kanade, T., & Baker, S. (2008). Multi-PIE. In *2008 8th IEEE International Conference on Automatic Face & Gesture Recognition* (pp. 1–8). IEEE. doi:10.1109/AFGR.2008.4813399
- Hadid, A., & Pietikäinen, M. (2013). Demographic classification from face videos using manifold learning. *Neurocomputing, 100*, 197–205. doi:10.1016/j.neucom.2011.10.040

- Hernández Sampieri, R. (2010). *Metodología de la Investigación*. (J. M. Chacón, Ed.) (5ta ed.). McGraw-Hill.
- Huang, G. B., Ramesh, M., Berg, T., & Learned-miller, E. (2007). *Labeled Faces in the Wild : A Database for Studying Face Recognition in Unconstrained Environments*. Amherst.
- Jan Erik Solem. (2012). *Programming Computer Vision with Python*. *Programming Computer Vision with Python*. O'Reilly Media.
- Li, B., Lian, X.-C., & Lu, B.-L. (2012). Gender classification by combining clothing, hair and facial component classifiers. *Neurocomputing*, 76(1), 18–27. doi:10.1016/j.neucom.2011.01.028
- Lowe, D. G. (2004). Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2), 91–110. doi:10.1023/B:VISI.0000029664.99615.94
- Mäkinen, E., & Raisamo, R. (2008). An experimental comparison of gender classification methods. *Pattern Recognition Letters*, 29(10), 1544–1556. doi:10.1016/j.patrec.2008.03.016
- Martinez, A. M., & Kak, A. C. (2001). PCA versus LDA. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(2), 228–233. doi:10.1109/34.908974
- Minear, M., & Park, D. C. (2004). A lifespan database of adult facial stimuli. *Behavior Research Methods, Instruments, & Computers*, 36(4), 630–633. doi:10.3758/BF03206543
- Mirza, A. M., Hussain, M., Almuzaini, H., & Muhammad, G. (2013). Gender Recognition Using Fusion of Local and Global Facial Features. In G. Bebis, R. Boyle, B. Parvin, D. Koracin, B. Li, F. Porikli, ... D. Gotz (Eds.), *Advances in Visual Computing* (Vol. 8034, p. pp 493–502). Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-642-41939-3
- Moghaddam, B., & Yang, M.-H. Y. (2002). Learning gender with support faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5), 707–711. doi:10.1109/34.1000244
- NEC. (2014). Surveillance for Critical Infrastructure Protection. Retrieved from [http://www.nec.com/en/global/solutions/security/solutions/surveillance\\_for\\_cip.html](http://www.nec.com/en/global/solutions/security/solutions/surveillance_for_cip.html)

- Ng, C. B., Tay, Y. H., & Goi, B. (2012). Recognizing Human Gender in Computer Vision: A Survey. In P. Anthony, M. Ishizuka, & D. Lukose (Eds.), *PRICAI 2012: Trends in Artificial Intelligence* (Vol. 7458, pp. 335–346). Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-642-32695-0
- Phillips, P. J., Rizvi, S. A., & Rauss, P. J. (2000). The FERET evaluation methodology for face-recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10), 1090–1104. doi:10.1109/34.879790
- Phillips, P. J., Wechsler, H., Huang, J., & Rauss, P. J. (1998). The FERET database and evaluation procedure for face-recognition algorithms. *Image and Vision Computing*, 16(5), 295–306. doi:10.1016/S0262-8856(97)00070-X
- Ramón-Balmaseda, E., Lorenzo-navarro, J., & Castrillón-Santana, M. (2012). Gender Classification in Large Databases. In L. Alvarez, M. Mejail, L. Gomez, & J. Jacobo (Eds.), *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications* (Vol. 7441, pp. 74–81). Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-642-33275-3
- Ricanek, K., & Tesafaye, T. (2006). MORPH: A Longitudinal Image Database of Normal Adult Age-Progression. In *7th International Conference on Automatic Face and Gesture Recognition (FGR06)* (pp. 341–345). IEEE. doi:10.1109/FGR.2006.78
- Rosin, P. L. (1999). Measuring Corner Properties. *Computer Vision and Image Understanding*, 73(2), 291–307. doi:10.1006/cviu.1998.0719
- Rosten, E., & Drummond, T. (2006). Machine learning for high-speed corner detection. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 3951 LNCS, pp. 430–443). doi:10.1007/11744023\_34
- Ruble, E., Rabaud, V., Konolige, K., & Bradski, G. (2011). ORB: An efficient alternative to SIFT or SURF. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 2564–2571). doi:10.1109/ICCV.2011.6126544
- Salton, G., Wong, A., & Yang, C. S. (1975). A Vector Space Model for Automatic Indexing, 18(11).

- Shakhnarovich, G., Viola, P. a., & Moghaddam, B. (2002). A unified learning framework for real time face detection and classification. *Proceedings of Fifth IEEE International Conference on Automatic Face Gesture Recognition*, 16–23. doi:10.1109/AFGR.2002.1004124
- Shan, C. (2012). Learning local binary patterns for gender classification on real-world face images. *Pattern Recognition Letters*, 33(4), 431–437. doi:10.1016/j.patrec.2011.05.016
- Sun, Z., Bebis, G., Xiaojing, Y., & Louis, S. J. (2002). Genetic feature subset selection for gender classification: a comparison study. In *Sixth IEEE Workshop on Applications of Computer Vision, 2002. (WACV 2002). Proceedings.* (pp. 165–170). IEEE Comput. Soc. doi:10.1109/ACV.2002.1182176
- Theodoridis, S., & Koutroubas, K. (2008). *Pattern Recognition, Fourth Edition*. Press, Academic.
- Toews, M., & Arbel, T. (2009). Detection, localization, and sex classification of faces from arbitrary viewpoints and under occlusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(9), 1567–81. doi:10.1109/TPAMI.2008.233
- Ullah, I., Hussain, M., Muhammad, G., Aboalsamh, H., Bebis, G., & Mirza, A. M. (2012). GENDER RECOGNITION FROM FACE IMAGES WITH LOCAL WLD DESCRIPTOR, (April), 11–13.
- Viola, P., & Jones, M. J. (2001). *Robust Real-time Object Detection*. *International Journal of Computer Vision - IJCV*. doi:10.1.1.23.2751
- Viola, P., & Jones, M. J. (2004). Robust Real-Time Face Detection. *International Journal of Computer Vision*, 57(2), 137–154. doi:10.1023/B:VISI.0000013087.49260.fb
- Webb, A. R. (2002). *Statistical Pattern Recognition Second Ed*. John Wiley & Sons, Ltd.
- Xu, Z., Lu, L., & Shi, P. (2008). A hybrid approach to gender classification from face images. *2008 19th International Conference on Pattern Recognition*, 1–4. doi:10.1109/ICPR.2008.4761883

Yang, J., & Yang, J. (2003). Why can LDA be performed in PCA transformed space? *Pattern Recognition*, 36(2), 563–566. doi:10.1016/S0031-3203(02)00048-1

Yang, M.-H., Kriegman, D. J., & Ahuja, N. (2002). Detecting faces in images: a survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(1), 34–58. doi:10.1109/34.982883

Zheng, J., & Lu, B.-L. (2011). A support vector machine classifier with automatic confidence and its application to gender classification. *Neurocomputing*, 74(11), 1926–1935. doi:10.1016/j.neucom.2010.07.032