

UNIVERSIDAD AUTÓNOMA DE MADRID  
ESCUELA POLITÉCNICA SUPERIOR



– TRABAJO FIN DE MÁSTER –

# ESTIMADOR AUTOMÁTICO DE ATRIBUTOS CORPORALES A DISTANCIA.

Máster Universitario en Ingeniería de  
Telecomunicación

Patricia Marín Belinchón

Febrero 2017

Copyright © 2017 por Patricia Marín Belinchón. Todos los derechos reservados. Ninguna parte de esta publicación puede ser reproducida o transmitida en cualquier forma o por cualquier medio, electrónico o mecánico, incluyendo fotocopia, grabación, o cualquier sistema de almacenamiento y recuperación de información sin el permiso por escrito del autor. La Universidad Autónoma de Madrid tiene derechos para poder reproducir y distribuir electrónicamente este documento.

# ESTIMADOR AUTOMÁTICO DE ATRIBUTOS CORPORALES A DISTANCIA.

AUTORA: Patricia Marín Belinchón

DIRECTOR: Rubén Vera Rodríguez

PONENTE: Julián Fierrez Aguilar



ATVS - Biometric Recognition Group  
Escuela Politécnica Superior  
Universidad Autónoma de Madrid  
Febrero 2017



# Resumen y Palabras Clave

## Resumen

Dado el interés creciente en las Soft Biometrics y su aplicación en numerosas áreas relacionadas con la biometría, en este trabajo se ha elegido este campo haciendo referencia a aquellas características que se basan únicamente en los atributos corporales de un sujeto. Los atributos Soft Biometrics seleccionados en este proyecto han sido: Altura, Ancho de Hombros, Ancho de Caderas, Largo de los brazos, Envergadura y Color de Pelo de una persona.

Se parte de la premisa de que la estimación automática de estos atributos corporales mejoraría el reconocimiento de personas en diversos sistemas, como pueden ser los de vídeo vigilancia, y aportaría una mayor robustez en los sistemas biométricos clásicos.

Sin embargo, dada la dificultad y limitaciones existentes en la extracción de dichos atributos, en este trabajo se pretende desarrollar un estimador automático de Soft Biometrics corporales.

Durante el desarrollo del trabajo se ha diferenciado el tratamiento para atributos objetivos y subjetivos, se han modelado estas Soft Biometrics extrayendo features propias y features utilizadas en trabajos relacionados y se han entrenado clasificadores SVM para la estimación de a qué grupo pertenece cada atributo correspondiente al sujeto.

Finalmente se han evaluado los resultados obtenidos utilizando un protocolo experimental y dividiendo su análisis según las features utilizadas para la estimación de los atributos, demostrando cuál funciona mejor para la clasificación de cada uno de ellos.

En definitiva, el principal objetivo del proyecto ha sido implementar un estimador que, a partir de una imagen de una persona tomada a distancia, extraiga de forma automática sus principales atributos corporales.

## Palabras Clave

Soft Biometrics, Atributos corporales, Vídeo Vigilancia, Estimador automático, Clasificadores SVM.



# Abstract and Keywords

## Abstract

Given the growing interest in Soft Biometrics and its application in many areas related to biometrics, in this work this field has been chosen referring to those characteristics that are based only on the body attributes of a subject. The selected Soft Biometrics attributes in this project have been the following: Height, Shoulder width, Hip width, Arms length, Span and Hair color of a person.

The automatic estimation of these corporal attributes would improve the people's recognition in different systems, such as video surveillance, and would provide greater robustness in classical biometric systems.

However, given the existing difficulty and limitations in the extraction of these attributes, this work intends to develop an automatic estimator of Body Soft Biometrics.

During the development of this work, the treatment for objectives and subjective attributes has been differentiated, these Soft Biometrics have been modelled by extracting own features and related work's features, and then SVM classifiers have been trained for estimating to which group each attribute corresponding to the subject belongs.

Finally, the obtained results have been evaluated using an experimental protocol and dividing the analysis according to the features that have been used for estimating the correspondent attributes. It shows which feature works better for classifying each one of them.

In conclusion, the main objective of the project is to implement an estimator that, based on an image of a person taken at a distance, automatically extract its main body attributes.

## Keywords

Soft Biometrics, Corporal features, Video Surveillance, Automatic Estimator, SVM Classifiers.





## Agradecimientos

*En primer lugar, quiero agradecer al ATVS de la Escuela Politécnica Superior y, en especial, a Rubén Vera, por la oportunidad de haberme permitido realizar el Trabajo Fin de Máster en su área. Gracias por tus consejos, por darme únicamente facilidades en todo momento y por haberme dejado la flexibilidad que necesitaba para poder compaginar este proyecto con la vida laboral.*

*El camino no ha sido fácil hasta llegar aquí, puede que me haya encontrado más obstáculos de los que esperaba y que en algún momento han hecho que aflojara las fuerzas para concluir merecidamente con un proyecto de Fin de Máster. Sin embargo, escribiendo estas líneas siento una enorme satisfacción de haberlo logrado al fin, con constancia, dedicación y esfuerzo. Jamás se debe dejar algo sin terminar.*

*Por ello, debo mencionar a todas las personas que me han dado ese empujón de energía que he podido necesitar, aquellas que pase lo que pase nunca dejan de creer en mí.*

*Para empezar, mis mejores amigas. Gracias chicas por preocuparos siempre por mí, por entender mis fines de semana de no poder salir de casa por trabajar, y por estar siempre dispuestas a ser las primeras en celebrar lo que haga falta y compartir mis alegrías.*

*A mis padres y mi hermana, que siempre han creído y creerán en mí a ciegas. Ya lo he dicho y lo vuelvo a repetir, si estoy orgullosa de ser como soy es gracias a vosotros y vuestro apoyo incondicional.*

*Y por último a ti. Eres mi fuente de energía y el principal culpable de todo lo bueno que me pueda estar pasando, y eso incluye el hecho de que hoy esté pudiendo entregar este proyecto. No hay ningún agradecimiento que te haga justicia ni que quepa en esta página, pero gracias por ser siempre mi motor de vida.*

*A todos, os dedico este trabajo. Gracias.*

*Patricia Marín Belinchón  
Febrero 2017*



# Índice general

<b>Índice de figuras</b>	<b>XI</b>
<b>Índice de tablas</b>	<b>XIV</b>
<b>Glosario de acrónimos</b>	<b>XVII</b>
<b>1. Introducción</b>	<b>1</b>
1.1. Motivación . . . . .	1
1.2. Objetivos . . . . .	2
1.3. Estructura de la memoria . . . . .	2
<b>2. Estado del arte: Soft Biometrics</b>	<b>3</b>
2.1. Soft Biometrics . . . . .	3
2.2. Trabajos Relacionados . . . . .	4
2.2.1. Identificación mediante Descripciones Comparativas . . . . .	4
2.2.2. Autenticación continua de Usuario basada en Atributos . . . . .	6
2.2.3. Estimación de Género a partir de features Corporales y Faciales. . . . .	8
2.2.4. Reconocimiento de atributos humanos mediante Contextos de escena . . . . .	9
2.3. Aplicaciones de las Soft Biometrics . . . . .	10
2.3.1. Vídeo Vigilancia . . . . .	10
2.3.2. Aplicaciones Comerciales . . . . .	13
<b>3. Sistema, Diseño y Desarrollo</b>	<b>19</b>
3.1. Extracción de puntos Automática . . . . .	19
3.1.1. Segmentado de la imagen . . . . .	19
3.1.2. Extracción de puntos . . . . .	23
3.1.2.1. Altura . . . . .	24
3.1.2.2. Ojos, Nariz y Barbilla . . . . .	25
3.1.2.3. Hombros . . . . .	25
3.1.2.4. Caderas . . . . .	26
3.1.2.5. Manos . . . . .	28

3.2.	Normalización . . . . .	30
3.3.	Extracción de features . . . . .	33
3.3.1.	Atributos objetivos: Distancia entre Puntos . . . . .	34
3.3.2.	Atributo subjetivo: Envergadura . . . . .	35
3.3.3.	Atributo subjetivo: Color del pelo . . . . .	35
3.3.4.	Forma del cuerpo ( <i>Body Shape</i> ) . . . . .	36
3.4.	Clasificadores . . . . .	37
3.4.1.	SVM (Support Vector Machines) . . . . .	37
3.4.2.	Entrenamiento . . . . .	39
3.4.3.	Test . . . . .	39
3.5.	Procesado Manual Inicial . . . . .	39
<b>4.</b>	<b>Experimentos realizados y Resultados</b>	<b>45</b>
4.1.	Base de Datos . . . . .	45
4.2.	Protocolo Experimental . . . . .	46
4.3.	Evaluación de Resultados . . . . .	52
4.3.1.	Evaluación de etiquetado . . . . .	53
4.3.2.	Evaluación del Sistema Real . . . . .	55
<b>5.</b>	<b>Conclusiones y Trabajo futuro</b>	<b>69</b>
5.1.	Conclusiones . . . . .	69
5.2.	Trabajo futuro . . . . .	70
	<b>Bibliografía</b>	<b>73</b>

# Índice de figuras

2.1. Recopilación de etiquetas comparativas. Fuente: [1] . . . . .	5
2.2. Soft Biometrics utilizadas para comparar Sujetos. Fuente: [1] . . . . .	6
2.3. Overview del método de Autenticación basado en atributos. Fuente: [2] . . . . .	7
2.4. Clasificadores de atributos para imágenes faciales de ejemplo. Fuente: [2] . . . . .	7
2.5. Esquema general para la estimación del género basado en features corporales y faciales. Fuente: [3] . . . . .	8
2.6. Algoritmo para la forma del cuerpo y la textura, para una distancia lejana. Fuente: [3] . . . . .	9
2.7. Arquitectura de red para el reconocimiento de atributos humanos utilizando contextos jerárquicos. Fuente: [4] . . . . .	10
2.8. Región del cuerpo visible para 3 distancias consideradas. Una persona caminando de frente a la cámara es capturada con una vídeo cámara de alta resolución (10 fps y resolución de $1600 \times 1200$ ) y se extraen las etiquetas soft disponibles visualmente en cada escenario. Fuente:[5] . . . . .	12
2.9. Google Cloud Vision API. Ejemplo de aplicación sobre una imagen. Fuente:[6] . .	15
2.10. Microsoft Computer Vision API. Ejemplo de aplicación sobre una imagen. Fuente:[7]	16
2.11. Objetivo de mejora en APIs tras añadir los resultados de un estimador de atributos corporales. . . . .	18
3.1. Diagrama de Implementación del Estimador de Atributos Corporales. . . . .	20
3.2. Vecindad de píxeles en procesamiento de imágenes: Conectividad 4 y 8. . . . .	21
3.3. Resultados obtenidos en el proceso de Segmentado de la imagen. . . . .	22
3.4. Puntos necesarios a extraer en la imagen para permitir estimar atributos corporales.	23
3.5. Extracción automática de puntos de Altura. . . . .	24
3.6. Extracción automática de puntos de Hombros. . . . .	26
3.7. Proporciones antropométricas de la figura humana. Fuente: Andrew Loomis [8]. .	27
3.8. Extracción automática de puntos para la cadera: derecha e izquierda. . . . .	28
3.9. Extracción automática de puntos del final de las Manos. . . . .	29
3.10. Transformaciones afines aplicadas sobre una imagen: escalado, cizallado y rotación. Fuente: [9] . . . . .	31
3.11. Normalización horizontal y vertical llevada a cabo sobre las coordenadas de un punto extraído en la imagen. . . . .	33

3.12. Soft Biometrics corporales que ya son extraídas por la API de Microsoft: edad y género. . . . .	34
3.13. Extracción automática de una medida para el Color de Pelo. . . . .	36
3.14. Procesado de extracción automática de una feature para la Forma del Cuerpo. . .	37
3.15. Representación de muestras pertenecientes a dos clases distintas. a) observaciones y posibles hiperplanos de separación. b) hiperplano de separación óptimo e hiperplanos H1 y H2 . . . . .	38
3.16. Clasificación Manual de features según la umbralización de sus histogramas. Ejemplo mostrado para el atributo de la altura. . . . .	41
3.17. Diagrama general de la implementación del estimador de atributos corporales: procesado automático y procesado manual. . . . .	43
4.1. Tunnel Database setup. Existen 8 cámaras de adquisición de la forma de caminar y una única cámara de alta resolución adquiriendo el frontal de los sujetos. Fuente: [5] . . . . .	45
4.2. Ejemplos de TunnelDB. Fuente: [5] . . . . .	46
4.3. Protocolo experimental 5-fold cross validation. Ejemplo de la distribución de conjuntos en la selección de los conjuntos de entrenamiento y test en cada caso. . . .	47
4.4. Protocolo experimental: Fase de entrenamiento y Test. . . . .	50
4.5. Features unidimensionales: Kernel lineal vs Kernel gaussiano. Ejemplo: Altura. . .	51
4.6. Función Kernel Gaussiano aplicado a datos no separables linealmente. Fuente: Wikipedia. . . . .	52
4.7. Error del Etiquetado Automático de los puntos vs Etiquetado Manual. Distancia media en píxeles. . . . .	53
4.8. Ejemplos más frecuentes de errores en el marcado automático de puntos sobre la imagen. . . . .	54
4.9. Error en la Medida de las features Objetivas: Medidas resultantes del etiquetado Automático vs Medidas resultantes del etiquetado Manual. Diferencia media en píxeles. . . . .	55
4.10. Entrenamiento y Test para la evaluación del Sistema Real. . . . .	56
4.11. Sistema Real. Resultados del Estimador de atributos corporales. . . . .	57
4.12. Matriz de confusión de un clasificador binario. . . . .	58
4.13. Sensibilidad, Especificidad y Precisión del Sistema Real para las features de la fase I. . . . .	60
4.14. Sensibilidad, Especificidad y Precisión del Sistema Real para las features de la fase II. . . . .	60
4.15. Curva ROC de todos los atributos del Sistema, utilizando las features de la fase I. .	62
4.16. Curva ROC de todos los atributos del Sistema, utilizando las features de la fase II. .	63
4.17. Ejemplo 1. Resultado obtenido por el Estimador de Atributos Corporales con las features de la Fase I y Fase II. . . . .	64
4.18. Ejemplo 2. Resultado obtenido por el Estimador de Atributos Corporales con las features de la Fase I y Fase II. . . . .	65

4.19. Ejemplo 3. Resultado obtenido por el Estimador de Atributos Corporales con las features de la Fase I y Fase II. . . . .	66
4.20. Ejemplo 4. Resultado obtenido por el Estimador de Atributos Corporales con las features de la Fase I y Fase II. . . . .	67





## Índice de tablas

2.1. Soft biometrics para vídeo vigilancia. Basada en [10]. . . . .	11
2.2. Rasgos Soft biometrics disponibles (marcados con X) visualmente en cada escenario a distancia. . . . .	13
4.1. Error cometido por el Sistema Real para las features de medidas entre puntos, color de pelo y envergadura. Número de imágenes mal clasificadas y % Acierto. . . . .	57
4.2. Error cometido por el Sistema Real para la feature de forma del cuerpo de la Fase II. Número de imágenes mal clasificadas y % Acierto. . . . .	57
4.3. Valores de la Sensibilidad, Especificidad y Precisión para cada clase de los atributos del sistema para las features de la fase I. . . . .	59
4.4. Valores de la Sensibilidad, Especificidad y Precisión para cada clase de los atributos del sistema para las features de la fase II. . . . .	59



## Glosario de acrónimos

- **API**: Application Programming Interface (Interfaz de Programación de Aplicaciones)
- **SVM**: Support Vector Machine (Máquina de Soporte Vectorial)
- **BBDD**: Base de Datos
- **CCTV**: Circuito Cerrado de Televisión
- **OCR**: Optical Character Recognition (Reconocimiento Óptico de Caracteres)
- **ROI**: Region of Interest (Área de Interés)
- **ROC** Receiver Operating Characteristic
- **HOG** Histogram of Oriented Gradients
- **CNN** Convolutional Neural Network





# 1 Introducción

## 1.1. Motivación

---

En los últimos años, en el ámbito de la biometría ha crecido notablemente el interés sobre las **Soft biometrics**, las cuales hacen referencia a los atributos físicos o de comportamiento que pueden ser inferidos de un sujeto al margen de la biometría clásica [11, 12]. Algunos de estos atributos incluyen la altura, anchura de hombros o cadera, la envergadura, el color del pelo, su longitud, la raza, entre otros.

Estos atributos han ganado mucha importancia debido a la gran cantidad de **beneficios** que conllevan, como es la habilidad de cruzar la brecha que separa a las personas y las máquinas en el contexto de reconocimiento de personas [5].

Con el fin de superar las limitaciones actuales existentes para la estimación de estos atributos corporales, la **principal motivación** de este trabajo tiene como objetivo desarrollar un **estimador automático de estas Soft biometrics** a partir de imágenes de personas situadas a una determinada distancia.

De esta forma, se conseguiría un avance para su uso en **aplicaciones de vídeo vigilancia** [13], sistemas de recuperación de imágenes [14] y, en definitiva, cualquier sistema que suponga una interacción de personas con ordenadores. Además, en el área de biometría, la consecución de esta estimación automática de características corporales puede suponer un incremento de la robustez de los sistemas biométricos clásicos [15].

Actualmente ya existen APIs desarrolladas en esta misma línea, como es el caso de la '*Cloud Vision Api*' de Google [6]. Esta API realiza un análisis de imagen y permite, mediante el uso de modelos de aprendizaje máquina, clasificar rápidamente el contenido de la imagen en una gran variedad de categorías, pudiendo incluso detectar contenido inapropiado o analizar atributos emocionales a partir de las caras de los individuos.

Basándonos en la funcionalidad de la '*Cloud Vision Api*' de Google, en este proyecto se persigue conseguir un resultado similar, pero específico para features corporales de personas. Por tanto, otra parte de la motivación de este trabajo es implementar un sistema que, a partir del tratamiento de imagen, pueda servir como **aplicación comercial** para la estimación de atributos corporales a partir de cualquier imagen dada, mejorando sustancialmente el alcance y funcionalidad del estimador desarrollado.

Por todo lo anterior, en el actual proyecto se propone la implementación de un estimador que, tras una fase previa de entrenamiento de las soft biometrics seleccionadas, sea capaz de describir los atributos corporales de cualquier persona de forma automática a distancia.

Este proyecto plantea una primera fase de estudio de trabajos relacionados con las Soft Biometrics, una implementación de la extracción automática de features que se puedan utilizar para entrenar el sistema de estimación de atributos, y una última parte de evaluación de resultados, tras haber desarrollado clasificadores para cada uno de estos atributos corporales.

## 1.2. Objetivos

---

De forma más concreta, los objetivos que se persiguen conseguir en este TFM son:

- Extracción automática de features a partir de una imagen de una persona a distancia.
- Entrenamiento de clasificadores SVMs que, a partir de unas features extraídas de la imagen, sean capaces de calcular scores de pertenencia a un grupo para diferentes atributos corporales previamente seleccionados.
- Obtención de un estimador de características que indique todas las puntuaciones que una imagen de una persona tiene para los diferentes posibles atributos corporales.

En definitiva, el principal objetivo del trabajo es implementar un **estimador** que, a partir de una imagen de una persona tomada a distancia, extraiga de forma automática sus principales **atributos corporales**.

## 1.3. Estructura de la memoria

---

A lo largo del presente documento se desarrollará el trabajo realizado de la siguiente forma:

- **Capítulo 2.** En este capítulo se describe el estado del arte. Se define el concepto de Soft Biometrics en el que está basado este trabajo, se relacionan con su implicación en el sector de la vídeo-vigilancia y, por último, su aplicación en el ámbito comercial.
- **Capítulo 3.** En él se desarrolla el detalle de la implementación del estimador de atributos corporales, incluyendo el diseño del sistema y el desarrollo realizado.
- **Capítulo 4.** En este capítulo se desarrolla el protocolo experimental y se evalúan los resultados obtenidos por el estimador implementado, teniendo en cuenta la BBDD utilizada, el método de entrenamiento llevado a cabo y los resultados observados.
- **Capítulo 5.** Finalmente, se describen las conclusiones analizadas en este trabajo y se proponen líneas de trabajo futuro para su mejora.



## 2 Estado del arte: Soft Biometrics

### 2.1. Soft Biometrics

El primer sistema en la historia que intentó describir e identificar a las personas basándose en rasgos morfológicos y fisiológicos fue el sistema desarrollado por Alphonse Bertillon [16]. Este sistema estaba basado en características tales como medidas corporales (antropometría), descripciones morfológicas sobre la apariencia, forma del cuerpo, y marcas peculiares observadas sobre el cuerpo. Este sistema fue muy utilizado en sus comienzos para la persecución de criminales, pero tenía una inaceptable alta tasa de falsa identificación. Esto ocurría debido a que estas características no eran lo suficientemente permanentes ni discriminantes para identificar de forma unívoca a un individuo durante un periodo de tiempo. Por lo tanto, las características suaves, del inglés **Soft Biometrics**, se definen como aquellas características que proveen algún tipo de información sobre un individuo que permite discriminar de forma suficientemente diferenciada entre dos personas distintas.

Los rasgos Soft biometrics pueden ser tanto continuos como discretos. Los rasgos como el género, color de ojos, etnia, etc. son discretos por naturaleza. Por otro lado, los rasgos como la altura y peso son variables continuas. En principio un sistema que está basado completamente en características soft biometrics no pueden proveer la precisión requerida en el reconocimiento de individuos por sí solas. Sin embargo, los rasgos soft biometrics se pueden utilizar para **mejorar un sistema tradicional de biometría** (por ejemplo, forma de caminar, cara, etc.) de muchas formas.

Los primeros trabajos en soft biometrics [17, 15, 18] intentaron utilizar información demográfica (por ejemplo, el género y la etnia) y atributos poco discriminantes (soft) como el color de ojos, altura, peso, otras marcas visibles como cicatrices [19, 20] y tatuajes [21] como información auxiliar para mejorar el rendimiento de los sistemas biométricos. Con ello se demostró que las soft biometrics pueden complementar los identificadores biométricos tradicionales (primarios) y también pueden ser útiles como fuente de evidencia en juicios legales, ya que son **más descriptivos** que los scores numéricos generados, por ejemplo, por un identificador facial tradicional.

Sin embargo, como se ha comentado, en la mayor parte de los casos esta información secundaria no es suficiente por sí misma para reconocer a un usuario.

En este trabajo se incluye la aplicación de soft biometrics para permitir describir a cualquier individuo en función de los rasgos corporales que se pueden observar sobre él en un vídeo o imagen.

Las Soft Biometrics se pueden obtener a partir de sensores de baja calidad, incluyendo cámaras de vigilancia. En comparación, un rasgo biométrico 'hard', como puede ser el reconocimiento automático de caras, puede entrar en conflicto con una baja resolución y un punto de vista que no sea frontal. Además, las Soft Biometrics requieren menos cálculos comparadas con las 'hard biometrics', no necesitan cooperación por parte del usuario y no son invasivas, lo cual las hace ideales para aplicaciones de vídeo vigilancia.

El uso de rasgos de soft biometrics en sistemas automatizados de reconocimiento biométrico tiene **muchos beneficios**. Por tanto, es esencial investigar cuidadosamente temas relacionados con su capacidad de extracción y de reconocimiento. Por ejemplo en vídeo vigilancia, los fotogramas de las cámaras de seguridad son generalmente de baja calidad y por ello las formas tradicionales de identificación a distancia no pueden ser utilizadas fácilmente. Las Soft Biometrics ofrecen una solución con respecto a este punto, pero carecen de la discriminación que se espera de los rasgos biométricos.

Esta tecnología es ideal para aplicaciones con alcances menores, como por ejemplo perseguir personas con una red de cámaras o identificar un individuo para el que se conoce que se encuentra en un área en concreto. En estas aplicaciones la invarianza del punto de vista es un factor clave cuando se trabaja con fotogramas de vídeo vigilancia. Una técnica exitosa debe identificar los rasgos soft biometrics desde **cualquier punto de vista del sujeto**.

El énfasis se debe poner en encontrar formas prácticas de obtener rasgos invariantes con el punto de vista (similar a [22]) o en desarrollar métodos para predecir rasgos ocultos basados en lo que se puede observar.

Más adelante en este capítulo se presenta un set en concreto para soft biometrics, aquél para el que se va a aplicar este trabajo: **soft biometrics para vídeo vigilancia**, las cuales son extraídas visualmente a partir de un sujeto a distancia (por ejemplo el género, altura, longitud de brazo, etc).

## 2.2. Trabajos Relacionados

### 2.2.1. Identificación mediante Descripciones Comparativas

Una de las principales ventajas de las Soft Biometrics es su relación con la descripción de las personas, es decir, las Soft Biometrics son utilizadas de forma natural para identificar y describir a cada individuo. Esto permite la identificación basándonos en la descripción del sujeto, posiblemente dada por un testigo.

Dado que las propias personas pueden ser imprecisas cuando predicen medidas exactas, las etiquetas absolutas se vieron como un método más robusto para obtener descripciones humanas (por ejemplo, decir que alguien es alto). El mayor problema de este método absoluto es su naturaleza altamente subjetiva. El significado de una etiqueta se basa en los propios atributos de la persona y su propia percepción de la media y variación de las medidas para un individuo. Esto hace que las etiquetas sean menos fiables, y discriminen de una forma muy pobre.

Por todo ello, se introdujo un nuevo método [1] para obtener descripciones de personas explotando el proceso de realizar **comparaciones visuales** entre varios sujetos.

Comparar la apariencia de dos sujetos es un proceso muy natural y puede ser más fiable que el uso de etiquetas absolutas porque las comparaciones se asignan basándose en un 'benchmark' específico, dando un resultado más objetivo. Se explota la facilidad de realizar comparaciones para conseguir descripciones más fiables y robustas.



En este método [1] se describen un conjunto de medidas relativas que describen al sujeto y que se pueden inferir a partir de etiquetas comparativas (comparando una persona con otra, por ejemplo 'más alto que'). Este conjunto se utiliza como 'firma' biométrica de la persona en cuestión, la cual ha demostrado ser altamente discriminante.

Una comparación entre dos personas únicamente explicará la diferencia entre ambas. Por ello, los rasgos físicos que se le asignen a una persona dependerán del sujeto con el que ha sido comparada. Un método práctico para obtener comparaciones entre un sujeto objetivo (representando al sospechoso en una aplicación) y otros sujetos es presentar videos de los individuos al anotador. Esto permite múltiples comparaciones con el mínimo equipamiento y personal. Previamente se habrían coleccionado categorías absolutas para la misma base de datos, permitiendo observar las diferencias según estas categorías.



Figura 2.1: Recopilación de etiquetas comparativas. Fuente: [1]

En la figura 2.1 se puede observar cómo se han obtenido todas estas comparaciones en una website diseñada para mostrar al anotador videos de un sujeto y del sujeto objetivo de forma simultánea. Las cajas desplegadas para cada atributo permite a los usuarios describir cómo difiere el sujeto de la persona objetivo.

Una única comparación consiste en 16 características a ser comparadas (mostradas en la figura 2.2), cada una definida según las 5 posibles etiquetas correspondientes. Tres características (género, etnia y color de la piel) fueron anotadas utilizando etiquetas absolutas, ya que no pueden ser comparativas por naturaleza. Por ello no son analizadas durante la comparación, únicamente para reconocimiento.

Se identificó que las anotaciones comparativas difieren de las etiquetas absolutas en un 17% de las ocasiones. Las grandes diferencias entre estas etiquetas para características para las que se ha demostrado que es difícil describir con etiquetas absolutas, podría ser un indicador de mejoras potenciales al utilizar etiquetas comparativas. Un ejemplo es el largo del brazo o la forma de los hombros, que difieren un 30% y 43% comparados con etiquetas absolutas. Esta diferencia sugiere que las **anotaciones comparativas contienen información más detallada**. Sin embargo, las pequeñas diferencias para características que se describían bien con etiquetas absolutas (por ejemplo, largo del pelo), demuestra que las comparaciones funcionan igual de bien que ellas.

Estas 19 medidas relativas han sido explotadas para identificación biométrica. El reconocimiento de sujetos de una base de datos de soft biometrics demostró el poder discriminatorio de las medidas relativas, alcanzando una precisión de reconocimiento del 92% con 10 comparaciones. Esta precisión incrementó al 95% con 20 comparaciones.

Trait	Description Type	Labels
Arm Length	Comparative	[Much Shorter, Shorter, Same, Longer, Much Longer]
Arm Thickness	Comparative	[Much Thinner, Thinner, Same, Thicker, Much Thicker]
Chest	Comparative	[Much Smaller, Smaller, Same, Bigger, Much Bigger]
Figure	Comparative	[Much Smaller, Smaller, Same, Larger, Much Larger]
Height	Comparative	[Much Shorter, Shorter, Same, Taller, Much Taller]
Hips	Comparative	[Much Narrower, Narrower, Same, Broader, Much Broader]
Leg Length	Comparative	[Much Shorter, Shorter, Same, Longer, Much Longer]
Leg Thickness	Comparative	[Much Thinner, Thinner, Same, Thicker, Much Thicker]
Muscle Build	Comparative	[Much Leaner, Leaner, Same, More Muscular, Much More Muscular]
Shoulder Shape	Comparative	[More Square, Same, More Rounded]
Weight	Comparative	[Much Thinner, Thinner, Same, Fatter, Much Fatter]
Age	Comparative	[Much Younger, Younger, Same, Older, Much Older]
Ethnicity	Absolute	[European, Middle Eastern, Far Eastern, Black, Mixed, Other]
Gender	Absolute	[Female, Male]
Skin Color	Absolute	[White, Tanned, Oriental, Black]
Hair Color	Comparative	[Much Lighter, Lighter, Same, Darker, Much Darker]
Hair Length	Comparative	[Much Shorter, Shorter, Same, Longer, Much Longer]
Neck Length	Comparative	[Much Shorter, Shorter, Same, Longer, Much Longer]
Neck Thickness	Comparative	[Much Thinner, Thinner, Same, Thicker, Much Thicker]

Figura 2.2: Soft Biometrics utilizadas para comparar Sujetos. Fuente: [1]

Las descripciones comparativas han demostrado contener más información discriminante y presentan una aproximación diferente para obtener descripciones robustas de personas mediante Soft Biometrics y posibles testigos.

Esta metodología es muy diferente a la que se ha llevado a cabo en este proyecto, puesto que no se han utilizado etiquetas comparativas. Sin embargo, es una buena referencia para la detección de Soft Biometrics corporales.

### 2.2.2. Autenticación continua de Usuario basada en Atributos

Los métodos tradicionales de autenticación de usuarios en dispositivos móviles están basados en contraseñas, números pin, patrones secretos o huellas. Mientras que están encendidos, los dispositivos no tienen mecanismos para verificar que el usuario que realizó la autenticación es todavía el que lo está utilizando, lo cual puede conllevar a que personas no autorizadas obtengan acceso a información personal del usuario.

Para tratar este problema, se han desarrollado varios sistemas de autenticación continua en los que los usuarios son validados continuamente después del acceso inicial al móvil.

En concreto, se presenta un sistema de autenticación continua basada en atributos para usuarios de smartphones [2]. La figura 2.3 da una idea global del método propuesto. Dada una imagen facial detectada por la cámara frontal, se determina un score comparando las features extraídas con las features que se corresponden con el usuario inscrito (el dueño del dispositivo). Estos valores de score son los que se utilizan para la autenticación del usuario.

Para entrenar los clasificadores de atributos, se ha utilizado una BBDD de imágenes faciales.

1. Partes faciales. Los componentes faciales que se consideran para el entrenamiento son los ojos, nariz, boca, pelo, combinaciones entre todas ellas, y la cara completa.
2. Features. Para cada atributo será necesario un tipo de feature. Por ejemplo, para el color de pelo se utilizarán features relacionadas con el color.
3. Localización de features. Para capturar información local, se consideran diferentes tamaños de celdas para las features.

Un SVM aprende de todas estas features, se ejecuta exhaustivamente para entrenar todos los modelos posibles. Para cada clasificador, se utiliza para el entrenamiento el 80 % de los datos disponibles y para el test se utiliza el 20 %, siempre sin solapar las imágenes faciales de test y entrenamiento.

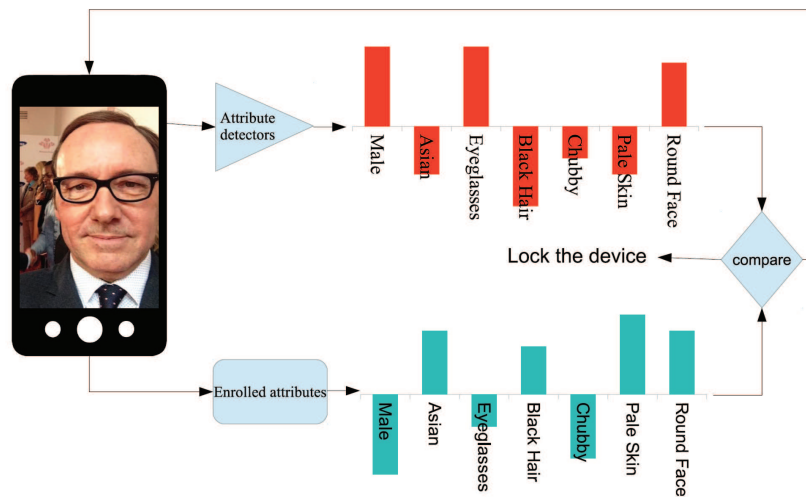


Figura 2.3: Overview del método de Autenticación basado en atributos. Fuente: [2]

El problema de autenticación se considera un problema de verificación en el que dadas dos imágenes, se determina si corresponden a la misma persona o no. La precisión resultante para la mayoría de los atributos es alta.

Además, en la figura 2.4 se muestran algunos ejemplos de la extracción de los clasificadores de atributos, escalados entre -0.5 y 0.5, que corresponden a cada una de las caras de las personas.

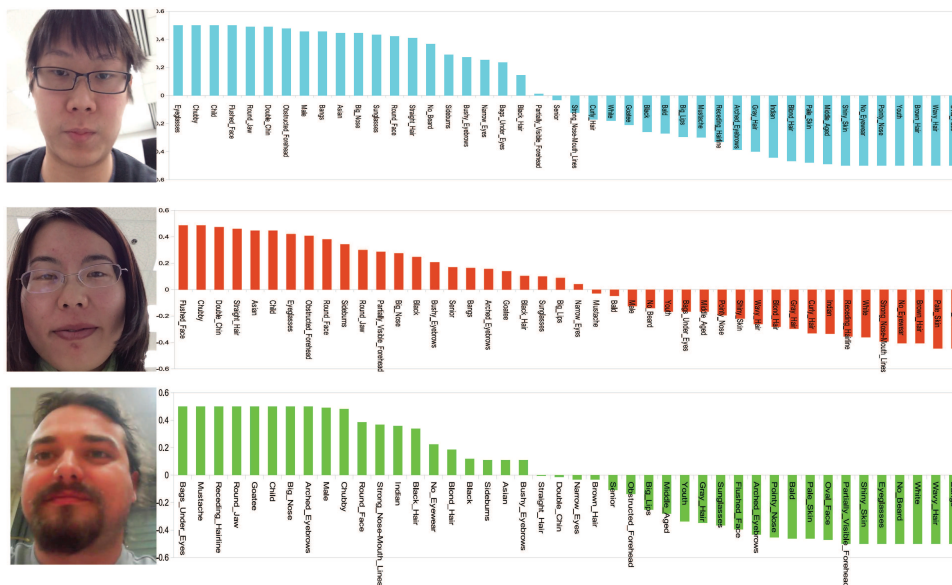


Figura 2.4: Clasificadores de atributos para imágenes faciales de ejemplo. Fuente: [2]

Con este trabajo [2] se muestra que los scores basados en atributos pueden mejorar los resultados de verificación por sí solos. Además, en situaciones en las que las features también son fiables, los resultados de verificación también se pueden mejorar fusionándose con estos scores.

Más adelante se propuso una extensión de este trabajo [23], en la cual se pretendía resolver este problema de autenticación continua en dispositivos mediante el uso de arquitecturas de redes neuronales. Cada red predice atributos multi faciales a partir de un componente facial. El funcionamiento de esta ampliación del sistema de autenticación funciona mejor que el anterior

método propuesto.

Este proyecto busca obtener unos resultados de una forma parecida a la que se ha mostrado en la figura 2.4, y también utilizando clasificadores SVM. Sin embargo, no se restringe solamente a la zona facial, sino que se extraerán Soft Biometrics corporales.

### 2.2.3. Estimación de Género a partir de features Corporales y Faciales.

Otro trabajo relacionado con las Soft Biometrics es la estimación del género de un individuo a partir del análisis conjunto de su cuerpo y cara [3]. El algoritmo de estimación utilizado extrae features geométricas y de textura a partir de una silueta dada.

Dado un sujeto, el género se estima a través de un algoritmo facial desarrollado en otro trabajo relacionado, OpenBR [24]. Lo que se propone en [3] es estimar adicionalmente el género a través de información corporal, mejorando el rendimiento en entornos no controlados, como pueden ser imágenes de baja resolución o caras ocultas.

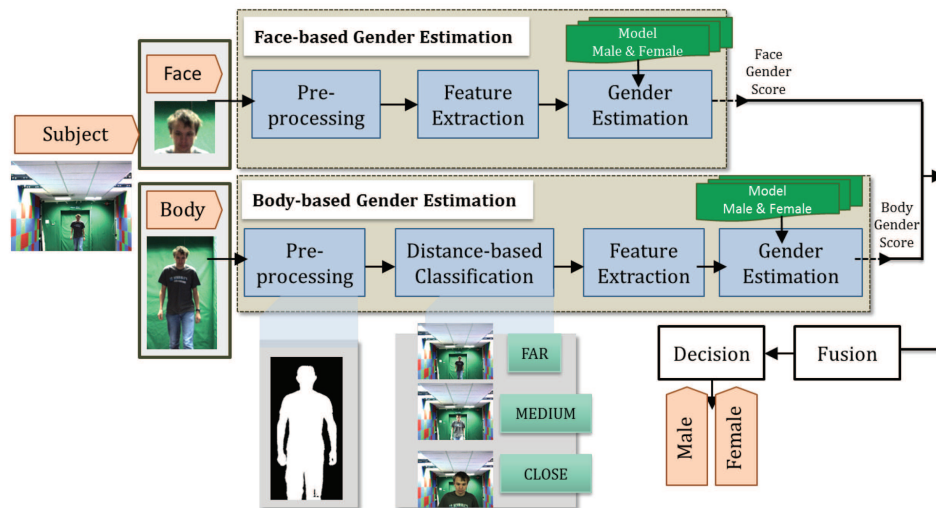


Figura 2.5: Esquema general para la estimación del género basado en features corporales y faciales. Fuente: [3]

En la figura 2.5 se presenta el esquema general presentado en este trabajo. Dada la imagen de un sujeto, se detecta la cara y se extraen y clasifican las features de apariencia con el algoritmo de estimación facial de género OpenBR. Adicionalmente se estima el género a través de la información corporal ya incluida en la misma imagen para mejorar el rendimiento. Para el cuerpo, se extrae la silueta del individuo para determinar en primer lugar la distancia a la cámara y luego predecir el género de acuerdo a esa distancia. Después, se utiliza la información de la forma y de texturas. Por último, ambas se combinan para dar una estimación mejorada del género en las tres situaciones de distancia sujeto-cámara posibles.

Para la estimación del género basado en el cuerpo:

- Se extraen features basadas en la forma del cuerpo, utilizando una medida continua relacionada con la silueta. Una vez que se tiene la imagen segmentada, se procede a seleccionar la región en la que se encuentra la figura y se realiza un cálculo de la suma de los píxeles de frente según las filas y según las columnas de la imagen. El vector final es la concatenación de ambas sumas de píxeles.

- Se extrae información de la textura, utilizando HOG (Histogram of Oriented Gradients). Se utiliza la región recortada previamente y se divide en bloques, sobre los que se aplica HOG. El vector final será la concatenación de los histogramas calculados en cada bloque.

Estos dos algoritmos aplicados sobre el cuerpo se pueden observar en la figura 2.6.

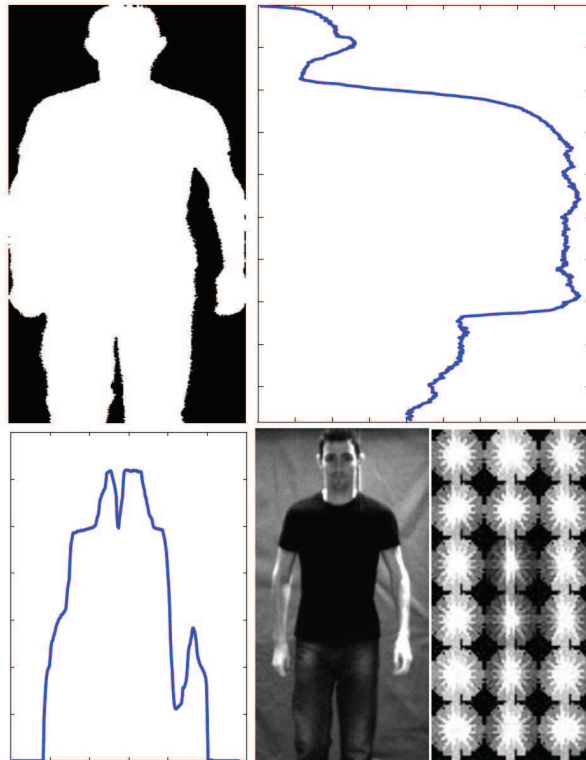


Figura 2.6: Algoritmo para la forma del cuerpo y la textura, para una distancia lejana. Fuente: [3]

Por último, se utilizan clasificadores SVM para ambos algoritmos, y los scores resultantes se fusionan. Se utiliza un modelo de estimación distinto en función de la distancia a la cámara: lejos, medio y cerca, ya que el modelo de género depende mucho de ella.

Añadiendo información corporal a la información facial, el rendimiento del sistema de estimación de género mejora sustancialmente, y además es robusto a condiciones no controladas.

La BBDD utilizada en este trabajo relacionado, es la que se ha utilizado para el desarrollo de este proyecto, como se verá más adelante.

#### 2.2.4. Reconocimiento de atributos humanos mediante Contextos de escena

El reconocimiento preciso de atributos humanos como el género y el estilo de ropa pueden beneficiar a muchas aplicaciones como es la identificación de personas en vídeos. Sin embargo, esto aún sigue siendo un desafío en escenarios no controlados en los que las imágenes tienen variaciones del punto de vista, pose, iluminación y oclusión.

Para ello, se propone el uso de contextos jerárquicos. El método propuesto en [4] está basado en partes, combina partes humanas y *deep learning* para permitir robustez al tratar con imágenes de personas no controladas. En este método se entrenan features CNN (Convolutional Neural Network) y módulos de scoring de atributos, al contrario de lo que se realiza en los métodos que optimizan un clasificador SVM lineal para predecir.

La figura 2.7 muestra un esquema global de la arquitectura de la red de este método.

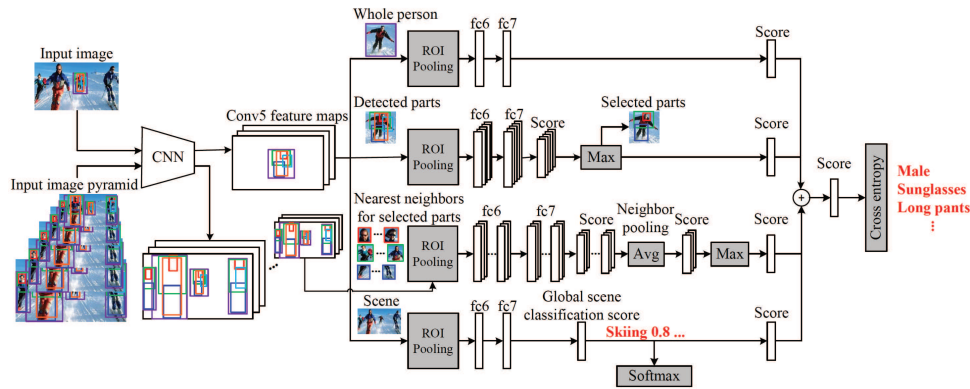


Figura 2.7: Arquitectura de red para el reconocimiento de atributos humanos utilizando contextos jerárquicos. Fuente: [4]

Dada una imagen de entrada, cada persona está asociada con una hipótesis de *bounding-box* y un conjunto de detectores de partes de personas. La imagen de entrada y su pirámide Gaussiana se pasan a través de la red para obtener un mapa de features convolucional. Luego se dividen 4 caminos se scoring de atributos utilizando diferentes regiones en el mapa de features. En los primeros dos caminos, se utilizan respectivamente la *bounding box* de la persona objetivo y cajas por partes que cubren la escala completa del cuerpo y partes locales.

El contexto jerárquico se incorpora en el tercer y cuarto camino. Se selecciona el contexto en el que se encuentra la persona en cuestión a partir de otras personas vecinas a ella en la pirámide de features. El contexto a nivel de escena se asigna al score de clasificación de la escena, antes de recalculer el score de atributos humanos de forma consciente.

Es decir, se desarrolla un método para reconocer atributos humanos, partiendo de la simple observación de que el contexto de la imagen puede revelar más pistas para facilitar el reconocimiento. Para ello no sólo se extrae conjuntamente el score del cuerpo humano y sus partes mediante un CNN, sino que se consiguen dos scores que capturan contextos jerárquicos (vecindad). En conclusión, se captura tanto a la persona objetivo como el **contexto de la escena** en la que se encuentra.

## 2.3. Aplicaciones de las Soft Biometrics

### 2.3.1. Vídeo Vigilancia

Los dos rasgos biométricos más populares para la identificación a distancia son la cara [25] y la forma de caminar [26]. Estas características pueden verse afectadas por la **baja calidad** del sensor de la mayoría de las cámaras CCTV. La baja resolución de las cámaras puede impactar seriamente en el reconocimiento facial, y tasas muy bajas de frame (algunas veces incluso con lapsos de tiempo) confunden el movimiento del cuerpo humano requerido para el reconocimiento de la forma de caminar.

Por el contrario, existen muchos otros **rasgos soft** que pueden ser obtenidos a partir de vídeos o imágenes **con calidad muy pobre**. Estas Soft Biometrics tienen un gran potencial para el uso inmediato en el mundo real sin tener que mejorar la gran infraestructura de vídeo vigilancia ya desplegada, de ahí su gran utilidad. Este conjunto de rasgos se indican en la tabla 2.1.

<b>Cuerpo</b>	
<b>Rasgo</b>	<b>Rango de Valores</b>
1. Largo de brazo	Muy corto, Corto, Normal, Largo y Muy Largo
2. Grosor de brazo	Muy delgado, Delgado, Normal, Grueso y Muy Grueso
3. Tórax	Muy delgado, Delgado, Normal, Grande y Muy Grande
4. Envergadura	Muy pequeña, Pequeña, Normal, Grande y Muy Grande
5. Altura	Muy pequeño, Pequeño, Normal, Alto y Muy Alto
6. Caderas	Muy estrecho, Estrecho, Normal, Ancho y Muy Ancho
7. Largo de pierna	Muy corto, Corto, Normal, Largo y Muy Largo
8. Dirección de pierna	Muy encorvado, Encorvado, Recto, Patizambo y muy Patizambo
9. Grosor de pierna	Muy delgado, Delgado, Normal, Grueso y Muy Grueso
10. Constitución muscular	Muy flaco, Flaco, Normal, Musculado y muy Musculado
11. Proporciones	Normal e Inusual
12. Forma de hombros	Muy redondeados, Redondeados, Normal, Cuadrados y Muy cuadrados
13. Peso	Muy delgado, Delgado, Normal, Pesado y Muy pesado

<b>Global</b>	
<b>Rasgo</b>	<b>Rango de Valores</b>
14. Edad	Infante, Pre Adolescente, Adolescente, Adulto Joven, Adulto, Edad Media, Senior
15. Etnia	Europea, Este Medio, India/Pakistaní, Este lejano, Negro, Mixto, Otro
16. Sexo	Femenino, Masculino

<b>Cabeza</b>	
<b>Rasgo</b>	<b>Rango de Valores</b>
17. Color de piel	Blanco, Moreno, Oriental y Negro
18. Color de vello facial	Ninguno, Negro, Marrón, Rojo, Rubio y Gris
19. Largo de vello facial	Ninguno, Poca barba, bigote, Perilla y Barba completa
20. Color de pelo	Negro, Marrón, Rojo, Rubio, Gris y Teñido
21. Largo de Pelo	Ninguno, Afeitado, Corto, Medio y Largo
22. Largo de Cuello	Muy corto, Corto, Medio y Largo
23. Grosor de Cuello	Muy delgado, Delgado, Normal, Grueso y Muy grueso

Tabla 2.1: Soft biometrics para vídeo vigilancia. Basada en [10].

En este trabajo se han utilizado varias de las soft biometrics descritas en la tabla 2.1, cuyo principal valor radica en que son discernibles por humanos a distancia.

Estas etiquetas de rasgos físicos fueron propuestas por Sina Samangoeei [10] y están disponibles en la base de Datos conocida como *Southampton Multibiometric Tunnel Database (TunnelDB)* [27].

Esta BBDD presenta como limitación que sólo tiene labels obtenidos por personas de forma visual, por tanto se debe tener en cuenta que cada persona tiene una opinión subjetiva sobre estas etiquetas. Este hecho ha sido una motivación para realizar una etapa inicial de etiquetado manual por parte de la autora de este proyecto (Sección 3.5), ignorando las etiquetas que ya vienen dadas, y tomando como ground-truth aquellas extraídas en este trabajo.

Dicha BBDD contiene información anotada a partir de grabaciones de individuos tomadas en condiciones controladas de laboratorio [28]. Presenta 23 rasgos: para cada etiqueta de un rasgo le asignaron un rango de valores discretos, por ejemplo, para 'Largo de Brazo' es posible marcar 1 (muy corto), 2 (corto), 3 (normal), 4 (largo) y 5 (muy largo). El proceso de anotación de cada etiqueta se describe en detalle en [10].

En la tabla 2.1 se muestra, por tanto, un resumen de estas etiquetas de rasgos y sus términos semánticos correspondientes. Estas etiquetas fueron diseñadas basándose en aquellos rasgos humanos capaces de ser utilizados consistentemente y de forma precisa para describir personas a distancia.

Como se puede ver, los rasgos están agrupados en 3 clases, las cuales son:

**Rasgos Corporales** . Describen el objetivo del somatotipo percibido [29] (altura, peso, etc).

Estos rasgos tienen una estrecha correlación entre el estilo y el tipo de ropa que lleva el

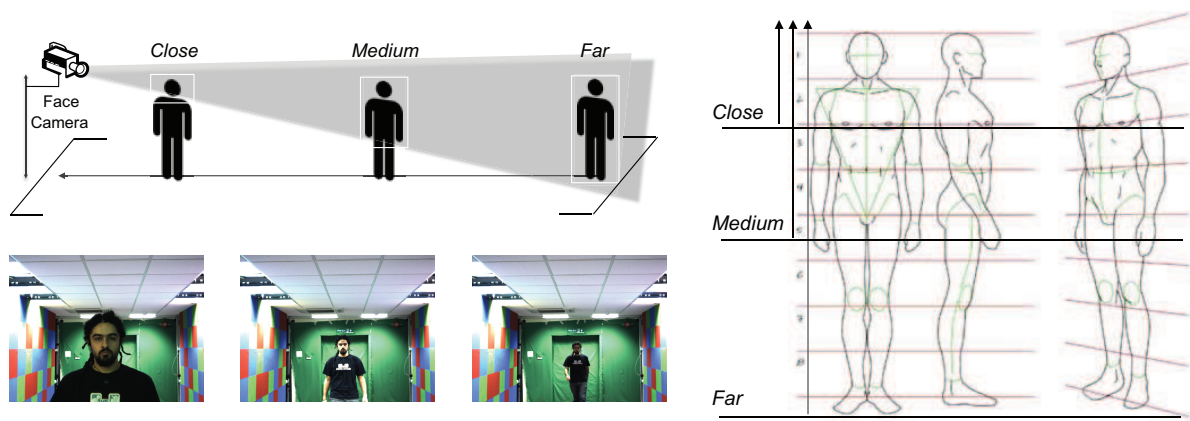


Figura 2.8: Región del cuerpo visible para 3 distancias consideradas. Una persona caminando de frente a la cámara es capturada con una vídeo cámara de alta resolución (10 fps y resolución de  $1600 \times 1200$ ) y se extraen las etiquetas soft disponibles visualmente en cada escenario. Fuente:[5]

sujeto en el proceso de anotación. Por ejemplo, ropa ajustada permitirá obtener etiquetas más estables que ropas más holgadas.

**Rasgos Globales** . Edad, etnia y sexo. La información demográfica como el género o la etnia de una persona normalmente no cambia durante su vida, de tal forma que pueden ser usados para filtrar en la base de datos y reducir el número de candidatos. Por otro lado, la edad se puede estimar a partir de los rasgos físicos a distancia y también puede ser usada para filtrar sospechosos.

**Rasgos de la Cabeza** . Área del cuerpo humano a la que se presta mucha atención siempre que sea visible [30] (color del pelo, barbas, etc). Estas soft biometrics son muy interesantes para ser fusionadas con sistemas de reconocimiento facial.

Siguiendo las definiciones en el capítulo 11 de [31], se puede ver que algunos de los rasgos son *nominales*, es decir, sus valores no se pueden ordenar de una forma coherente (por ejemplo la etnia (15), sexo (16), piel (17), vello facial (18) y color del pelo (20)), mientras que otros son *ordinales*, es decir, sus valores sí pueden ser ordenados con significado (por ejemplo el largo del brazo (1), grosor del brazo (2), altura (4), peso (13) y largo del pelo (21)).

Se asume que estas soft biometrics para vídeo vigilancia son **discernibles** por los humanos a distancia y son extraídas manualmente por un anotador o incluso de forma automática por un sistema automatizado. La otra asunción que se realiza en biometría a distancia es que la cantidad de rasgos soft biometrics en la escena es **variable en función de la distancia**, tal y como se muestra en la figura 2.8.

En este sentido, de forma general se pueden definir tres posibles escenarios variantes con la distancia entre el sujeto y la cámara. Los tres escenarios se definen a continuación (ver figura 2.8):

- **Distancia cercana** Incluye tanto la cara como los hombros.
- **Distancia media** Incluye la mitad superior del cuerpo.
- **Distancia lejana** Incluye el cuerpo completo.



		<i>Cerca</i>	<i>Medio</i>	<i>Lejos</i>
<b>Cuerpo</b>	1. Largo del brazo		X	X
	2. Grueso del brazo		X	X
	3. Pecho		X	X
	4. Envergadura		X	X
	5. Altura			X
	6. Caderas			X
	7. Largo de pierna			X
	8. Dirección de la pierna			X
	9. Grueso de pierna			X
	10. Constitución muscular		X	X
	11. Proporciones	X	X	X
	12. Forma de hombros	X	X	X
	13. Peso		X	X
<b>Global</b>	14. Edad	X	X	X
	15. Etnia	X	X	X
	16. Sexo	X	X	X
<b>Cabeza</b>	17. Color de piel	X	X	X
	18. Color vello facial	X	X	X
	19. Largo vello facial	X	X	X
	20. Color de pelo	X	X	X
	21. Largo de pelo	X	X	X
	22. Largo de cuello	X	X	X
	23. Grueso de cuello	X	X	X

Tabla 2.2: Rasgos Soft biometrics disponibles (marcados con X) visualmente en cada escenario a distancia.

La razón fundamental detrás de la metodología propuesta es el hecho de que dependiendo del escenario particular, algunas etiquetas pueden no verse presentes y otras pueden estar escondidas. Como resultado de ello, la información discriminante de las soft biometrics variará dependiendo de la distancia. La tabla 2.2 muestra las etiquetas soft disponibles para cada escenario definido.

De forma más concreta, en este proyecto se ha trabajado con imágenes de personas para las que no se mostraba la figura completa, motivo por el cual ha sido importante tener en cuenta la variabilidad de los atributos corporales en función de la distancia del sujeto a la cámara.

En este trabajo, dado que el objetivo es estimar **atributos corporales** de las personas, se ha seleccionado únicamente la **distancia más lejana** posible de la base de datos (que en este caso incluye la parte del cuerpo de rodillas hacia arriba), ya que es la que permite observar la mayor información posible en cuanto a soft biometrics se refiere. Por tanto, de esta forma se podrían extraer prácticamente la totalidad de los rasgos descritos en la tabla 2.2.

### 2.3.2. Aplicaciones Comerciales

Dada la naturaleza de las Soft Biometrics, además de su uso en sistemas de vídeo vigilancia, tienen un gran potencial en el ámbito comercial.

Actualmente ya existen numerosos avances en esta línea. En esta sección se van a describir dos ejemplos de APIs reales, actualmente en funcionamiento, hacia las que se ha conducido parte de la motivación de este proyecto, el cuál ha tomado como base los principios de dichas APIs pero los ha redirigido hacia atributos únicamente corporales.

La visión final de los resultados calculados por el estimador implementado en este trabajo comparte similitud con las aplicaciones ya presentes en el mercado que se describen a continuación: *Google Cloud Vision API* y *Microsoft Computer Vision API*.

## 1 - Google Cloud Vision API

Google Cloud Vision API permite comprender el contenido de una imagen gracias al encapsulado de modelos potentes de *machine learning* en una REST API (API cuya arquitectura está diseñada en la red) fácil de utilizar.

Esta API clasifica rápidamente imágenes en miles de categorías posibles (por ejemplo barcos, animales, construcciones, personas, etc). Además, detecta objetos y caras de forma individual en las imágenes.

El usuario que la utilice puede introducir metadatos en la imagen, moderar contenido ofensivo, o incluso analizarla a nivel sentimental.

En definitiva, las principales **funcionalidades** de la Google Cloud Vision API son:

- **Detección de etiquetas.** Esta API permite detectar amplios conjuntos de objetos en las imágenes, incluyendo miles de categorías de objetos que comúnmente se pueden encontrar en las imágenes.
- Detección de **Contenido inapropiado.** A través de Google SafeSearch, se puede moderar fácilmente el contenido de imágenes obtenidas de múltiples fuentes. La API permite detectar distintos tipos de contenido inapropiado, como puede ser adulto o violento.
- **Análisis Sentimental** de la cara. La API puede analizar atributos emocionales en las caras de las personas que aparecen en las imágenes, como por ejemplo alegría, tristeza o enfado. Este punto puede ser utilizado en tareas de marketing, ya que combinándolo con detección de objetos y detección de logos, se puede valorar cómo se puede sentir la gente con respecto a dicho logo.
- **Landmarks.** Puede detectar en la imagen estructuras populares, ya sean naturales o realizadas por el hombre (arquitecturas).
- **Extracción de texto.** Esta API, mediante OCR (*Optical Character Recognition*), permite detectar texto en las imágenes, junto con la identificación automática del idioma.
- **Atributos Generales** de la imagen. Puede detectar atributos genéricos, como por ejemplo el color predominante en la imagen.

En la figura 2.9 se puede observar un ejemplo concreto de la funcionalidad de esta API de Google. A partir de una imagen cualquiera seleccionada se puede extraer, tal y como se muestra en dicha figura y en este orden:

1. Detección de la cara en la imagen y emociones según los gestos localizados en la cara.
2. Etiquetas con las que se puede definir la imagen en su conjunto.
3. Rango de colores predominantes en la imagen.
4. Detección de contenido inapropiado.
5. Localización de todas las coordenadas de los puntos característicos de la imagen.

Ejemplos de extracción de landmarks y texto se pueden encontrar en la figura 2.11, para la que se ha utilizado una imagen que incluye estos dos atributos.

## 2 - Microsoft Computer Vision API

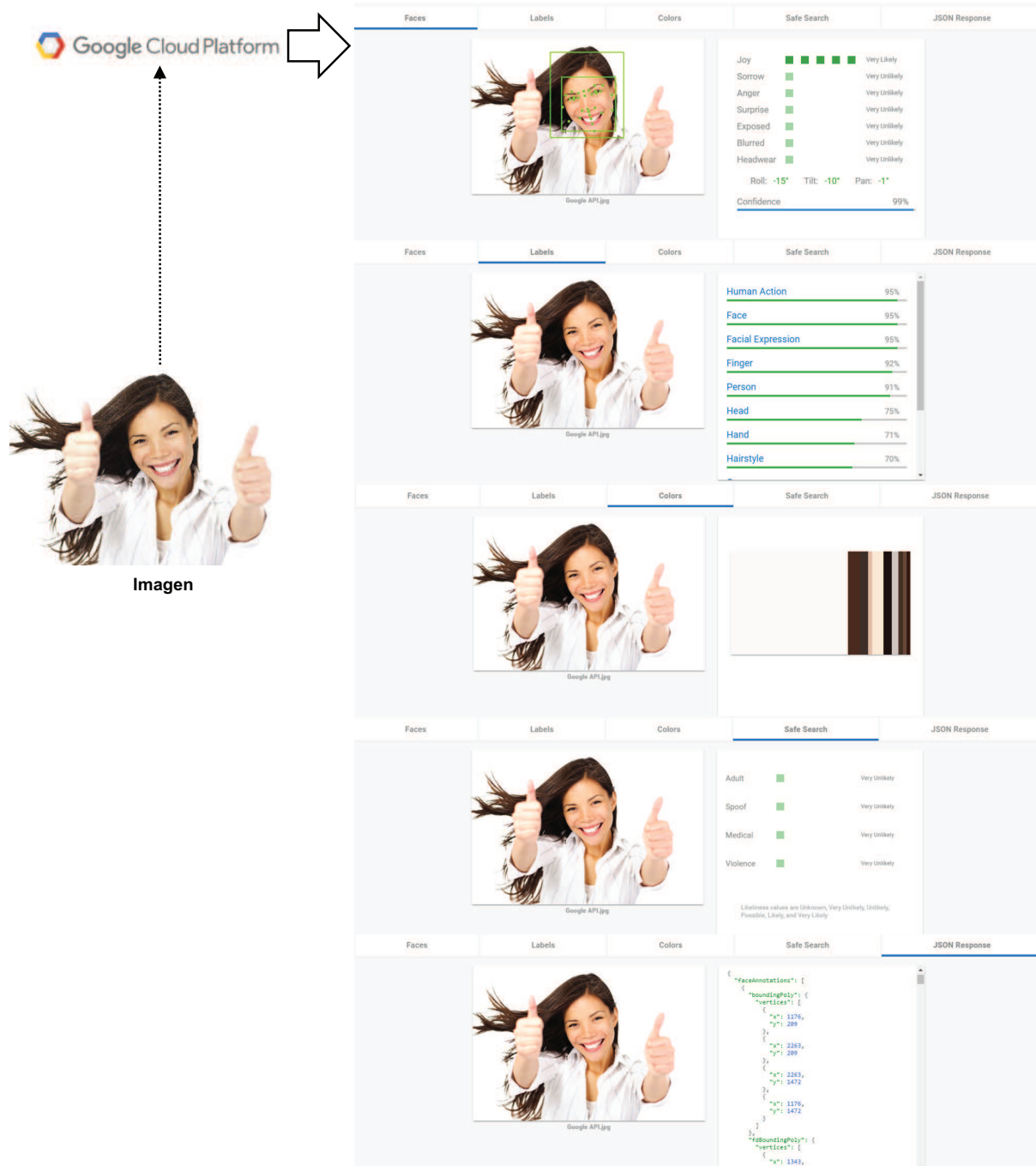


Figura 2.9: Google Cloud Vision API. Ejemplo de aplicación sobre una imagen. Fuente:[6]

Microsoft Computer Vision API [7] permite el acceso a algoritmos avanzados de procesamiento de imágenes y extracción de información a partir de ellas. Tiene muchos aspectos en común con Google Cloud Vision API, descrita en el punto anterior.

Los algoritmos de esta API pueden analizar el contenido visual de diferentes formas basándose en las imágenes que se introduzcan o las preferencias de los usuarios.

Las principales funcionalidades de Microsoft Computer Vision API son:

- **Generación de Descripciones.**

Los algoritmos de Computer Vision API analizan el contenido de la imagen, lo cual ge-

nera una descripción mostrada en lenguaje legible en forma de **frases completas**. Esta descripción resume aquello que se visualiza en la imagen.

Se generan varias diferentes, cada una evaluada con un score generado de fiabilidad. Se muestran ordenadas de mayor a menor fiabilidad.

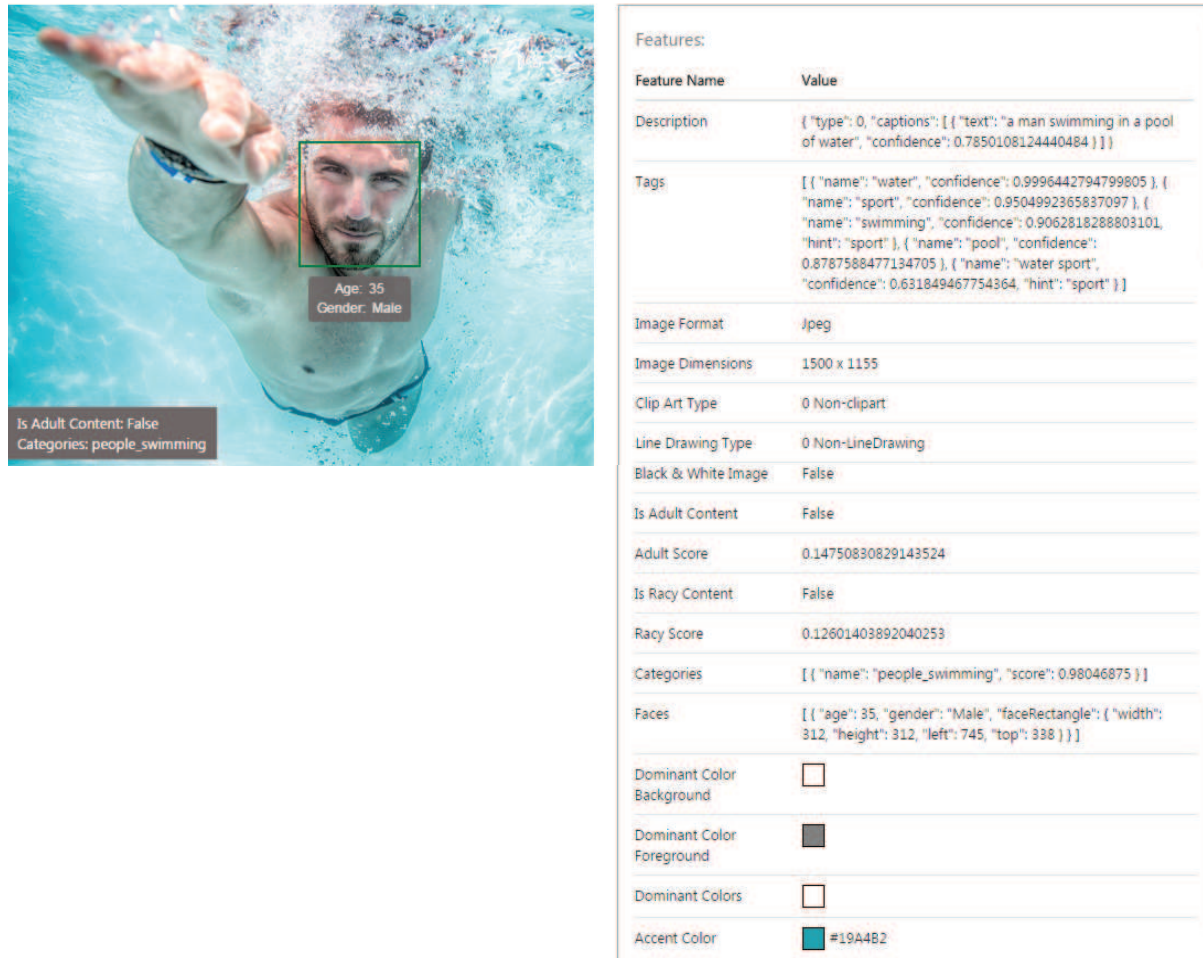


Figura 2.10: Microsoft Computer Vision API. Ejemplo de aplicación sobre una imagen. Fuente:[7]

- **Etiquetado de Imágenes.**

La API devuelve etiquetas basadas en más de 2000 objetos, escenarios y acciones reconocibles. Las etiquetas no están organizadas como una taxonomía ni existe jerarquía entre ellas.

Un conjunto de etiquetas construyen la base para la descripción de una imagen, es decir, a partir de una imagen dada, el algoritmo extrae etiquetas basadas en el contenido de la misma. Estas etiquetas no están limitadas al sujeto principal, como por ejemplo la persona que pueda aparecer en el fondo de la escena, sino que también incluye la situación (interior o exterior), muebles, herramientas, plantas, accesorios, etc.

- **Categorización de Imágenes.**

Además del etiquetado y descripción, Computer Vision API devuelve categorías basadas en un modelo jerárquico. Estas categorías están organizadas como una taxonomía con jerarquías hereditarias padre/hijo, es decir, que las características encontradas en una imagen se pueden categorizar partiendo de un conjunto más amplio y avanzando hacia uno más específico. Por ejemplo: Objeto - Comida - Pan.

Esta API cuenta con una lista de 86 conceptos categóricos.

- **Identificación del Tipo de Imagen.**

La API puede utilizar una flag para indicar si una imagen está en **blanco y negro** o en color, y puede utilizar el mismo método para indicar si se trata de un **dibujo** o no. Incluso puede definir si se trata de un clipart y, en ese caso, cuál es su calidad en una escala del 0 al 3.

- **Detección de caras.**

Es capaz de detectar las caras en una imagen y generar sus coordenadas, el rectángulo en el que se encuentra, **el género y la edad**. Estas características visuales son un subconjunto de metadatos generados para la cara.

- **Percepción de colores.**

El algoritmo es capaz de extraer el color predominante de la imagen, analizándolo en 3 contextos diferentes: fondo, frente y color general. Además, extrae el color más acentuado a la vista en la imagen, utilizando una mezcla de los colores predominantes y la saturación.

- **Detección de contenido inapropiado.**

Agrupar contenido adulto o racista, lo cual permite la detección de material relacionado con estas categorías y restringe su visualización. El umbral de detección de este contenido puede fijarse según la preferencia del usuario.

Todas estas funcionalidades se pueden observar en la figura de ejemplo 2.10.

Partiendo de las dos APIs descritas previamente, este proyecto se puede plantear de manera que tenga cabida en el alcance de la funcionalidad de cualquiera de estas aplicaciones comerciales.

Es decir, utilizando la estimación de **Soft Biometrics corporales** a partir de una imagen dada, el resultado se puede combinar para dar lugar a una **mejora de API comercial**, más completa y con un alcance mayor.

En la figura 2.11 se muestra cómo un estimador de atributos corporales podría **añadir valor** a una API de las previamente descritas. En concreto, se ha escogido Google Cloud Vision API como base para esta demostración.

Tal y como se aprecia en dicha figura, además de los resultados que ya se obtendrían con la API a día de hoy, tras la inclusión del estimador de features corporales podríamos obtener información acerca de:

- Altura
- Anchura de Hombros
- Ancho de Cadera
- Largo de los brazos
- Envergadura de la persona
- Color del pelo

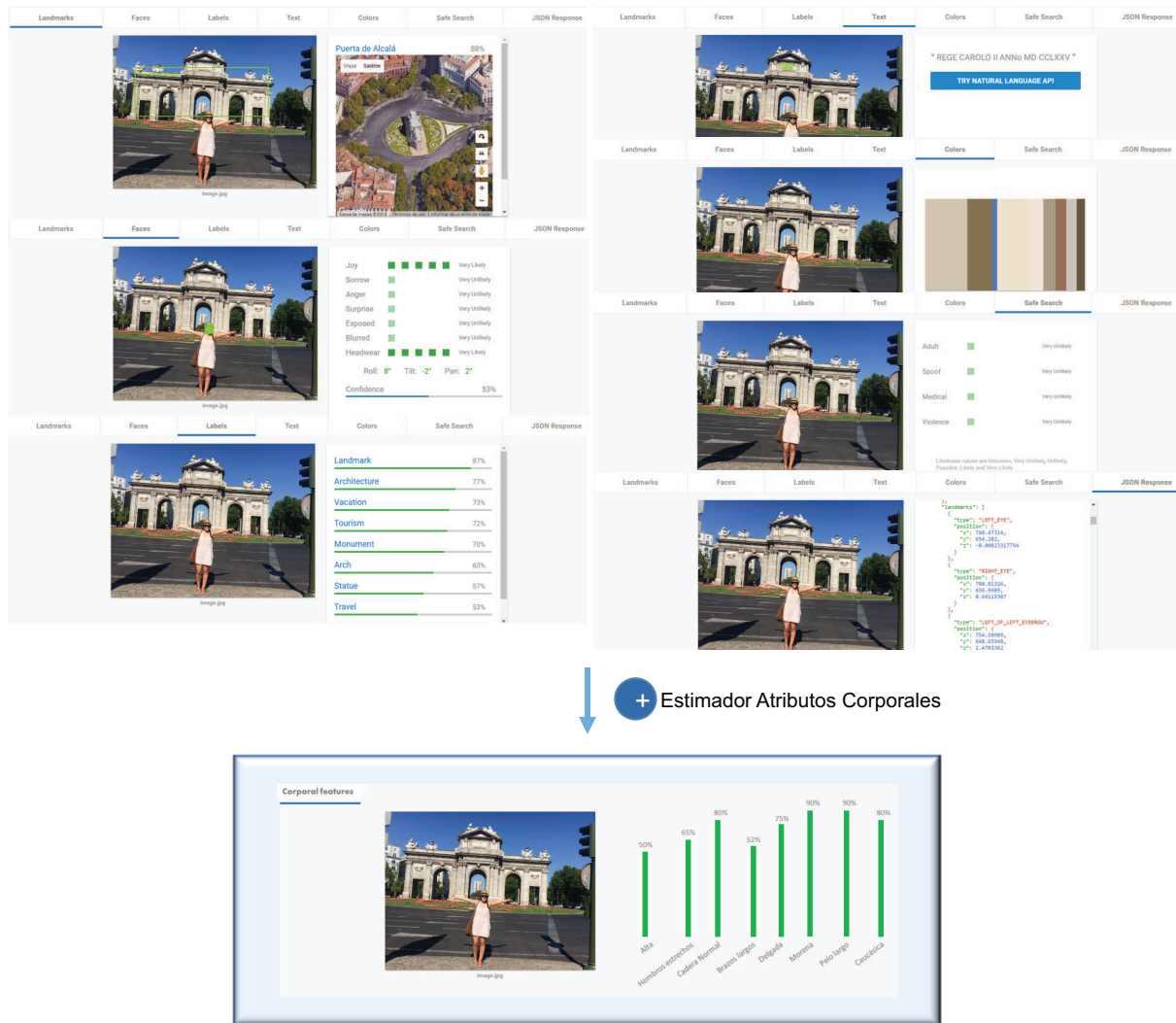


Figura 2.11: Objetivo de mejora en APIs tras añadir los resultados de un estimador de atributos corporales.

- Longitud del pelo
- Etnia, etc.

Para todos estos atributos se mostrará siempre la cualidad para la que se tenga una mayor probabilidad. Es decir, si la envergadura más probable para un individuo es 'grueso', o el color de pelo detectado es 'rubio', éstas serán las características que se mostrarán en los atributos corporales, en forma de porcentaje.

Por tanto, este proyecto conllevaría una mejora sustancial no sólo en **aplicaciones de vídeo vigilancia**, sino también en cuanto a **aplicaciones comerciales** se refiere.



## 3 Sistema, Diseño y Desarrollo

En este capítulo se describirá en detalle cómo se ha llevado a cabo la implementación del estimador de atributos corporales a distancia. Es decir, se describirá la parte más importante del proyecto, sobre la que luego se evaluarán los resultados.

El objetivo del estimador es conseguir, para una imagen dada, la extracción de sus medidas y características físicas de forma automática. De esta forma, para cada atributo corporal entrenado, se proporcionará un score que definirá la probabilidad de que el sujeto posea dicha característica en mayor o menor medida.

Es decir, el estimador de atributos corporales permitirá conocer de forma automática cuáles son los rasgos físicos estimados para cada usuario, mostrándolos como una medida de probabilidad.

La implementación de este proyecto se ha desarrollado siguiendo el diagrama mostrado en la imagen 3.1.

Partiendo del diagrama previo, a lo largo del capítulo se irá desarrollando de forma detallada el trabajo contenido en cada uno de los bloques que componen la implementación del estimador de atributos corporales.

### 3.1. Extracción de puntos Automática

---

El primer paso en el procesado del estimador es el que va a permitir extraer de forma automática todos los puntos característicos sobre la imagen de entrada.

Para llevarlo a cabo, primeramente se realiza una binarización de la imagen que extraiga únicamente la silueta del sujeto y, después, se extraen los puntos sobre esta imagen segmentada.

A continuación se explica cómo se ha realizado cada uno de estos procesos.

#### 3.1.1. Segmentado de la imagen

El objetivo del segmentado es obtener el frente de la imagen, es decir, únicamente la persona que se encuentra en la escena.

Para ello, partiendo del fondo y la imagen completa, tal y como se muestra en la figura 3.1, se ha llevado a cabo una técnica de *Background Subtraction* siguiendo los pasos descritos a continuación:

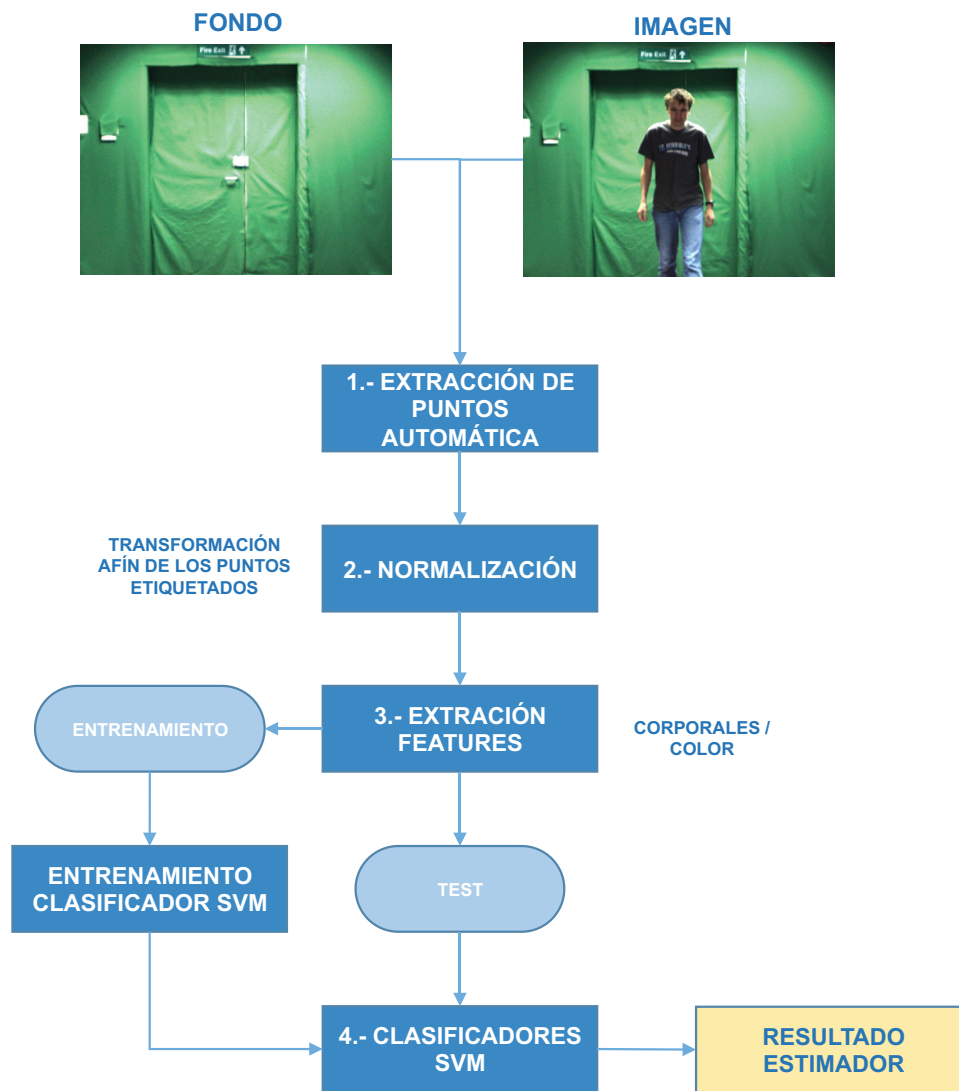


Figura 3.1: Diagrama de Implementación del Estimador de Atributos Corporales.

1. Para cada canal de color de la imagen (R, G, B), se realiza **una resta** entre el valor de los píxeles en la imagen de fondo y el valor de los píxeles en la imagen completa.

Como resultado se obtiene una **imagen con el fondo sustraído** y la figura del frente compuesta por la diferencia entre los píxeles de color.

2. Se convierte la imagen RGB obtenida en el paso anterior a **escala de grises**, es decir, se elimina la información de color.
3. A continuación se **binariza** la imagen, convirtiendo el valor de sus píxeles a 0 o 1. Para ello se ha establecido un umbral según el cual todos los píxeles que tengan un valor superior a él, quedarán en blanco. El resto, quedarán en negro.

El umbral seleccionado en este proyecto que mejor funcionaba para esta BBDD, ha sido  $Th = 20$ .

4. Se **elimina el ruido fino** de la imagen binarizada, es decir, los espurios menores. Para conseguirlo:
  - Paso 1: Se extrae una matriz conteniendo etiquetas para los objetos en la imagen según su conectividad. Se ha utilizado conectividad 8, es decir, para cada píxel se ha



buscado conectividad con todos los píxeles de su alrededor. Los píxeles etiquetados como 0, se corresponden con el fondo (no tienen valor en ningún píxel de su alrededor); los etiquetados como 1, con el primer objeto encontrado; los etiquetados como 2, con el segundo; etc.

En la figura 3.2 se puede observar cómo se realiza el procesado de imagen con conectividad 8, y cuál es su diferencia con la conectividad 4.

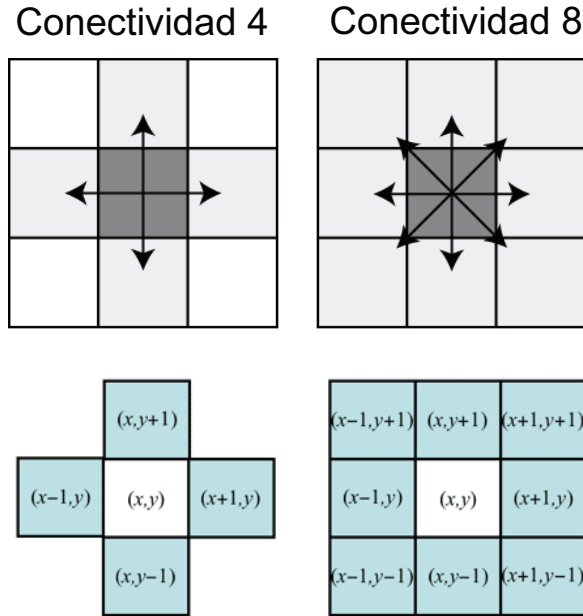


Figura 3.2: Vecindad de píxeles en procesamiento de imágenes: Conectividad 4 y 8.

- Paso 2: Se mide el área para cada componente conectado (cada objeto) encontrado en el paso anterior, calculándose como el número de píxeles contenido en su región.
- Paso 3: Se selecciona un umbral para el área de los objetos encontrados, que sea suficientemente discriminante como para eliminar el ruido fino (píxeles espurios). Se ha seleccionado  $Th = 4$ .

5. Se **elimina el resto de ruido** que pueda quedar en la imagen, como pueden ser pequeñas figuras, manchas o líneas más gruesas.

Para ello, para empezar, se repiten los dos primeros pasos que se han seguido para la eliminación de ruido fino: etiquetado de objetos según conectividad 8 y medida del área contenida en cada uno de ellos.

Una vez hecho esto, únicamente se selecciona el objeto encontrado con **mayor área**, el cual se corresponderá justamente con la silueta del sujeto.

En la figura 3.3 se puede observar el resultado de la aplicación de cada uno de los pasos del segmentado de la imagen, desde la imagen y fondo iniciales, hasta llegar a obtener el resultado final.

Tras la ejecución de todos estos pasos, se habrá conseguido segmentar la imagen y por tanto **obtener la ROI** (área de interés), sobre la que posteriormente se extraerán todos los puntos necesarios para la obtención de medidas de features corporales.

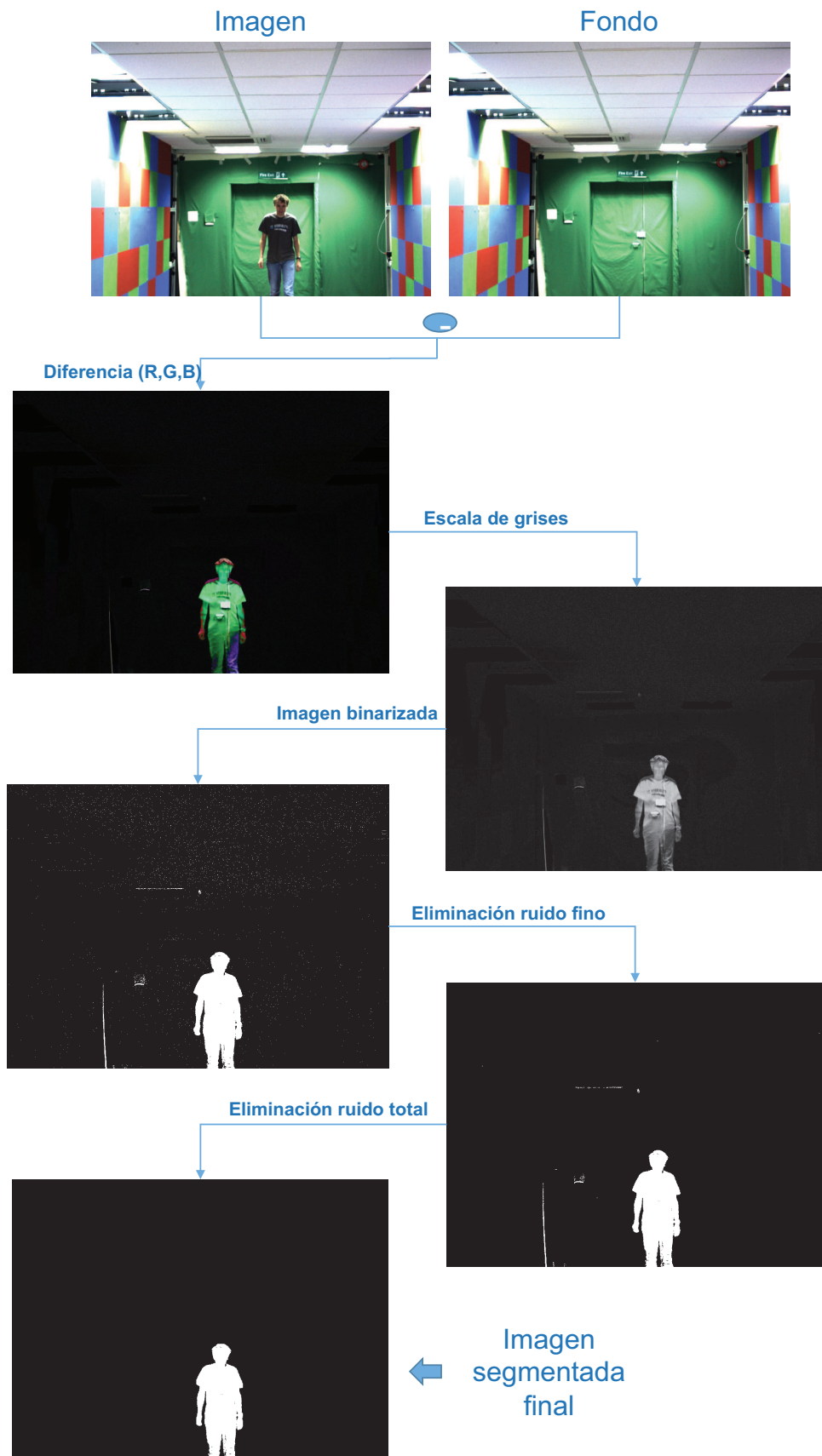


Figura 3.3: Resultados obtenidos en el proceso de Segmentado de la imagen.

### 3.1.2. Extracción de puntos

Una vez segmentada la imagen, se procede a la detección automática de los puntos que van a permitir extraer las medidas y atributos corporales del usuario.

Sin embargo, es importante destacar que no todos los atributos necesitan la extracción de coordenadas de puntos en la imagen. Esto ocurre, por ejemplo, para las features subjetivas, como pueden ser el color del pelo o la envergadura.

En la figura 3.4 se muestran los **puntos que se van a extraer de cada imagen**, necesarios para estimar más adelante los atributos corporales que dependen de estas coordenadas. La imagen que se puede ver en ella es un ejemplo de un usuario de la Base de datos utilizada en este proyecto.

Todos los puntos mostrados en dicha figura han sido extraídos tanto en el procesamiento automático de las imágenes, como en el procesamiento manual que se ha llevado a cabo en paralelo (descrito en la sección 3.5.).

Sin embargo, los codos son los únicos dos puntos que se han excluido del procesamiento automático, por no haber encontrado una proporción corporal lo suficientemente discriminante como para poder aplicársela a todos los usuarios y que extraiga correctamente los puntos en cuestión.



Figura 3.4: Puntos necesarios a extraer en la imagen para permitir estimar atributos corporales.

Como se puede observar en la figura 3.4, las imágenes de la Base de Datos de este proyecto están cortadas en la parte inferior, no muestran el cuerpo de manera completa. Sin embargo, dado que la altura se medirá como una relación de la distancia entre los dos puntos correspondientes y todas las imágenes se encuentran cortadas por el mismo lugar, no resultará ningún problema en la estimación de dicho atributo, ya que siempre será proporcional.

Para muchos de estos puntos se ha utilizado la localización del **centroide** correspondiente al frente en la imagen binarizada.

El centroide hace referencia al centro de masa de la región específica, y por tanto se trata

del primer punto que se extrae en este proceso. En la figura 3.5 se puede observar un ejemplo de la ubicación de un centroide en una imagen de estas características.

A continuación se describe en detalle la forma en la que se ha llevado a cabo la **extracción automática para cada uno de los puntos** indicados en la figura 3.4.

### 3.1.2.1. Altura

Para extraer el punto que va a marcar la altura del usuario, se procede de la siguiente forma:

1. Se extraen todas coordenadas de los puntos que se encuentran binarizados como frente en la imagen.
2. Para el conjunto de puntos anteriores, se localiza la coordenada 'y' mínima, es decir, la altura en la imagen a la que comienza la figura del usuario. Esta altura definirá el valor de la **coordenada 'y'** para el atributo de la altura.
3. Dado que pueden existir varios puntos con la misma coordenada 'y', es decir, varios puntos que van a marcar la altura del usuario, es necesario decidir qué coordenada 'x' se va a seleccionar. Para ello, siempre se tomará como referencia la **coordenada 'x'** del centroide.
4. Por último, para extraer el punto inferior que determinará la medida de la altura, se replicará la misma coordenada 'x' definida para el punto superior ya descrito. Para la coordenada 'y', se seleccionará automáticamente el máximo posible según el tamaño de la imagen en concreto.

El proceso de extracción de los puntos que determinan la altura se muestra en la figura 3.5.

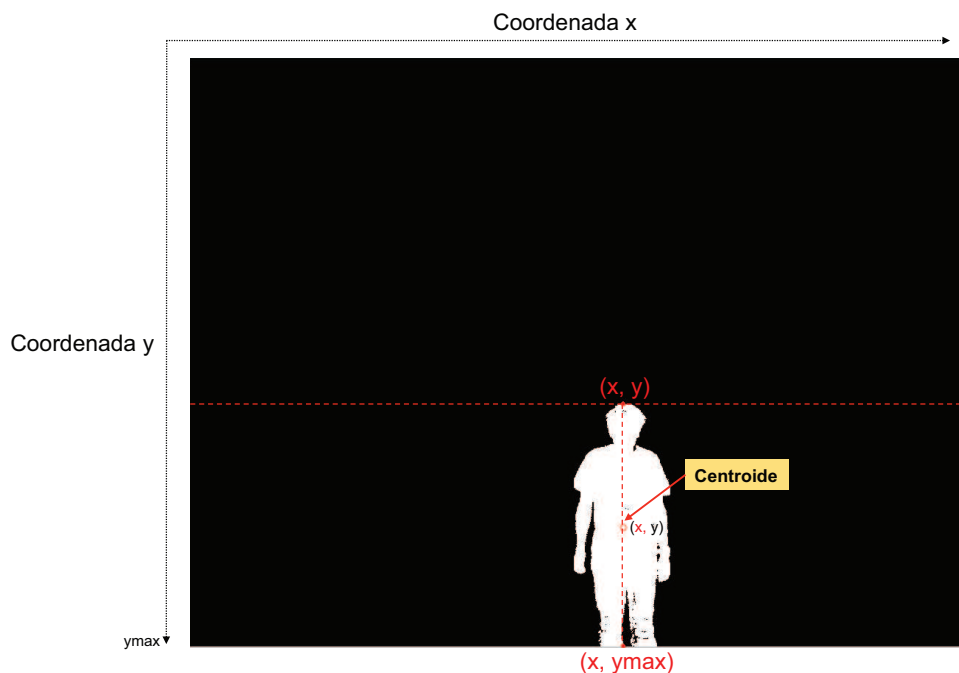


Figura 3.5: Extracción automática de puntos de Altura.

### 3.1.2.2. Ojos, Nariz y Barbilla

Debido a la magnitud del proyecto, y dado que actualmente existen APIs de código abierto que extraen correctamente los puntos de la cara, para la localización de los puntos de los ojos, la nariz y la barbilla se ha decidido utilizar el Sistema comercial basado en **API de Microsoft** explicado en el capítulo anterior (capítulo 2).

Con ello nos aseguramos de obtener unos resultados muy precisos para la extracción automática de estos puntos faciales.

Para poder hacer uso de dicha API de Microsoft, se ha utilizado un script desarrollado en lenguaje Python. Dicho script es capaz de extraer los puntos faciales detectados en una imagen, teniendo como entrada únicamente la localización (ruta) de dicha imagen.

Para cada imagen de los usuarios de la BBDD utilizada, se ha invocado al **APIRest de Microsoft**, la cual por medio de algoritmos de procesado de imagen devuelve un total de 27 landmarks/puntos faciales, con sus coordenadas x,y. Los 27 puntos faciales se obtienen en la forma: p[x,y].

Sin embargo, de todo este conjunto de puntos faciales solamente nos hemos quedado con 4 de nuestro interés:

- Ojo derecho.
- Ojo izquierdo.
- Centro de la nariz.
- Barbilla

Dichos puntos se han incorporado a nuestro sistema automático de etiquetado de puntos.

El script desarrollado en Python ha sido invocado desde Matlab, integrándolo en la globalidad de nuestro sistema automático. De esta forma, con una imagen de entrada conseguimos obtener todos los puntos descritos en la figura 3.4, sin necesidad de realizar procesamientos paralelos.

### 3.1.2.3. Hombros

Los siguientes puntos que se van a extraer son los correspondientes a los hombros del usuario. Para llevarlo a cabo se ha tomado como referencia el punto extraído para la **nariz del sujeto** y, a partir de él, se han desarrollado el resto de pasos descritos a continuación:

1. En primer lugar se realiza una **selección rectangular** sobre la imagen, que incluya el detalle de la zona en la que se encuentra cada uno de los hombros.

Partiendo de la nariz, se seleccionan **dos cuadrados** (uno para el hombro derecho y otro para el izquierdo) con una dimensión que permita incluir el hombro en cuestión. La medida que se ha utilizado para los lados de los cuadrados ha sido el doble de la distancia nariz-altura, suficiente para asegurar que siempre se incluirá la zona de interés para todos los usuarios.

2. A continuación se realiza, para cada cuadrado, una **suma en horizontal de los píxeles contenidos en la zona seleccionada**. De esta forma se obtiene un vector que indica la cantidad de píxeles pertenecientes al usuario en cada una de las filas del cuadrado de interés, lo cuál facilitará la búsqueda del hombro.

3. Sobre el vector extraído de suma de píxeles calculamos la **pendiente de la curva** para cada pequeña variación de píxeles. Es decir, en cada mínimo tramo vamos calculando la pendiente correspondiente entre los dos puntos. Dicha pendiente se ha calculado según la ecuación:

$$m = \frac{\Delta y}{\Delta x} = \frac{y_2 - y_1}{x_2 - x_1} \quad (3.1)$$

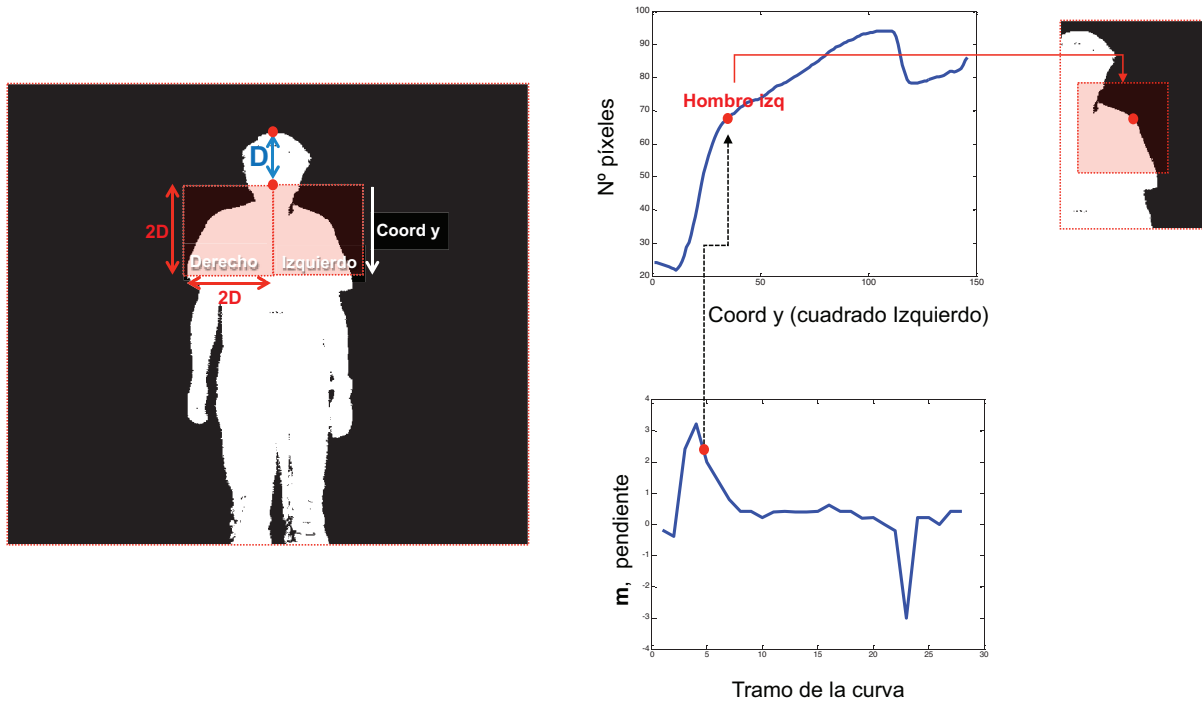


Figura 3.6: Extracción automática de puntos de Hombros.

4. El patrón que se calculará siempre será una pendiente que comienza siendo negativa (disminuyen los píxeles hasta llegar al cuello, equivalente a un mínimo local) y, a partir de ahí, comienza a crecer positivamente según se va bajando hacia el hombro, ya que el número de píxeles va aumentando cada vez más.

Al llegar al hombro la pendiente comenzará a disminuir. Esto se explica porque, aunque el número de píxeles siga aumentando, no lo hace de forma tan rápida como el recorrido cuello-hombro.

Por ello, seleccionaremos el punto en el que la pendiente pasa de ser la mayor de todas a disminuir su valor. Ése será el **punto extraído para el hombro**.

En la figura 3.6 se puede observar un esquema del proceso descrito, en el que se visualiza cómo se lleva a cabo la extracción del punto para uno de los hombros.

#### 3.1.2.4. Caderas

Los siguientes puntos que se van a extraer son los pertenecientes a las caderas del usuario.

Para ello, se ha recurrido al análisis antropométrico de Alphonse Bertillon [16] y Andrew Loomis [8], en el cual se pueden ver las proporciones del cuerpo humano medidas por dichos autores.

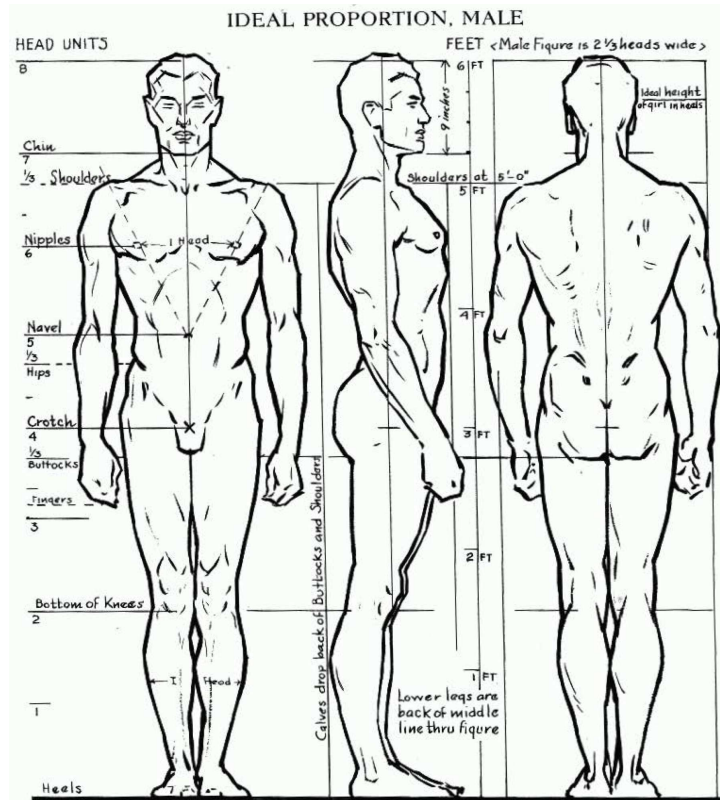


Figura 3.7: Proporciones antropométricas de la figura humana. Fuente: Andrew Loomis [8].

Atendiendo a las **proporciones corporales** observadas en la figura 3.7, la extracción de los puntos de la cadera se ha llevado a cabo de la siguiente forma:

1. Partimos del punto extraído para el máximo de la altura (el final de la cabeza) y del punto extraído para la barbilla.

Con estos dos puntos calculamos la **distancia del límite de la cabeza a la barbilla**, la cual nos va a permitir localizar la altura aproximada de la cadera del usuario.

2. Calculamos la **altura a la que se encontrarán las caderas** a partir de la barbilla, y sumando dos cabezas y media por debajo de ella, tal y como se puede ver en la figura 3.7. Es decir, la 'coordenada y' para la cadera se calculará mediante la ecuación:

$$Altura_{Cadera} = Altura_{Barbilla} + (Distancia_{Barbilla/Cabeza} * 2,5)$$

En el caso de un niño pequeño, esta operación deberá realizarse multiplicando por un factor de 2, en lugar de por 2.5, siguiendo de nuevo las proporciones dadas por las referencias anteriores.

3. Una vez obtenida la altura de las caderas, únicamente falta conocer cuáles son las **dos coordenadas horizontales**, coordenadas x, para los dos puntos correspondientes a la cadera derecha y cadera izquierda.

Para ello, lo primero que llevamos a cabo es la **extracción de la silueta** del sujeto, dibujando los bordes de la misma y conociendo así dónde delimita la cadera.

4. Partiendo de la coordenada x del centro del cuerpo, es decir, del **centroide**, se avanza hacia la derecha y hacia la izquierda sobre la línea de altura de la cadera calculada en pasos previos.

El objetivo es encontrar en esta línea y en ambos sentidos, el primer punto de la silueta en cada dirección.

De esta forma, al encontrar la **intersección de la silueta con la altura de la cadera**, ya obtendremos las coordenadas completas de los dos puntos de la cadera.

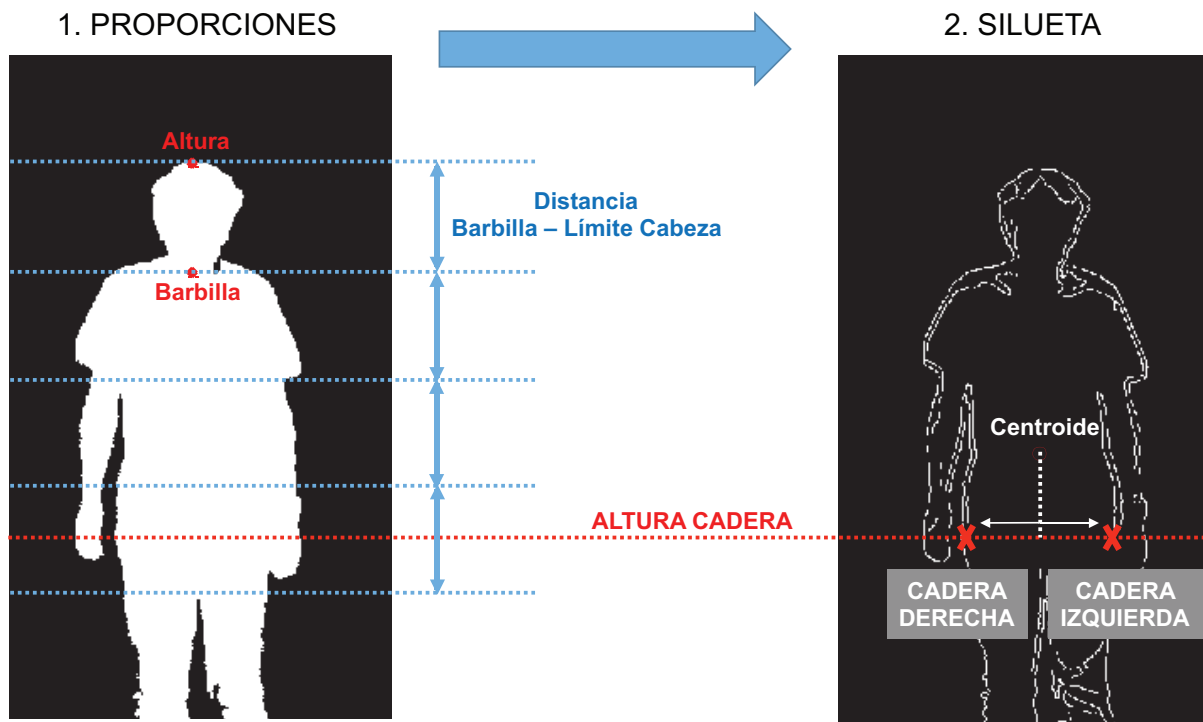


Figura 3.8: Extracción automática de puntos para la cadera: derecha e izquierda.

La extracción del punto de la cadera conllevará un pequeño error para los pocos casos en los que el usuario tenga los brazos completamente pegados al cuerpo.

En estos casos no se marcaría la silueta de la cadera y por tanto el primer punto de la silueta que se encontraría avanzando desde el interior del cuerpo hacia los bordes externos sería directamente el del brazo, con lo cual se calcularía una cadera más ancha.

Sin embargo, en una postura relajada o en movimiento, lo más común es que la gran mayoría de los usuarios tengan los brazos mínimamente despegados del cuerpo, con lo cual no es un problema relevante en este proyecto.

De todas formas se ha previsto la posibilidad de que se de este problema. En este caso se haría un segundo procesado en el que el borde de la silueta se extraería de una forma más fina, permitiendo diferenciar mejor la posible separación de la cadera al brazo y por tanto arreglando el problema.

El proceso que se lleva a cabo para la extracción de los puntos de la cadera se puede observar en el esquema de la figura 3.8.

### 3.1.2.5. Manos

Para extraer los puntos que marcan dónde finalizan las manos y, por tanto, los brazos, se ha utilizado un procesado similar al de la extracción de los puntos de los hombros.

Es decir, de nuevo se utiliza una curva que indique el número de píxeles de una región seleccionada y la pendiente de dicha curva para localizar puntos de inflexión.



De forma más concreta, los pasos que se han llevado a cabo han sido los siguientes:

1. En primer lugar se realiza una **selección rectangular** sobre la imagen, que incluya el detalle de la zona en la que se encuentra cada una de las manos.

Partiendo del **centroide** de la imagen, se seleccionan **dos cuadrados** (uno para la mano derecha y otro para la izquierda) con una dimensión que permita incluir la mano en cuestión.

La medida que se ha utilizado para los lados de los cuadrados ha sido justamente la distancia entre los hombros del usuario, suficiente para asegurar que siempre se incluirá la zona de interés para todos los sujetos.

2. Tal y como se ha realizado para los hombros, a continuación se realiza, para cada cuadrado, una **suma en horizontal de los píxeles contenidos en la zona seleccionada**. De esta forma se obtiene un vector que indica la cantidad de píxeles pertenecientes al usuario en cada una de las filas del cuadrado de interés, lo cuál facilitará la búsqueda del final de la 'coordenada y' la mano.

Para el caso de las manos también realizaremos una **suma en vertical de los píxeles contenidos en la zona seleccionada**, de tal forma que podamos extraer la 'coordenada x' del final de la mano.

Para los hombros no fue necesario realizar la suma en vertical, puesto que teniendo la 'coordenada y' podíamos extraer fácilmente la 'coordenada x', dado que los hombros forman una línea continua en el cuerpo.

Sin embargo, dado que las manos normalmente están separadas mínimamente del torso, la extracción de la 'coordenada x' no es trivial, podría caer justamente en el hueco entre el torso y la mano. Por ello esta coordenada necesita un procesado al margen.

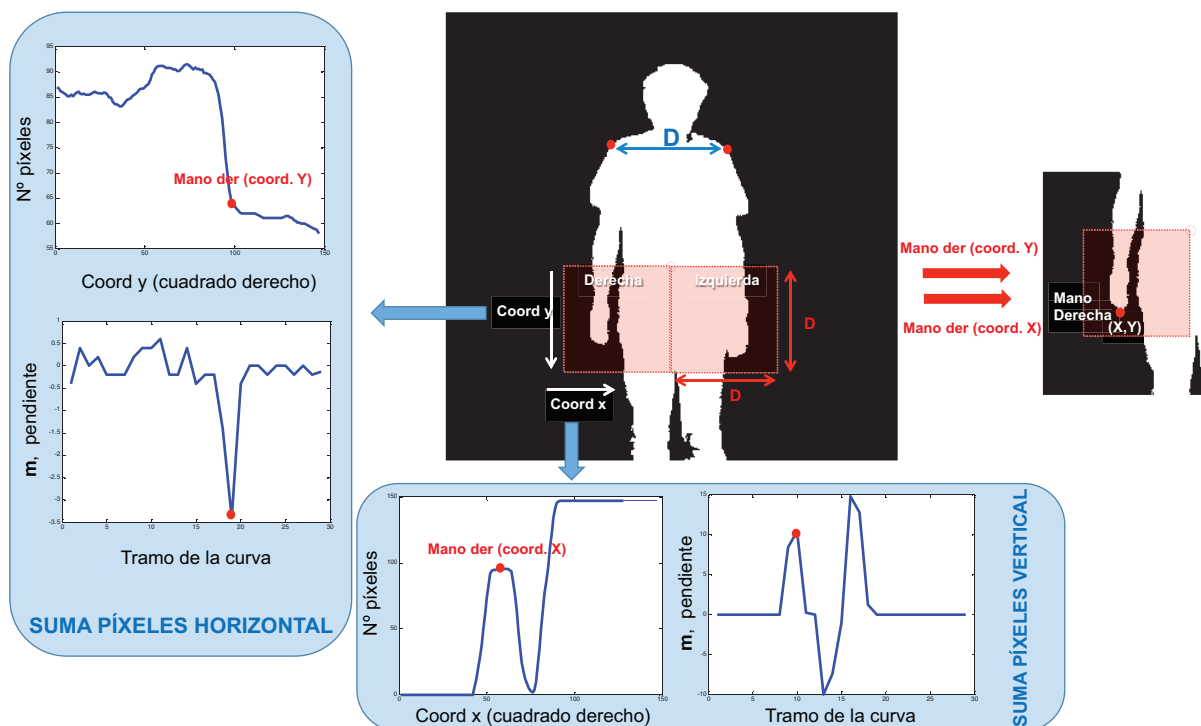


Figura 3.9: Extracción automática de puntos del final de las Manos.

3. También repitiendo el mismo paso que se aplicó sobre los hombros, sobre ambos vectores extraídos de suma de píxeles calculamos la **pendiente de la curva** para cada pequeña variación de píxeles. Es decir, en cada mínimo tramo vamos calculando la pendiente correspondiente entre los dos puntos mediante la ecuación:

$$m = \frac{\Delta y}{\Delta x} = \frac{y2 - y1}{x2 - x1} \quad (3.2)$$

4. Para el cálculo de la 'coordenada y' se utilizará la pendiente de la suma en horizontal de los píxeles.

El patrón que se calculará siempre, independientemente de cómo empiece, será una pendiente que en algún punto desciende de una forma muy notable. Es decir, será una pendiente muy pronunciada y negativa, lo cual se corresponderá con el final de la mano. Por ello, para extraer la '**coordenada y**', se buscará la **pendiente mínima** calculada sobre la curva.

Para el cálculo de la 'coordenada x' se utiliza la pendiente de la suma en vertical de los píxeles.

En este caso, si se trata de la mano derecha, siempre tendremos un primer máximo local, que se corresponderá justamente con el final de la mano, y un segundo y permanente máximo en la curva que será el cuerpo del usuario una vez superada la zona de la mano.

Por ello y para la mano derecha, para obtener la '**coordenada x**' se seleccionará el punto correspondiente al **primer pico (máximo) de la pendiente** calculada sobre la curva. Para la mano izquierda se traspone el vector y, de esta forma, también coincidirá con el primer pico de la pendiente.

De esta forma ya se habrá obtenido los dos puntos correspondientes al final de las manos.

En la figura 3.9 se indica cómo se realiza la extracción del punto de la mano derecha para un usuario de ejemplo, siguiendo el procesamiento de imagen descrito.

## 3.2. Normalización

---

Una vez extraídos los puntos necesarios de la imagen, se deben normalizar con respecto a una imagen referencia, de tal forma que siempre podamos tratar con una **misma distancia del usuario a la cámara**.

Es decir, las coordenadas de un punto etiquetado en la imagen inicial  $\mathbf{P} = (\mathbf{x}, \mathbf{y})$  serán normalizadas y calculadas para un nuevo espacio en el que el usuario se encuentre a una distancia de la cámara común para todos, obteniendo por tanto un nuevo punto  $\mathbf{P}' = (\mathbf{x}', \mathbf{y}')$ .

Gracias a esta normalización, las medidas corporales que se calculen como una distancia entre dichos puntos serán acordes al resto de medidas de los usuarios de la BBDD. Así, se evitarán errores de medida como podría ser, por ejemplo, tener dos usuarios a distinta distancia de la cámara y considerar más delgado al que se encuentra más lejos (aunque no fuera así realmente).

Para ello, en este proyecto se ha llevado a cabo una **Transformación Afin**, es decir, una transformación lineal de las coordenadas de los puntos, la cual se calcula de la siguiente manera:

$$P' = A * P$$

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = A * \begin{bmatrix} x \\ y \end{bmatrix}$$

donde A es una matriz 2x2, también denominada la pendiente de la transformación.

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} * \begin{bmatrix} x \\ y \end{bmatrix}$$

Geoméricamente, una transformación afín en un espacio euclídeo es una transformación que preserva:

- Las relaciones de **colinealidad entre puntos**, es decir, puntos que recaen sobre una misma línea antes de la transformación, siguen recayendo en ella tras la misma.
- Las **razones entre distancias** a lo largo de una línea, es decir, la relación entre distancias de puntos siguen manteniéndose después de la transformación.

Las transformaciones afines más comunes son el escalado, rotación y cizallado. A continuación se muestra cómo sería la matriz A para cada una de estas transformaciones:

$$\text{Escalado} : A = \begin{bmatrix} s_1 & 0 \\ 0 & s_2 \end{bmatrix}$$

$$\text{Cizallado} : A = \begin{bmatrix} 1 & sh_1 \\ sh_2 & 1 \end{bmatrix}$$

$$\text{Rotacion} : A = \begin{bmatrix} \cos(\theta) & -\text{sen}(\theta) \\ \text{sen}(\theta) & \cos(\theta) \end{bmatrix}$$

donde s1 y s2 son los factores de escala según los ejes X e Y respectivamente, sh1 y sh2 son los factores de cizalla paralela a los ejes X e Y respectivamente y donde  $\theta$  es el ángulo a rotar.

Los resultados que se obtendrían sobre una imagen aplicando cada una de las transformaciones indicadas, se muestra en la figura 3.10.

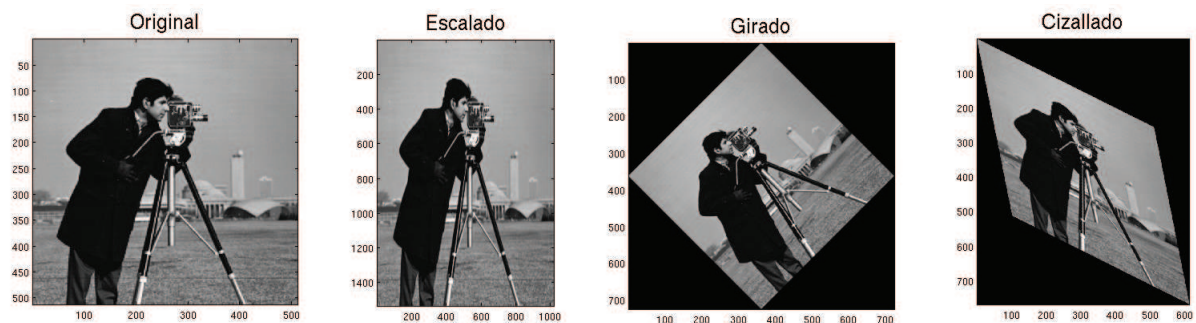


Figura 3.10: Transformaciones afines aplicadas sobre una imagen: escalado, cizallado y rotación. Fuente: [9]

En este trabajo únicamente se ha aplicado la **transformación afín de Escalado sobre ambos ejes**, ya que ha sido la única necesaria para poder normalizar las coordenadas de los

puntos etiquetados. Se ha asumido que todas las imágenes de la BBDD se han tomado con un mismo ángulo de giro de la cámara.

De forma más concreta, el **procesado de la normalización** de los puntos extraídos de la imagen se ha llevado a cabo de la siguiente manera:

1. Se han seleccionado dos distancias de referencia (una horizontal y otra vertical) con respecto a las que poder normalizar todas las imágenes.

Estas distancias de referencia han sido la distancia entre los ojos (**Referencia horizontal**) y la distancia de los ojos a la nariz (**Referencia vertical**), asumiendo que son medidas suficientemente iguales para todos los usuarios, y por tanto fiables para hacer la normalización respecto a ellas.

2. Se ha seleccionado un valor numérico para estas distancias de referencia a partir de los resultados obtenidos en el procesado manual inicial (desarrollado en el punto 3.5 de este capítulo).

Los **valores numéricos utilizados** en el proyecto han sido:  $d_{REF_H} = 25$  y  $d_{REF_V} = 12,5$ , para la normalización horizontal y vertical, respectivamente.

3. Para cada punto etiquetado en la imagen en el proceso anterior (etiquetado automático), se normalizan sus coordenadas con respecto a  $d_{REF_H}$  y  $d_{REF_V}$ .

Para ello se ha utilizado la **matriz de Escalado** utilizando dos factores de escala, horizontal y vertical, y aplicándolos sobre las coordenadas de los puntos:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} s_h & 0 \\ 0 & s_v \end{bmatrix} * \begin{bmatrix} x \\ y \end{bmatrix}$$

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} s_h * x \\ s_v * y \end{bmatrix}$$

donde los factores de escala dependen de las distancias de referencia  $d_{REF_H}$  y  $d_{REF_V}$ , y las distancias entre los mismos puntos (ojos y ojos-nariz) en la imagen actual  $d_H$  y  $d_V$ :

$$s_h = d_{REF_H}/d_H$$

$$s_v = d_{REF_V}/d_V$$

A modo resumen, se puede ver un esquema del proceso de normalización llevado a cabo en la figura 3.11.

Una vez aplicados todos estos pasos, hemos obtenido la totalidad de los puntos normalizados, y ya se podría proceder a la extracción de medidas corporales.

Esta normalización no solamente es útil para las imágenes de la BBDD utilizada en el proyecto, sino que permite mejorar la extracción de atributos en **imágenes de otras BBDD** que se puedan introducir en el estimador, para las que a priori no se sepa la distancia del sujeto a la cámara.

Por tanto, se trata de un paso clave en el proyecto.

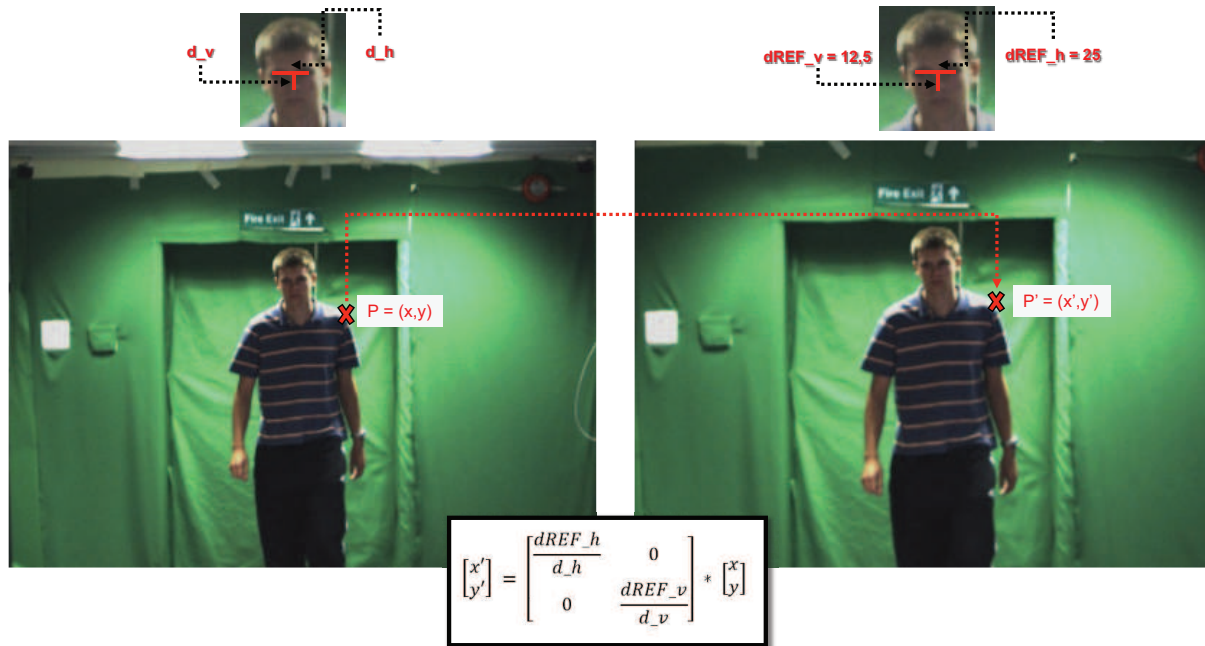


Figura 3.11: Normalización horizontal y vertical llevada a cabo sobre las coordenadas de un punto extraído en la imagen.

### 3.3. Extracción de features

Una vez normalizadas todas las coordenadas de los puntos característicos extraídos, el siguiente paso es conseguir una medida de los atributos corporales.

Para aquellas que dependen directamente de las coordenadas de los puntos etiquetados en la imagen (osease, **features objetivas**), la **extracción de la feature** se obtendrá mediante el cálculo de la **distancia euclídea** entre dichos puntos.

El resto de features son **atributos subjetivos**, y por tanto no se pueden extraer a partir de únicamente las coordenadas de algunos puntos en la imagen.

Los atributos subjetivos que se han extraído de forma automática sobre los usuarios son **la envergadura y el color del pelo**. Para este conjunto se ha procedido de manera diferente, como se indica más adelante.

El motivo por el cual únicamente hemos seleccionado la envergadura y color del pelo y se ha descartado la extracción de otras Soft Biometrics corporales subjetivas, como lo es la edad o el género, se basa en que éstos ya son extraídos por la API de Microsoft, tal y como se ha detallado en la Sección 2.3.2. De forma concreta, en la figura 3.13 se muestra cómo esta API extraería la edad y el género sobre los sujetos de la BBDD.

Adicionalmente, al margen de las features basadas en distancia o las features específicas para el color del pelo y la envergadura, en este proyecto se ha extraído una feature del estado del arte que se puede aplicar a todos los atributos corporales. Esta **feature está basada en la forma del cuerpo**, y es una medida continua que depende de la silueta del sujeto.

A continuación se describen todas las features extraídas en este proyecto.

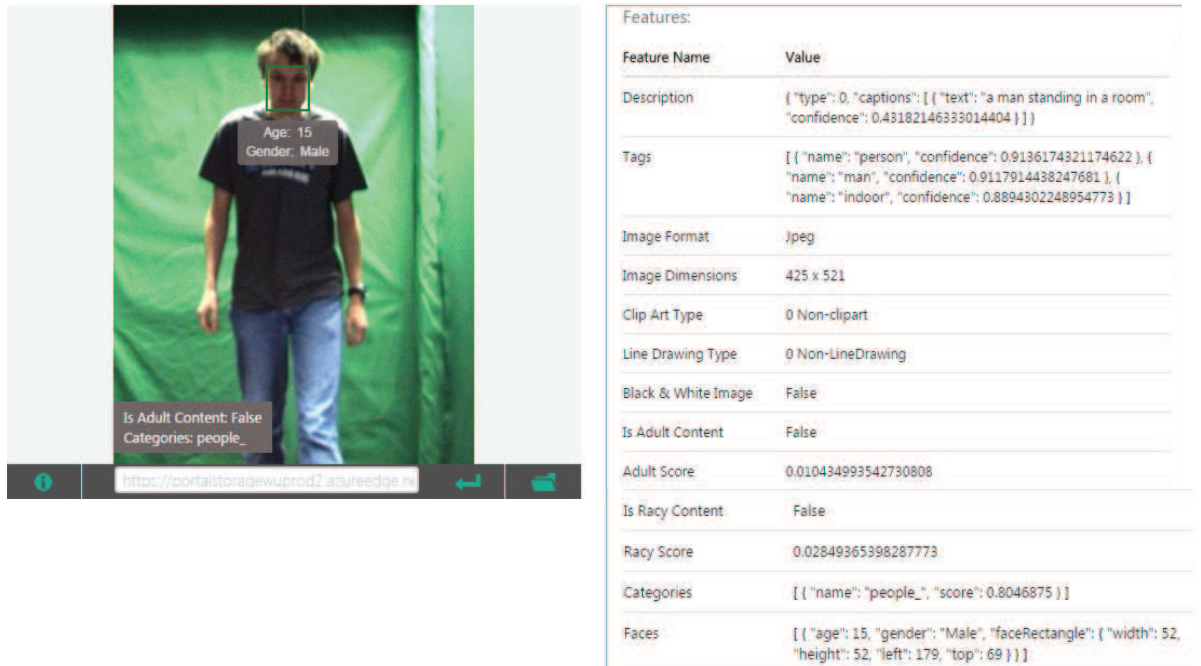


Figura 3.12: Soft Biometrics corporales que ya son extraídas por la API de Microsoft: edad y género.

### 3.3.1. Atributos objetivos: Distancia entre Puntos

La distancia euclídea entre dos puntos  $P_1$  y  $P_2$  se calculará según la ecuación representada en 3.3.

$$d_E(P_1, P_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (3.3)$$

A continuación se mencionan los atributos corporales que **se calcularán a partir de la distancia euclídea** de los puntos previamente extraídos en la imagen segmentada.

Además, se indica sobre qué puntos, referenciados en la figura 3.4, se realiza el cálculo de la distancia.

- **Altura.** Punto 1 y Punto 14.
- **Ancho de los hombros.** Punto 6 y Punto 7.
- **Ancho de la cadera.** Punto 11 y Punto 12.
- **Largo de brazo derecho.** Punto 6 y Punto 10, recordamos que en el caso del etiquetado automático no se ha extraído el punto perteneciente al codo. Por tanto la distancia se calcula directamente del hombro a la mano.
- **Largo de brazo izquierdo.** Punto 7 y Punto 13.

Tras realizar el cálculo de las distancias entre los puntos mencionados, ya se obtiene la medida para cada uno de los atributos corporales objetivos necesarios para el estimador.

Los demás puntos de la imagen 3.4 no se han utilizado para extraer medidas de atributos, sino que su cometido ha sido otro diferente.

Por ejemplo, los puntos 2 y 3, se han utilizado para normalizar la imagen en horizontal. El punto 4 se ha utilizado para normalizar la imagen en vertical, conjugándose con los puntos 2 y 3. Por último, el punto 5 se ha utilizado para extraer los puntos de la cadera mediante proporciones corporales, tal y como se ha explicado en la sección de extracción de puntos.

### 3.3.2. Atributo subjetivo: Envergadura

Este atributo permite definir a una persona según su complejión: si es delgada, normal, o gruesa.

Para definir este atributo subjetivo, únicamente se han llevado a cabo dos pasos:

1. Para empezar, se ha realizado la **suma de todos los píxeles** binarizados como frente en la imagen.
2. Con el objetivo de tener una medida de la envergadura **normalizada**, el número de píxeles calculado en el punto anterior se divide entre el **tamaño de la imagen**, es decir, se obtendrá este atributo como un porcentaje.

Con este paso conseguimos evitar que una persona que se encuentre más cerca de la cámara sea considerada más gruesa o que una que se encuentre más alejada sea considerada más delgada.

Consideramos que el tamaño de la imagen siempre sea igual para todas las posibles imágenes de sujetos que se introduzcan en el estimador.

Por tanto, la ecuación a aplicar para el **cálculo del porcentaje** de envergadura es la siguiente:

$$\%Envergadura = \frac{numpíxeles}{h * w} * 100\% \quad (3.4)$$

donde  $h$  y  $w$  son las dimensiones del alto y ancho de la imagen.

### 3.3.3. Atributo subjetivo: Color del pelo

Para extraer la información de color del pelo que nos permita definir si el usuario es **rubio, castaño o moreno**, se siguen los siguientes pasos de forma automática:

1. A partir del punto que marca la altura del usuario, se realiza una selección de **una pequeña área** por debajo del mismo.

El cuadrado de selección que se ha elegido ha sido de 10 píxeles de lado. De esta forma, será lo suficientemente grande como para poder seleccionar **parte del cabello del usuario** pero no tanto como para poder introducir otro tipo de información de color que puede provenir de sombras o cualquier otro aspecto no deseado.

2. Una vez que se ha seleccionado esta pequeña área de píxeles de cabello, se calcula el valor RGB medio de todos los píxeles contenidos en ella.

De esta forma tenemos el valor RGB de un píxel que nos sirve como referencia para tomar la decisión de si la persona es rubia, castaña o morena.

Como referencia, los valores RGB aproximados para cada tipo de cabello son:

- Rubio: RGB = (250, 240, 190)

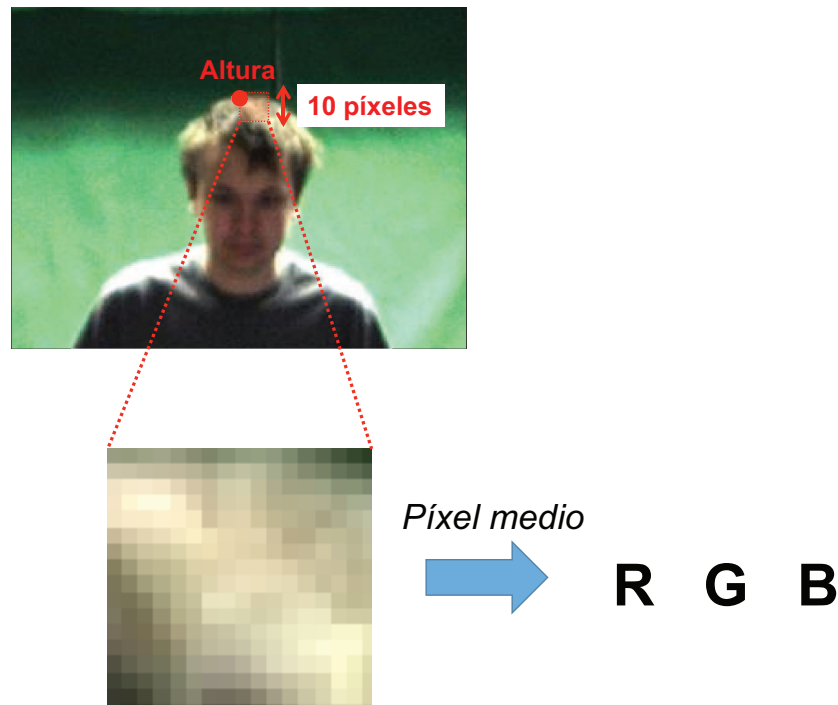


Figura 3.13: Extracción automática de una medida para el Color de Pelo.

- Castaño: RGB = (150, 75, 0)
- Moreno: RGB = (49, 50, 44)

En conclusión, el color del pelo se extraerá como una feature con 3 dimensiones: R, G y B. Esto será lo que nos ayude a tomar la decisión sobre el tipo de cabello más adelante.

En la imagen 3.13 se muestra un pequeño esquema sobre el procedimiento llevado a cabo para la obtención de una medida para el color del pelo del usuario, según lo descrito en este apartado.

### 3.3.4. Forma del cuerpo (*Body Shape*)

Por último se ha incluido una **feature basada en la forma del cuerpo**, utilizando el método desarrollado y explicado en la Sección 2.2.3, en la que se exponía un trabajo relacionado con las Soft Biometrics y la estimación del género del individuo [3].

Esta feature se podrá aplicar a diferentes atributos con **independencia de si son objetivos o no lo son**, siempre y cuando tengan relación con la silueta corporal.

Para este proyecto se ha utilizado como base el algoritmo para la forma del cuerpo expuesto en el trabajo de referencia mencionado, modificando algunos detalles para que se ajuste mejor a la estimación de Soft Biometrics corporales.

El procesado llevado a cabo para extraer esta feature de *Body Shape* ha sido el siguiente:

1. Partiendo de la imagen original segmentada del sujeto, se normaliza en base a la distancia entre los ojos, para que todos los individuos tengan la misma distancia entre ellos. Es decir, que todos los ojos de todos los sujetos de la BBDD estarán alineados y en la misma posición para poder comparar entre sí correctamente la forma de sus cuerpos.



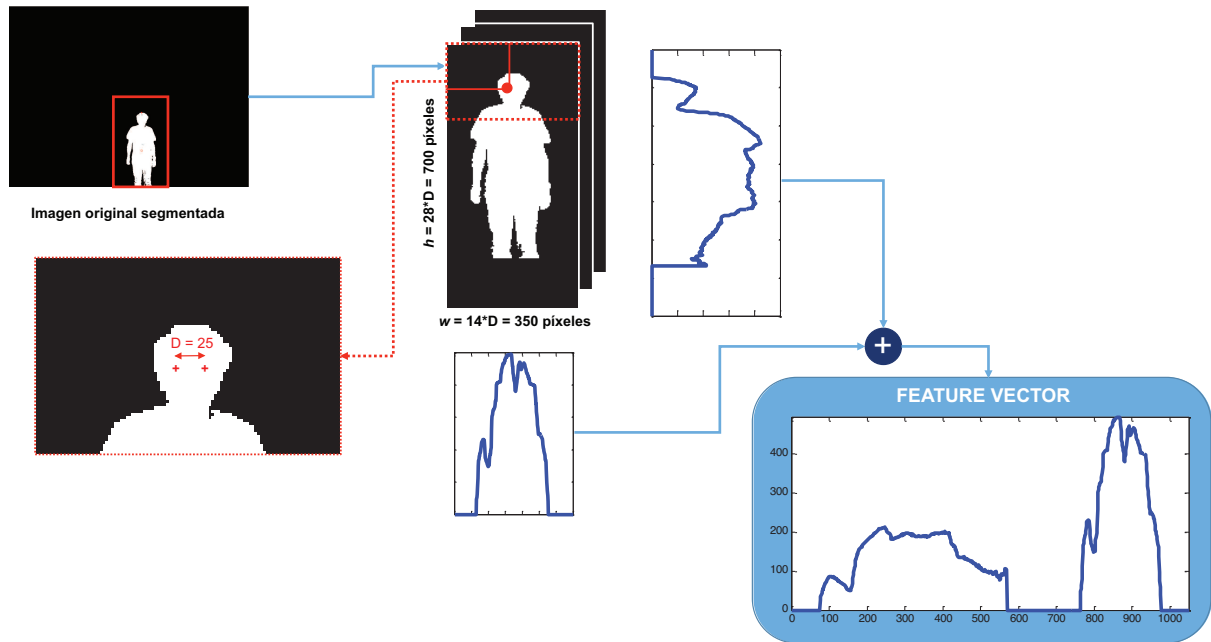


Figura 3.14: Procesado de extracción automática de una feature para la Forma del Cuerpo.

La distancia entre los ojos que se ha seleccionado ha sido  $D=25$  píxeles.

El algoritmo de normalización llevado a cabo ha sido el que se ha descrito en la Sección 3.2, realizando una transformación afín de las imágenes.

2. Se recorta un rectángulo, tomando como referencia el ojo derecho de cada sujeto, de forma que todos ellos tienen los ojos en la misma posición.

La dimensión del rectángulo será de  $h \times w$ , donde  $h = 28 * D$  y  $w = 14 * D$ . Estas dimensiones se han fijado en base a las proporciones corporales (ver figura 3.7), de tal forma que el rectángulo siempre contenga la figura de todos los sujetos, desde el más delgado al más grueso.

3. A continuación, se extrae un vector con la suma de los píxeles de frente según las filas, y otro vector con la suma de los píxeles según las columnas.
4. Estos dos vectores de suma de píxeles en horizontal y vertical se concatenan para obtener el **vector final de la feature** de la forma del cuerpo.

Este procesado al completo se puede ver en la figura 3.14.

## 3.4. Clasificadores

### 3.4.1. SVM (Support Vector Machines)

Las máquinas de vectores soporte son básicamente un algoritmo de clasificación de patrones binario, cuyo objetivo es asignar cada patrón a una clase [32]. Por ejemplo, si tenemos dos conjuntos de elementos, uno de ellos compuesto por ovejas blancas y otro por ovejas negras, el algoritmo tratará de diferenciar estas ovejas en función de su color (clase), clasificando cada una de las ovejas en el conjunto blanco o negro.

Comenzaremos la explicación de las máquinas de vectores soporte [Burges, 1998] haciendo uso del caso más simple, el caso donde los datos son linealmente separables. Más adelante extrapolaremos la solución a problemas donde los datos no cumplan esta característica.

Los datos con los que entrenaremos el sistema serán una serie de vectores etiquetados, de la forma:  $\{\vec{x}_i, \vec{y}_i\}$   $i = 1, \dots, l$ .

Donde  $\vec{x}_i \in R^d$  es el vector de observaciones en un espacio de dimensión  $d$ . Y  $\vec{y}_i \in \{-1, 1\}$  representa etiqueta de la clase a la que pertenece cada vector.

El problema consistirá en asignar cada vector a su clase correspondiente, 1 ó -1, para ello se construirá un hiperplano de separación que divida el espacio  $d R$  en dos regiones. Supongamos que tenemos dicho hiperplano, las muestras que caigan en una región pertenecerán a clase -1 (negativa) y las que caigan en la otra a la clase 1 (positiva). A este hiperplano se le conoce como hiperplano de separación.

Los puntos  $\vec{x}$  que caen justo en este hiperplano satisfarán la ecuación:  $\vec{w}\vec{x} + b = 0$ . Donde  $\vec{w}$  es un vector normal al hiperplano de separación y  $b$  es una constante.

A la hora de buscar este hiperplano de separación óptimo las distancias cobrarán una especial relevancia por ello se utiliza la distancia perpendicular desde el hiperplano al origen.

Veamos un ejemplo gráfico. Por comodidad y facilidad de interpretación se ilustrará en un espacio de 2 dimensiones,  $R^2$ . Los círculos negros representan muestras pertenecientes a la clase 1 y los cuadrados muestras de la clase -1. El objetivo del algoritmo será encontrar el hiperplano que separe estas muestras de una manera óptima. En la Figura 3.15 a) se ilustra la distribución espacial de las muestras y varios hiperplanos de separación posibles.

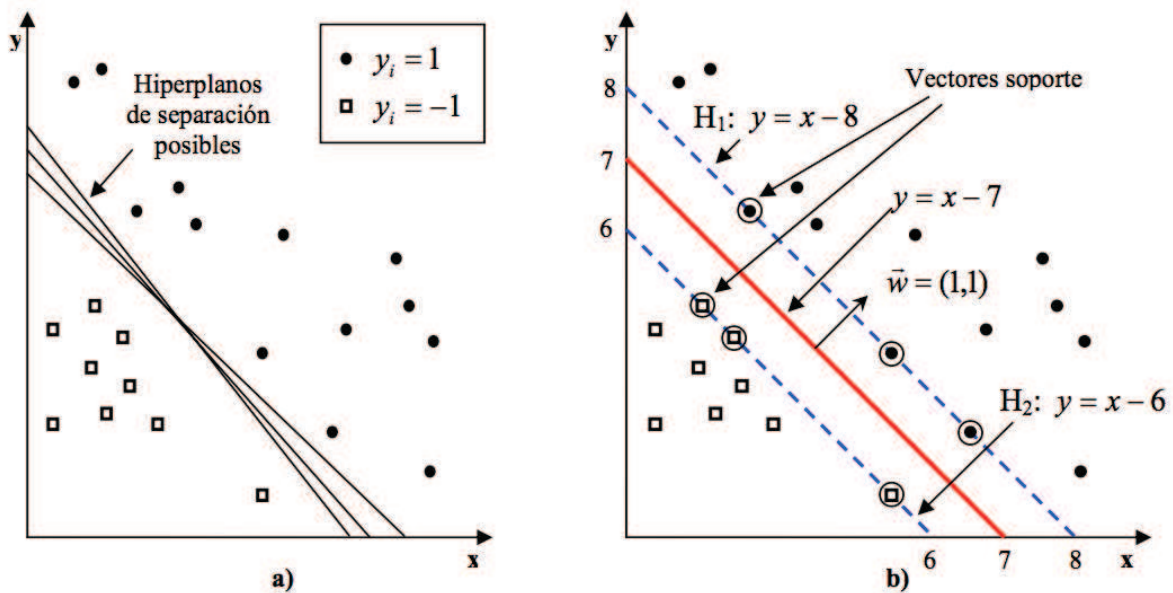


Figura 3.15: Representación de muestras pertenecientes a dos clases distintas. a) observaciones y posibles hiperplanos de separación. b) hiperplano de separación óptimo e hiperplanos  $H_1$  y  $H_2$

En el caso este proyecto la dimensión de nuestras características/features es de  $d=1$ , teniendo por tanto un espacio de 1 dimensión,  $R^d$ . Por esta razón se descarta el uso de SVMs lineales y se propone el uso de SVMs no lineales con un kernel de función de base radial Gaussiana, separado por un hiperplano en el espacio transformado.

Es decir, el kernel utilizado será del tipo:

$$K(x_i, x_j) = \exp(-(x_i - x_j)^2 / 2(\sigma)^2)$$

donde  $\sigma = 1$

De esta forma, al utilizar este tipo de núcleo (kernel) sobre características de una única dimensión, como puede ser el caso de utilizar la altura del sujeto para entrenar un clasificador de tipo de altura (bajo, medio, alto), permitimos que puedan alcanzar una convergencia.

### 3.4.2. Entrenamiento

El entrenamiento de los clasificadores SVMs para los atributos estudiados ha seguido la misma metodología.

Primero, se utiliza el conjunto de imágenes de entrenamiento para obtener los umbrales superiores e inferiores de cada atributo estudiado. Con estos umbrales se generan tres subconjuntos de datos sobre los que se entrenan dichos SVMs.

Por ejemplo, para la altura, se han creado los 3 subconjuntos de a) bajos, b) medios y c) altos. Dichos subconjuntos se han utilizado para entrenar los 3 SVMs buscados, cada uno de ellos para cada tipo de subconjunto.

Para llevarlo a cabo, se han utilizado clases positivas y negativas de la siguiente forma:

- Clasificador 1. Clase positiva: Bajos. Clase negativa: Normales y Altos.
- Clasificador 2. Clase positiva: Normales. Clase negativa: Bajos y Altos.
- Clasificador 3. Clase positiva: Altos. Clase negativa: Bajos y Normales.

Los clasificadores para el resto de atributos se han desarrollado de la misma manera.

### 3.4.3. Test

Para evaluar nuestro sistema simplemente se ha calculado la distancia de los atributos de la imagen de test al hiperplano de cada uno de los SVMs entrenados.

Por ejemplo, para la altura de una imagen de test, se obtienen 3 scores:

- $s_1$  = distancia del SVM altura baja
- $s_2$  = distancia del SVM altura media
- $s_3$  = distancia del SVM altura alta

El  $\max(s_1, s_2, s_3)$  será por tanto el atributo asignado a dicho sujeto de test.

En el próximo capítulo se ilustra en detalle el protocolo experimental y los resultados obtenidos con el estimador, donde se presenta de forma gráfica todo lo explicado anteriormente.

## 3.5. Procesado Manual Inicial

---

De forma paralela a todos los pasos descritos en este capítulo y de cara a obtener un ground-truth de todos los atributos corporales con el que poder comparar los resultados que se obtengan con este estimador, ha sido necesario procesar las imágenes de todos los sujetos de forma manual.

Para ello se han seguido los mismos pasos descritos hasta ahora, aunque de distinta forma:

1. **Etiquetado manual** de todos los puntos característicos en la imagen cuyas coordenadas van a permitir obtener una medida de una feature. Se corresponden al completo con los puntos que se encuentran marcados en la figura 3.4, incluidos los puntos de los codos. Se trata de atributos **objetivos**.

Todas estas coordenadas se han almacenado para cada imagen.

2. **Normalización** de todos estos puntos siguiendo exactamente el mismo proceso que el descrito previamente en este capítulo, sección 3.2.
3. **Extracción de las features.**

Para aquellas features que dependen de las coordenadas de los puntos etiquetados en la fase previa, se ha calculado la distancia euclídea entre ellos (tal y como se ha procedido con los puntos extraídos automáticamente, Sección 3.3). La única diferencia con respecto al procesado automático, es que en el manual el largo de los brazos se calcula como la suma del hombro al codo más la distancia del codo a la mano, con lo cual tenemos una medida más precisa del largo del brazo.

Para las demás, atributos **subjetivos** (envergadura y color del pelo), no se ha extraído ninguna feature como sí se ha hecho en la Sección 3.3 para la extracción automática, sino que se ha realizado la clasificación del sujeto directamente en una categoría (0-1-2) según el juicio de la autora. Adicionalmente a la envergadura y color del pelo, en la etapa manual se han clasificado otros atributos de esta naturaleza: han sido el largo del pelo y la etnia del usuario.

La feature de Body Shape utilizada en la extracción automática no se ha aplicado en el procesado manual, dado que para poder realizar una clasificación manual de los usuarios ha bastado con el etiquetado de puntos para los atributos objetivos y con una clasificación directa para los subjetivos. Por tanto, esta feature únicamente se aplicará en la estimación automática de la forma especificada en el protocolo experimental del Capítulo 4.

4. **Clasificación - Ground Truth.** Se han generado etiquetas de clasificación dividiendo en 3 posibles grupos cada una de las features.

Para los atributos subjetivos, tal y como se ha comentado en el punto anterior, la clasificación se ha realizado directamente a juicio de la autora.

Para los atributos objetivos, medidos con distancias entre puntos en la imagen, la clasificación se ha realizado a partir de los histogramas de sus valores repartidos en todos los usuarios de la BBDD, fijando dos umbrales para la separación de 3 grupos a la vista de los resultados. Estos umbrales han sido levemente ajustados mediante la observación directa de los usuarios próximos a los límites. Una vez definidos los umbrales, se ha clasificado a cada usuario según el grupo al que se corresponda su valor de medida de la feature en concreto.

En la figura 3.16 se muestra un ejemplo de la clasificación manual que se ha llevado a cabo con cada feature. En este caso, el ejemplo representa la forma en la que se ha clasificado el atributo de la altura.

En concreto, los atributos que se han extraído a partir de un etiquetado **manual** en las imágenes de la BBDD han sido:

- Altura. (*Objetiva*)
- Ancho de Hombros. (*Objetiva*)
- Ancho de Cadera. (*Objetiva*)

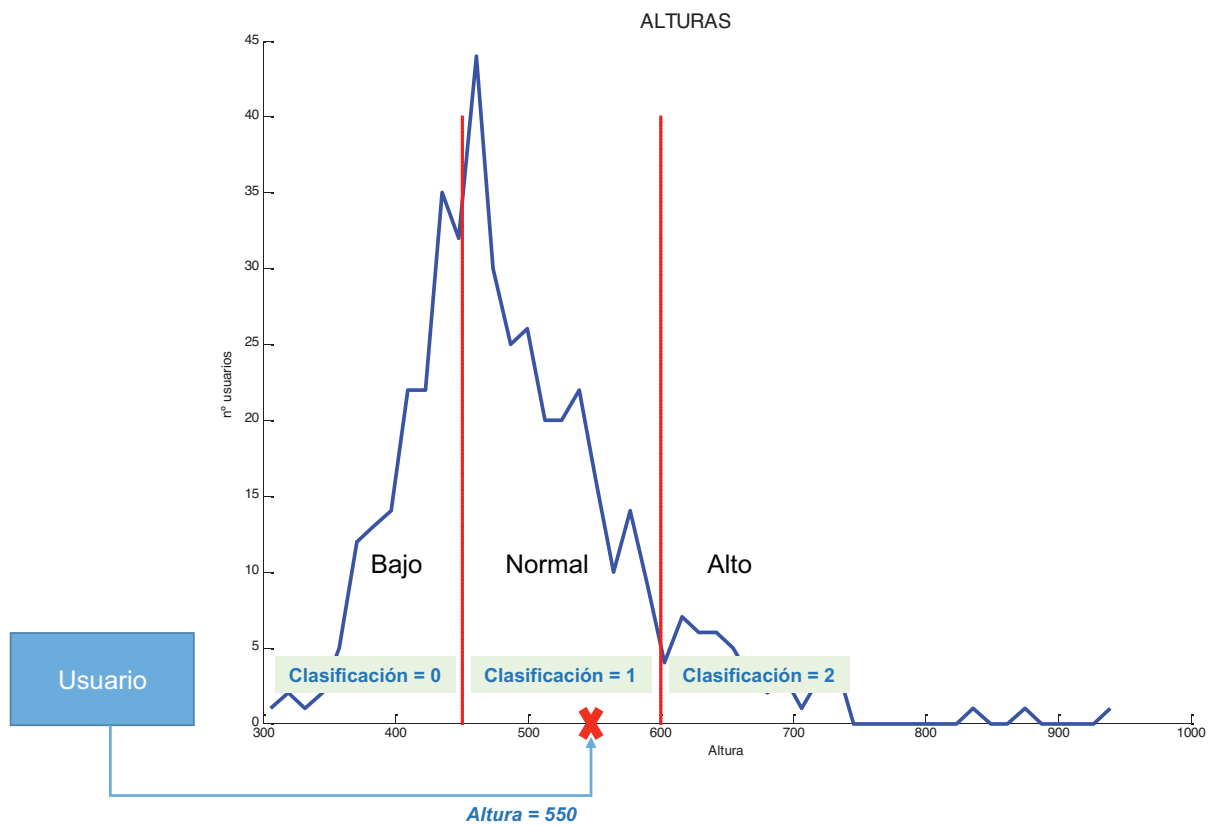


Figura 3.16: Clasificación Manual de features según la umbralización de sus histogramas. Ejemplo mostrado para el atributo de la altura.

- Distancia entre los Ojos. (*Objetiva*)
- Distancia Ojos - Nariz. (*Objetiva*)
- Largo Brazo derecho e izquierdo. (*Objetivas*)
- Envergadura. (*Subjetiva*)
- Color del pelo. (*Subjetiva*)
- Longitud del pelo. (*Subjetiva*)
- Etnia. (*Subjetiva*)

Es importante subrayar que se han encontrado varias **limitaciones o dificultades** durante el etiquetado/clasificación de los atributos. Fruto de ello, en posteriores etapas podrían encontrarse discrepancias entre los resultados de la clasificación. A continuación se detallan varios de estos problemas:

- Diferenciación de largo del pelo para las personas que llevan coleta. Aunque realmente tengan el pelo largo, en la imagen no se contempla por estar recogido.
- Estimación del color del pelo en gente con gorra. Dado que se encuentra oculto en gran parte, un sistema no sabrá estimar correctamente cuál es el color correcto.

- Estimación del color del pelo con una calidad pobre de la imagen. En ocasiones existen focos de luz directamente sobre la cabeza o la imagen se encuentra más oscura en dicha zona. Esto hace que se pueda confundir el color verdadero.
- Detección de color del pelo en personas calvas.
- Largo de los brazos cuando los individuos tienen las manos ocultas, por ejemplo, detrás de la espalda o en el bolsillo.
- Estimación de la envergadura en personas que llevan ropas muy anchas. Este hecho confundirá a cualquier sistema, detectando una mayor envergadura de la real.
- Diferenciación entre personas musculadas y gruesas. Dos individuos pueden ser igual de anchos, pero uno de ellos ser delgado o normal porque está musculado, y otro de ellos ser realmente grueso.

Al margen de dichos problemas, en conclusión, tal y como se describirá en el siguiente capítulo, gracias a este procesado manual inicial que deriva en un ground truth será posible realizar una correcta evaluación de los resultados del estimador implementado.

En definitiva, el **diagrama general** llevado a cabo en este proyecto se muestra en la figura 3.17. En dicha figura se incluye el procesado automático y el procesado manual, gracias al cual se obtiene un ground-truth sobre el que poder realizar una evaluación de los resultados dados por el estimador.

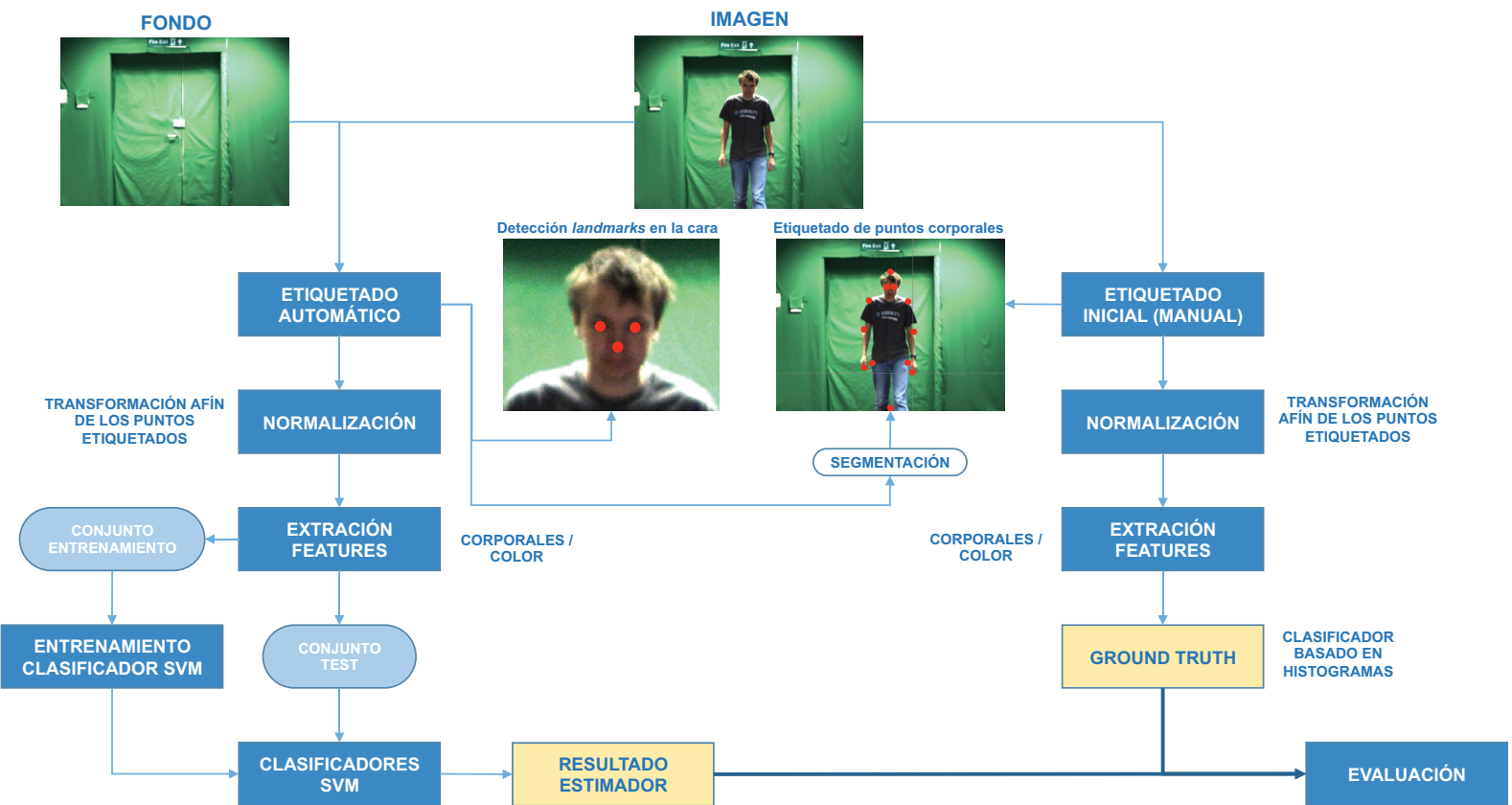


Figura 3.17: Diagrama general de la implementación del estimador de atributos corporales: procesado automático y procesado manual.





## 4 Experimentos realizados y Resultados

### 4.1. Base de Datos

La base de datos utilizada en este proyecto ha sido *Southampton Multibiometric Tunnel Database (TunnelDB)* [27]. Esta base de datos contiene ejemplos biométricos de 222 sujetos, para los cuales se han tomado 10 vídeos de ejemplos de forma de caminar utilizando desde 8 a 12 puntos de vista simultáneos. Estas muestras se han almacenado para extraer información acerca de la forma de andar en 3D.

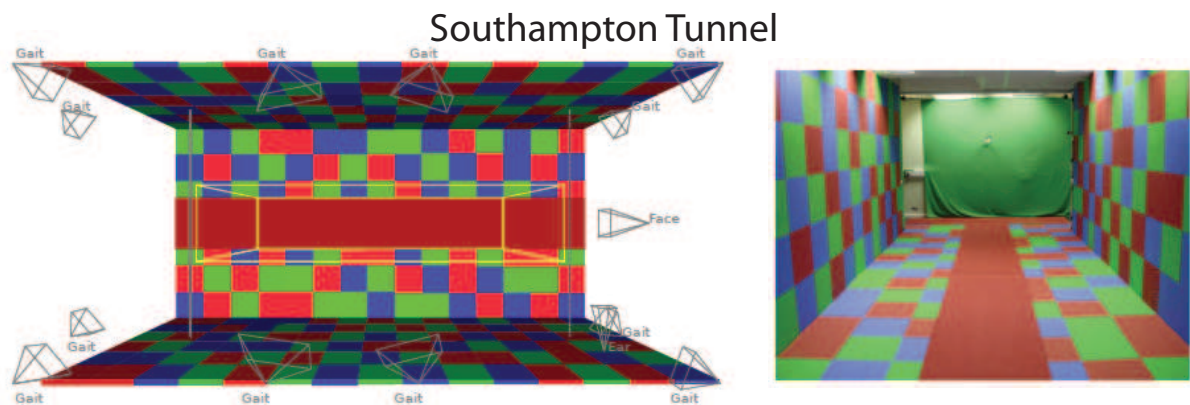


Figura 4.1: Tunnel Database setup. Existen 8 cámaras de adquisición de la forma de caminar y una única cámara de alta resolución adquiriendo el frontal de los sujetos. Fuente: [5]

TunnelDB también contiene imágenes frontales de alta resolución para extraer información sobre la cara, e imágenes de perfil para extraer rasgos biométricos de las orejas.

En total hay aproximadamente 10 conjuntos de información reunidos para cada sujeto en TunnelDB.

De los 222 sujetos, el 67% son hombres; la mayoría con una edad de entre 18-28 años y el 70% de origen europeo. Estas tendencias demográficas en la base de datos son las esperadas, ya que representan estudiantes.

El proceso de adquisición se muestra en las figuras 4.1 y 4.2, donde los sujetos fueron recogidos caminando desde la entrada de un pasillo recto con el suelo rojo hasta la salida, de frente a una cámara facial. Durante una única andada (una muestra), el sujeto fue capturado simultáneamente por las cámaras faciales y cámaras que obtienen la forma de andar. Cuando el sujeto alcanza el

final del pasillo, se utilizó un único disparo de la cámara para fotografiar la oreja derecha.

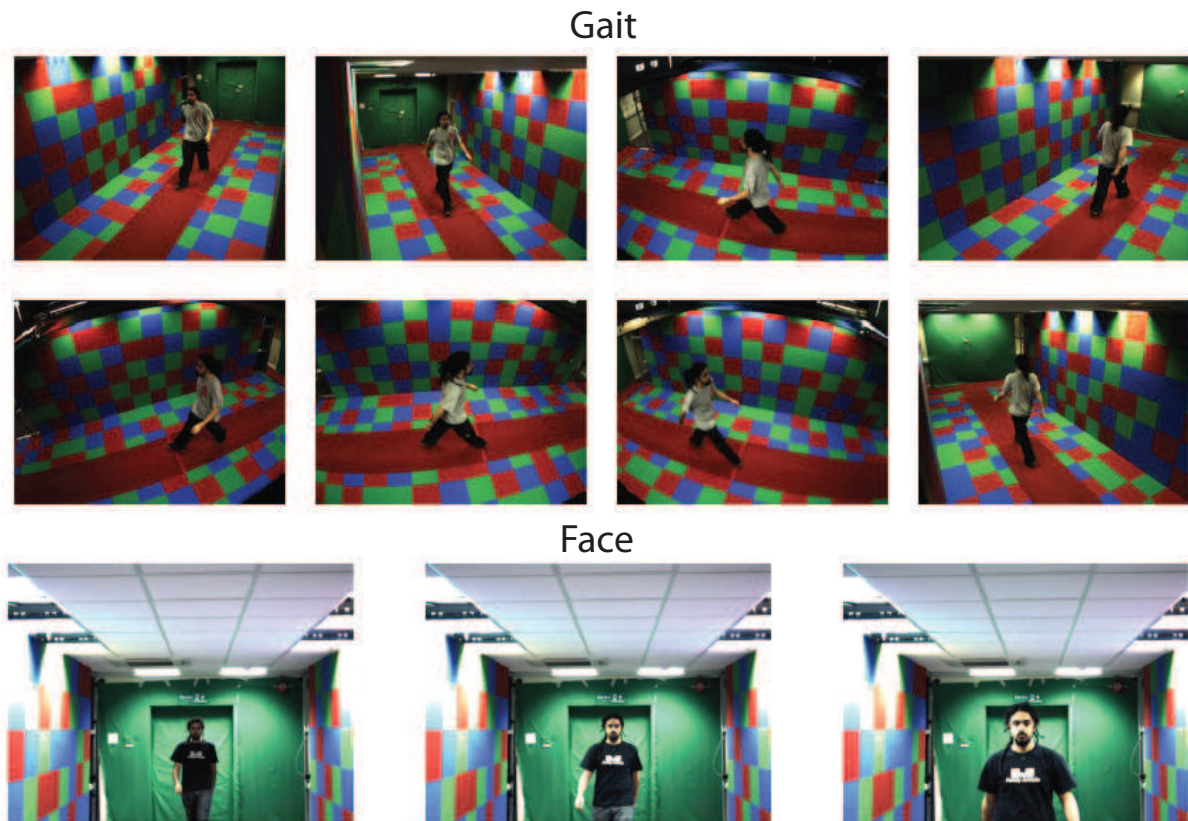


Figura 4.2: Ejemplos de TunnelDB. Fuente: [5]

La información de la andada está grabada por 12 cámaras, y a la vez que se tomaron estas imágenes de la forma de caminar, una única cámara de una mayor resolución capturaba al final del pasillo imágenes faciales con una resolución de 1600x1200, a 27 frames por segundo.

En el escenario del pasillo, la dirección de la mirada fue garantizada instruyendo a los sujetos, así como se hizo para la dirección en la que debían caminar. La iluminación y otras variables del entorno también fueron controladas.

Las figuras 4.1 y 4.2 muestran la forma de adquisición, junto con un ejemplo de cada rasgo biométrico capturado por sujeto.

## 4.2. Protocolo Experimental

El protocolo experimental adoptado en este proyecto se basa en aplicar una validación cruzada de  $K$  iteraciones (del inglés  $K$ -fold cross-validation). Esta técnica se basa en dividir los datos disponibles en  $K$  subconjuntos.

Uno de los subconjuntos se utiliza como **datos de test** y el resto ( $K-1$ ) como **datos de entrenamiento**.

En nuestro caso se ha seleccionado un  $K = 5$ , de forma que la base de datos se divide evaluando 5 diferentes conjuntos de test y de esta manera se logra completar totalmente el análisis de la misma.

De los 222 sujetos disponibles en la BBDD, en este proyecto se han utilizado 2 imágenes por cada usuario, cada una correspondiente a una sesión de vídeo distinta y coincidiendo con el

punto de vista más lejano. Por ello, el total de imágenes con el que se ha trabajado ha sido de 444.

El proceso se describe en al Fig. 4.3, donde como puede verse, de las 444 imágenes disponibles se han seleccionado aleatoriamente sub-conjuntos de 89 imágenes de test y 355 de entrenamiento.

**Nota:** A pesar de que en la figura 4.3 parezca que se han seleccionado las imágenes de test de forma secuencial y ordenada, en la práctica se ha realizado una selección aleatoria.

Para la primera iteración  $K = 1$ , se selecciona de forma aleatoria el conjunto de datos de test y se utiliza el resto de imágenes como datos de entrenamiento.

Para las siguientes iteraciones,  $K = 2$  a  $K = 5$ , el conjunto de datos de test se selecciona aleatoriamente de entre el conjunto de datos que aún no haya sido seleccionado como Datos de Test en las iteraciones anteriores, es decir, que únicamente hayan sido utilizados como Datos de Entrenamiento hasta el momento.

De esta forma se logra que todas las imágenes de la base de datos sean evaluadas.

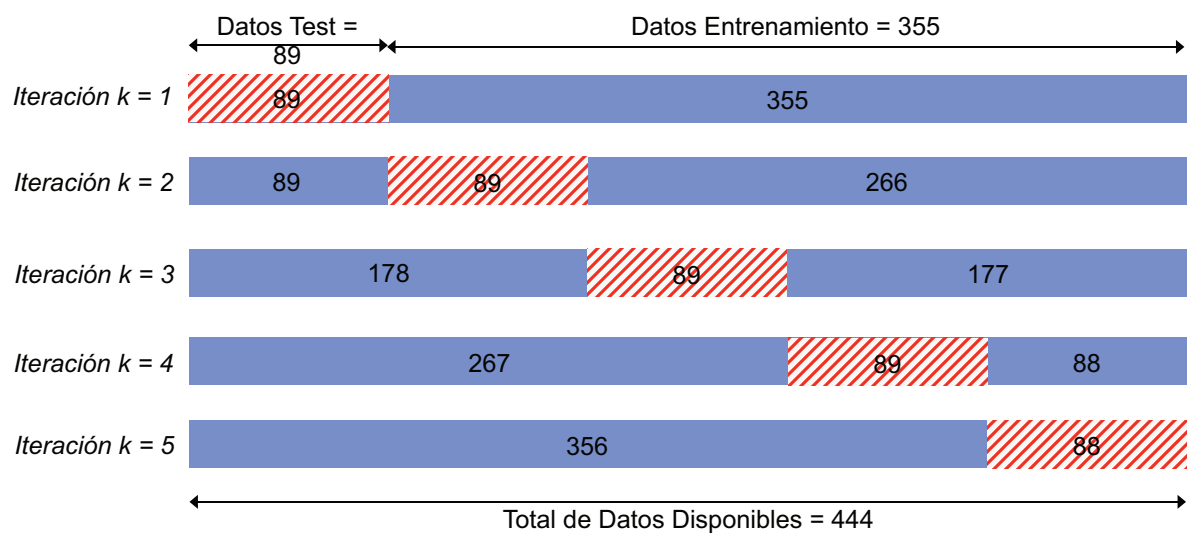


Figura 4.3: Protocolo experimental 5-fold cross validation. Ejemplo de la distribución de conjuntos en la selección de los conjuntos de entrenamiento y test en cada caso.

Adicionalmente, es importante destacar que **la evaluación de resultados se va a dividir en dos partes** claramente diferenciadas, en función de las features que se utilicen para la clasificación de atributos corporales:

- Features de **medidas, color del pelo y envergadura**.

Primeramente se entrena y testea al sistema utilizando como features la distancia entre puntos para los atributos objetivos, el valor RGB para el pelo y el porcentaje frente-tamaño imagen para la envergadura.

- Feature de la **Forma del Cuerpo (*Body Shape*)**.

En una segunda parte se realizará el entrenamiento y testeo del sistema con la feature de *Body Shape* para todos los atributos objetivos y la envergadura (el color de pelo no tiene relación con la forma del cuerpo, y por tanto no se ha evaluado con esta feature).

De esta forma se podrá evaluar **qué tipo de features funcionan mejor para cada atributo corporal**, diferenciando entre las propuestas en este proyecto y la propuesta en un trabajo relacionado respecto a la forma del cuerpo.

## ENTRENAMIENTO

El protocolo que se ha seguido para **realizar el entrenamiento** del sistema (figura 4.4) ha sido ligeramente distinto dependiendo de si los atributos son objetivos o subjetivos, o de si se trata de la primera fase de evaluación o la segunda. Sin embargo, en todos los casos se han utilizado como datos de entrenamiento los atributos resultantes del **etiquetado automático** y al mismo tiempo se ha utilizado el groundtruth manual para poder separar las imágenes en las 3 clases de cada atributo.

A continuación se describe en detalle el entrenamiento realizado en el sistema:

### 1. FASE I. Features de medidas, color del pelo y envergadura, desarrolladas en este proyecto.

- Para cada uno de los **atributos objetivos del estimador** (medibles como distancia entre puntos) se ha seguido el siguiente protocolo:

- a) Se utilizan las 355 imágenes de entrenamiento para extraer los atributos estimados a partir de los puntos etiquetados manualmente para cada una de ellas.
- b) Se obtiene la distribución que sigue cada atributo para el conjunto de imágenes de entrenamiento.
- c) Esta distribución se divide en tres subconjuntos, creando los **3 grupos de imágenes**.

Para ello ha sido necesario obtener dos umbrales que delimiten estos 3 grupos. Al igual que se hizo en la clasificación de las imágenes con etiquetado manual (Sección 3.5), la clasificación del conjunto de entrenamiento se ha realizado a partir de los histogramas de sus valores repartidos en todos los usuarios de la selección de entrenamiento, fijando dos umbrales para la separación de 3 grupos de forma equitativa.

Por ejemplo, para el atributo de la altura se crean 3 grupos denominados: bajos, medios y altos.

En la figura 3.16 de clasificación de la sección 3.5, se puede ver un ejemplo de dicha selección de subconjuntos para el entrenamiento de los SVMs.

Dichos grupos serán los utilizados para entrenar los SVMs, es decir, se entrenará **un SVM por cada grupo** de cada atributo:

- SVM 1: Primer grupo contra todos los que no son del primer grupo.
- SVM 2: Segundo grupo contra todos los que no son del segundo grupo.
- SVM 3: Tercer grupo contra todos los que no son del tercer grupo.

Para el entrenamiento, a los SVMs de los atributos objetivos se les introducirá directamente el **valor de la medida** de la feature en concreto, junto con el conocimiento de a qué grupo de los tres posibles pertenece (ground-truth a partir del etiquetado manual).

- Por otro lado, el protocolo que se ha seguido para **realizar el entrenamiento de atributos subjetivos** (color del pelo y envergadura) ha sido diferente, dado que no se basan en medidas tomadas como distancia entre puntos. Sin embargo, sí comparten la existencia de los 3 SVMs, cada uno para un grupo de cada atributo. De forma más concreta:

- **Color del pelo.** Se ha utilizado la información de color RGB obtenida en la extracción de features (Sección 3.3). Esta medida no tiene error de marcado, ya que no depende de etiquetado de puntos.

Este atributo, el cual se compone de 3 features diferentes (R, G y B), será el que se introducirá en el SVM. Es decir, será una **feature de 3 Dimensiones**.

Dado que es necesario introducir también a qué clase se corresponde dicho valor RGB, se utilizará la clasificación manual realizada sobre el sujeto al que pertenece dicha feature. De esta forma, se consigue obtener la **relación valor-clasificación** que se introduce en el SVM: Rubio, Castaño o Moreno.

- **Envergadura.** Se ha utilizado la información de proporción entre la figura de la persona y el tamaño de la imagen, dato obtenido en la Sección 3.3 (medida sin error de marcado), además del ancho de cadera, ancho de hombros y Altura medidos a partir de los puntos etiquetados automáticamente (éstos sí contendrán error de marcado). Se ha seleccionado esta combinación de features porque se ha considerado que todas ellas influyen a la hora de determinar la envergadura de una persona.

El conjunto de estas 4 features distintas será lo que se introducirá en el SVM. Es decir, se trata de una **feature de 4 Dimensiones**.

Al igual que para el resto de atributos, se utilizará la clasificación manual realizada sobre el sujeto al que pertenecen dichas features. De esta forma, se consigue obtener la **relación valores-clasificación** que se introduce en el SVM: Delgado, Normal o Grueso.

## 2. **FASE II.** Feature de la Forma del Cuerpo (*Body Shape*), basada en el Estado del Arte.

Esta feature es la única que se considera en esta fase.

En esta segunda parte que se va a tratar para la evaluación del sistema, se entrenan todos los atributos que puedan tener relación con la forma del cuerpo. Por tanto, se seleccionan todos los atributos objetivos y la envergadura, ya que el color del pelo no guarda relación con esta feature.

El procesado es exactamente igual para todos ellos. Se utiliza, para cada imagen, el vector resultante de la concatenación de la suma de píxeles de frente en horizontal y vertical. Esta feature multidimensional (hasta 1050 dimensiones), junto con la clasificación manual realizada sobre el sujeto, se introducen en los 3 SVMs posibles, cada uno para un grupo de cada atributo.

Una vez que se haya introducido en cada caso la feature específica en el SVM, el siguiente paso en la fase de entrenamiento es común a todos los atributos y a ambas fases. Este paso tiene como objetivo clasificar a los sujetos de las imágenes según los diferentes grupos de sus atributos, mediante la **búsqueda del hiperplano** equidistante entre las diferentes clases.

Es decir, se busca, para cada uno de los tres SVMs utilizados, cuál es el hiperplano que separa a cada clase del resto.

Para llevar a cabo la clasificación de las imágenes, es importante destacar que todas las **features objetivas de medidas entre puntos son unidimensionales**, ya que cada atributo se describe únicamente con una medida. Es decir, determinar la altura de una persona o el ancho de sus caderas solamente depende de la medida de la distancia que se haya calculado entre dos puntos etiquetados.

Sin embargo, **el resto de features son multidimensionales**. En el caso de los atributos subjetivos en la Fase I, tanto el color del pelo como la envergadura se van a clasificar a partir de varias medidas o datos independientes. En el caso de la Fase II, el vector de forma del cuerpo aplicado a todos los atributos tiene hasta 1050 dimensiones.

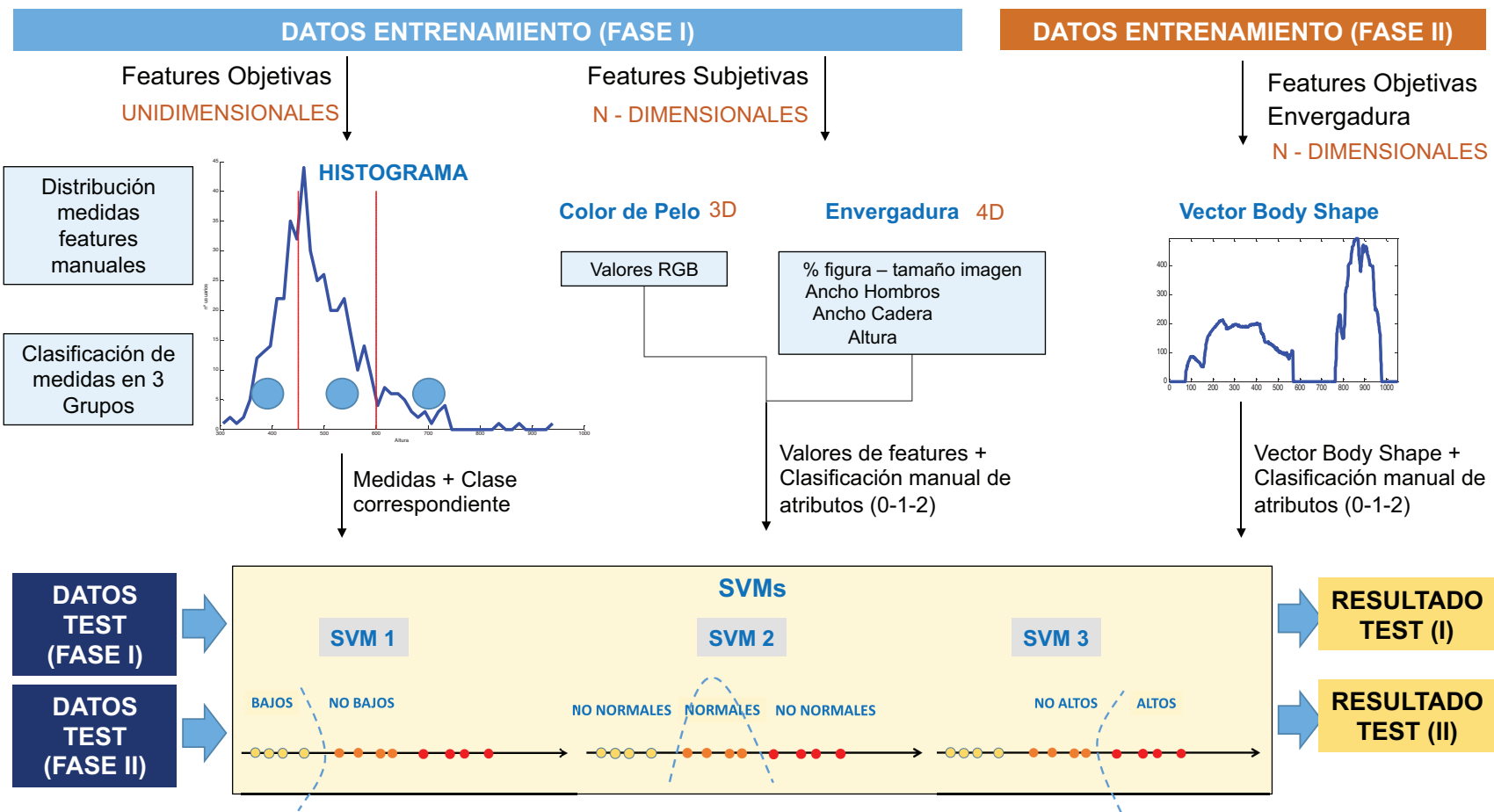


Figura 4.4: Protocolo experimental: Fase de entrenamiento y Test.

Centrándonos en el caso de la unidimensionalidad de nuestras features objetivas en la Fase I, nos enfrentamos a un problema con la **selección del kernel del SVM**:

La selección de un kernel lineal en el entrenamiento de los SVMs funcionaría solamente para el primer y tercer grupo de cada atributo (por ejemplo, para el atributo de altura, los grupos de bajo y alto), ya que para estos casos es posible converger en la búsqueda del hiperplano lineal (en este caso una recta) que separa a dicho grupo del resto.

Es decir, utilizando como ejemplo el atributo de altura, una recta es suficiente para separar los bajos de todo aquel que no es bajo (normales y altos), o para separar a los altos de todo aquel que no es alto (bajos y normales).

Sin embargo, el **kernel lineal en features unidimensionales no funcionaría** para el grupo intermedio, ya que no es posible realizar una separación lineal con los otros dos posibles grupos, es decir, una sola recta no es suficiente para realizar la separación. En este caso, la búsqueda de la recta que actúa como hiperplano no podría converger nunca, por **no ser linealmente separable**.

Este caso se encuentra representado en la figura 4.5, donde se muestra la dificultad que implicaría utilizar un kernel lineal en features unidimensionales.

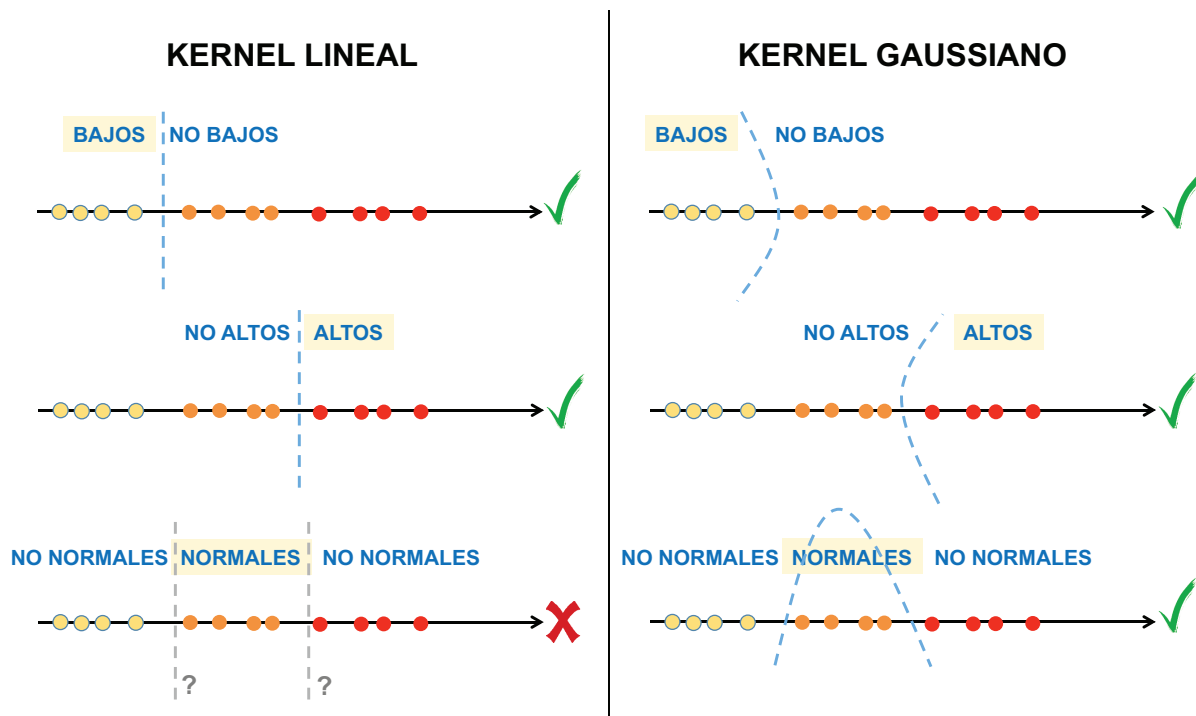


Figura 4.5: Features unidimensionales: Kernel lineal vs Kernel gaussiano. Ejemplo: Altura.

Por todo ello, los SVMs son entrenados con un **kernel gaussiano** [33] para poder funcionar correctamente en features de una sola dimensión.

Utilizar un tipo de función kernel (en este caso gaussiano) permite proyectar la información a un espacio de características de mayor dimensión, el cual aumenta la capacidad computacional de las máquinas de aprendizaje lineal. Es decir, se mapea el espacio de entradas  $X$  a un nuevo espacio de características de mayor dimensionalidad donde las clases puedan ser linealmente separables.

La función de base radial Gaussiana que separa por un hiperplano en el espacio transformado y que se puede observar en la figura 4.6 es la siguiente:

$$K(x_i, x_j) = \exp(-(x_i - x_j)^2 / 2\sigma^2)$$

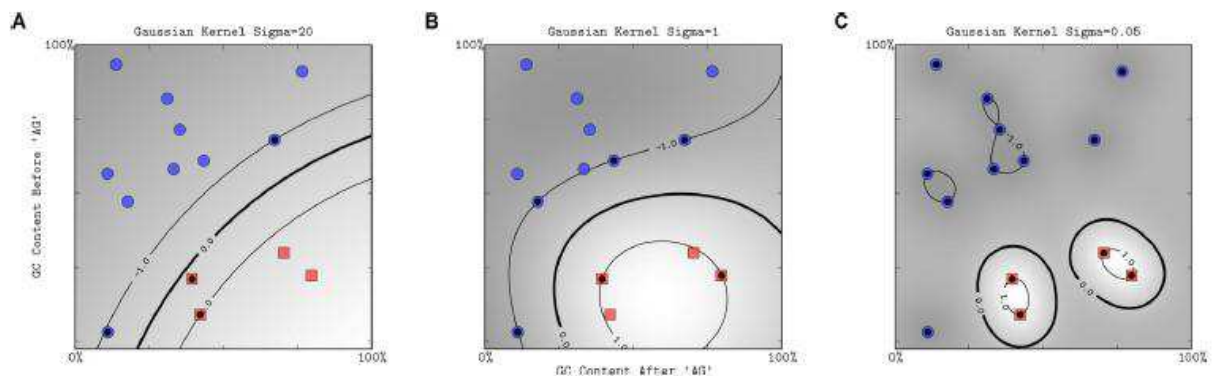


Figura 4.6: Función Kernel Gaussiano aplicado a datos no separables linealmente. Fuente: Wikipedia.

Al aplicar el kernel gaussiano a las features unidimensionales, las tres clases convergen al encontrar el hiperplano de separación. Se puede observar en la figura 4.5.

Aunque esta problemática no se aplica para el caso de las features multidimensionales (el resto), se ha seleccionado el mismo kernel gaussiano para unificar los SVMs utilizados.

## TEST

Por último, una vez que se han encontrado todos los hiperplanos de separación en todas las clases de los atributos, solamente queda introducir los **datos de test** para obtener su clasificación en base al entrenamiento que se ha llevado a cabo.

Como se ha comentado anteriormente, a los SVMs de los atributos objetivos en la fase I se les introducirá directamente el valor de la medida de la feature en concreto y su clasificación; a los de los atributos subjetivos en la fase I se les introducirán todas las features necesarias comentadas anteriormente (información de color, ancho de cadera, etc) y su clasificación; y en la fase II se introducirá el vector de información de forma del cuerpo y su clasificación.

De esta forma, una vez calculados los hiperplanos de separación entre clases, cuando se introducen los datos de test se calcula la distancia que tienen con respecto a los posibles hiperplanos, es decir, se devuelve un **score** de probabilidad de pertenecer a cada grupo. El score máximo indicará la clase a la que con más seguridad pertenezca el dato de test, consiguiendo por tanto clasificar cada uno de los atributos.

La globalidad del protocolo llevado a cabo en el proyecto, diferenciando la fase de entrenamiento y test para las dos fases de evaluación de resultados, se puede observar en la figura 4.4.

## 4.3. Evaluación de Resultados

Para evaluar los resultados obtenidos por el sistema desarrollado, a continuación se van a analizar las partes del sistema en las que se introduce error y se va a mostrar el detalle de la evaluación realizada sobre el sistema global implementado.



### 4.3.1. Evaluación de etiquetado

Una de las partes que introducen error en el sistema es la que se corresponde con el etiquetado de puntos para la extracción de features/landmarks en las imágenes. Un error en este punto afecta por encima de todo a los atributos objetivos en la fase I, ya que son aquellos cuya medida se calcula a partir de la distancia entre puntos etiquetados. Por tanto, un error muy alto en esta parte conllevará a un error notable en las medidas de las features con las que se entrena el sistema.

En esta sección se va a analizar el error que se comete en el marcado automático de puntos con respecto al etiquetado manual de los mismos.

Es decir, se evalúa la diferencia que existe entre las **coordenadas de los puntos** que se han marcado a mano y los puntos que se han marcado de forma automática siguiendo el procesado descrito en la Sección 3.1.2.

Los puntos etiquetados de forma manual se consideran el *ground-truth* del sistema.

Para extraer el error de marcado de puntos:

1. Para cada punto etiquetado de forma automática se ha calculado la distancia en píxeles con respecto al mismo punto etiquetado de forma manual.
2. Una vez que se ha realizado dicho cálculo para toda la BBDD y todos los puntos, se calcula la distancia media de error. De esta forma, obtenemos una media del error del etiquetado automático de puntos.

Los resultados obtenidos se pueden observar en la figura 4.7, donde se aprecia el error medio en el etiquetado automático de puntos. Para saber a qué punto hace referencia, los puntos se indican en la figura humana también contenida en la misma figura.

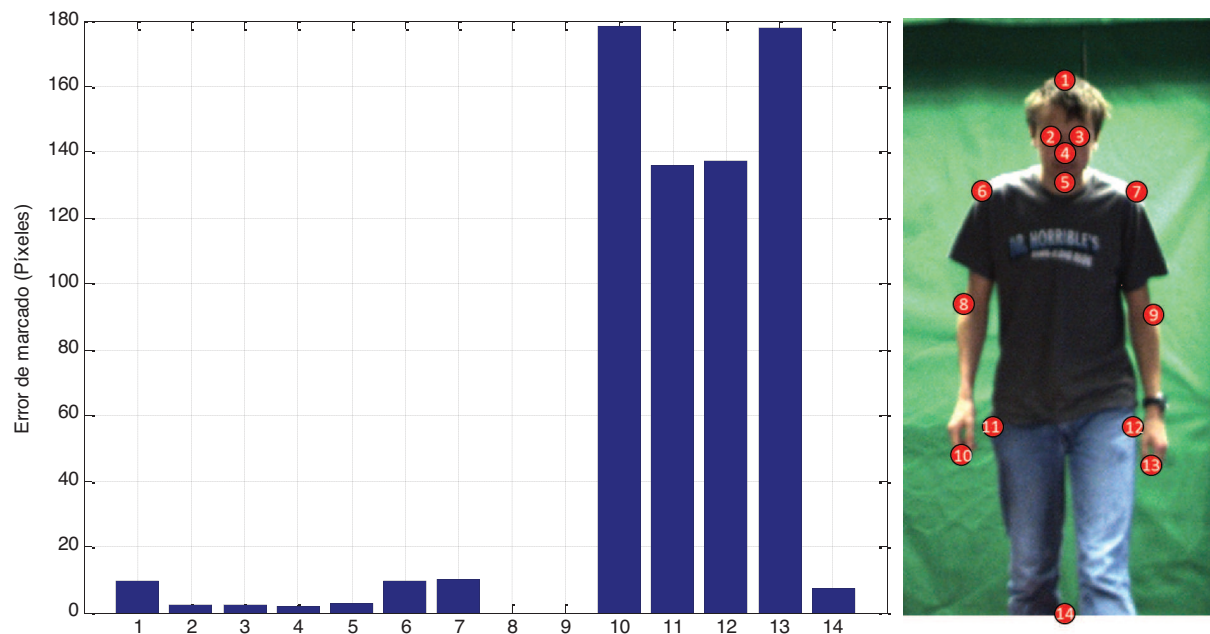


Figura 4.7: Error del Etiquetado Automático de los puntos vs Etiquetado Manual. Distancia media en píxeles.

Como se puede ver, los puntos de la altura, ojos, nariz, barbilla y hombros se etiquetan con un **error muy bajo**. Es decir, se ha conseguido localizar automáticamente estos puntos de una forma bastante precisa.

Los puntos referentes a los codos se han ignorado puesto que no se ha desarrollado ningún marcado automático para ellos.

Por último, el **mayor error de marcado se produce en las manos y las caderas**. En cuanto a las manos, esto se puede deber a múltiples motivos que se dan en la BBDD, como por ejemplo brazos que están doblados, ocultos detrás de la espalda, manos metidas en los bolsillos, etc. En cuanto a las caderas, existen dos problemas principales: Uno de ellos se debe a que en numerosas ocasiones es complicado separar los brazos de las caderas, no se localiza correctamente el borde de separación entre ellos, y por tanto el punto se etiqueta de forma incorrecta; y el otro de ellos se debe a la aparición de chaquetas o prendas de distinto color que en el procesado se consideran bordes de la figura. En el segundo caso, los puntos de la cadera se etiquetan prácticamente en el centro del cuerpo, donde se localiza este cambio brusco de color.

Tanto el punto de las manos como de la cadera tendrían menos error en una BBDD donde todos los individuos tuvieran los brazos estirados y visibles, y separados una mínima distancia del cuerpo.

Los errores de marcado más frecuentes que se han dado en esta BBDD se muestran en la figura 4.8, donde se indican imágenes de ejemplo en las que la estimación automática de los puntos no es correcta. En su mayoría suelen corresponderse con los puntos de las manos y las caderas.

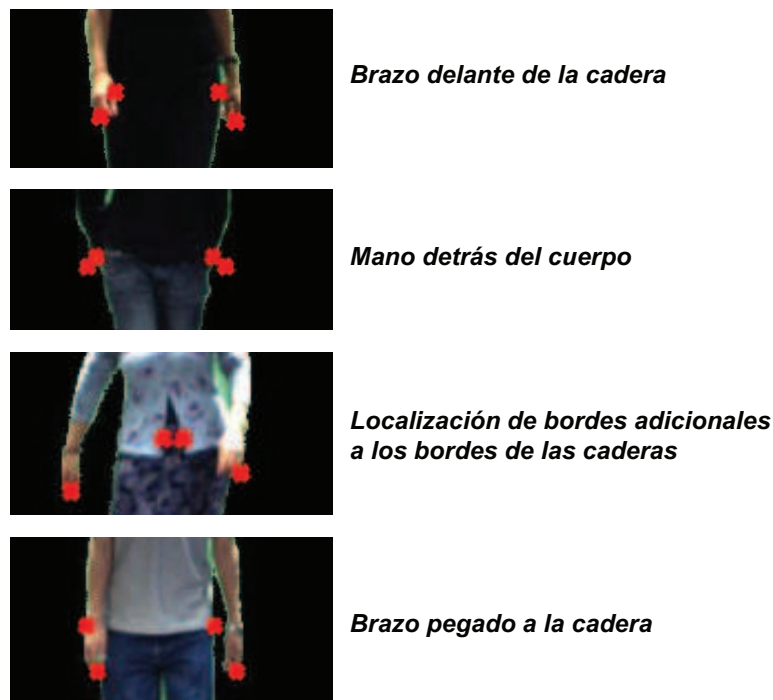


Figura 4.8: Ejemplos más frecuentes de errores en el marcado automático de puntos sobre la imagen.

Por otro lado, el etiquetado de puntos influye de forma directa en la **medida de las features** (Altura, Ancho de cadera, Ancho de hombros y Largo de brazos).

Por esta razón, también se ha evaluado el error que se obtiene en las medidas que se basan en las coordenadas de puntos etiquetados por el sistema de forma automática. Estos errores se pueden visualizar en la figura 4.9.

Para evaluar el error en las medidas:

1. Para cada distancia entre puntos etiquetados de forma automática (medidas automáticas) se ha calculado la diferencia en píxeles con respecto a la misma distancia entre puntos etiquetados de forma manual (medidas manuales, ground-truth).
2. Una vez que se ha realizado dicho cálculo para toda la BBDD y todas las medidas, se calcula una media del error del cálculo de medidas de features automáticas.

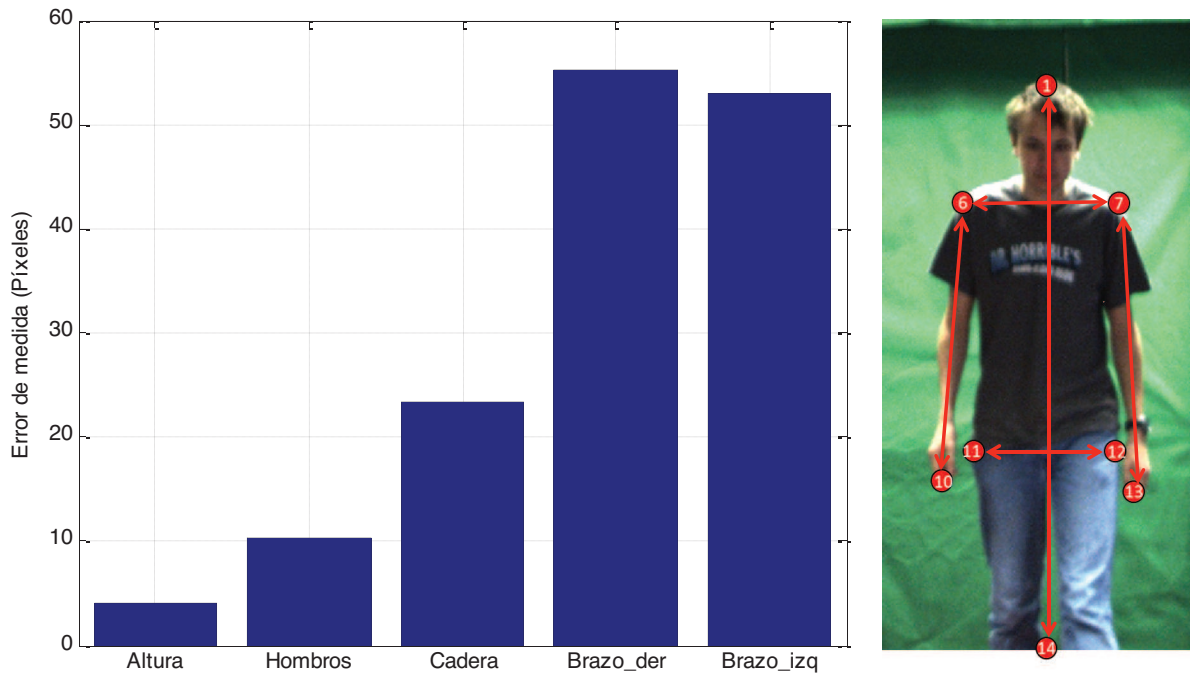


Figura 4.9: Error en la Medida de las features Objetivas: Medidas resultantes del etiquetado Automático vs Medidas resultantes del etiquetado Manual. Diferencia media en píxeles.

Basándonos en los resultados que se habían obtenido para el error del marcado de puntos, tiene sentido el error resultante en las medidas de features.

La altura y el ancho de hombros son bastante precisos, con un error muy bajo. El ancho de cadera introduce más error por los motivos explicados anteriormente en el error de marcado de puntos de la cadera. Por último, el largo de los brazos es la feature con más error, ya que además de todas las dificultades en el marcado de puntos de la mano ya descritos, no se han etiquetado los puntos de los codos. Dado que en las medidas manuales el largo del brazo se calculaba utilizando también el punto del codo, el error aumenta en la extracción de esta medida.

### 4.3.2. Evaluación del Sistema Real

En esta sección se realiza la evaluación global del sistema, teniendo en cuenta los **errores de etiquetado automático** descritos en el apartado anterior.

De esta manera vamos a obtener la precisión real del sistema desarrollado.

Para llevarlo a cabo, siguiendo el diagrama de entrenamiento y test descrito en la figura 4.4, vamos a entrenar al sistema con datos resultantes del etiquetado automático (como se ha detallado en la Sección 4.2) para obtener unos clasificadores lo mejor entrenados posibles para

los datos de test que van a tener que procesar. Además, se les entrenará con la clasificación del ground-truth correspondiente a cada uno de los usuarios según el atributo en concreto.

Los **Datos de Test** serán también aquellos resultantes de la extracción automática de features, que se introducirán en los SVMs previamente entrenados y obtendrán un score según la cercanía que tenga con respecto a cada grupo de atributos.

En este caso en el que se va a evaluar el Sistema Real, el método de entrenamiento y Test llevado a cabo ha sido el que se muestra en la figura 4.10.

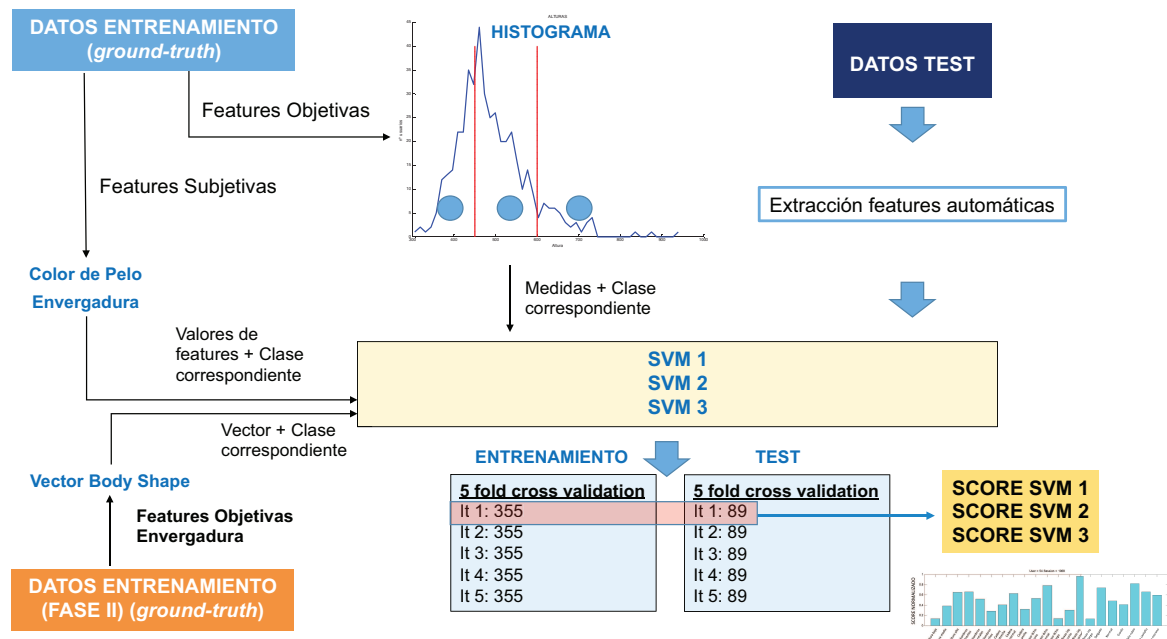


Figura 4.10: Entrenamiento y Test para la evaluación del Sistema Real.

Como se puede ver en dicha figura, los resultados del sistema consistirán en **3 scores, uno por cada SVM**. Es decir, para cada una de las 3 posibles clases se devolverá la probabilidad de que un atributo en concreto pertenezca a cada uno de los grupos posibles. Como es lógico, el máximo score calculado por el SVM para cada atributo determinará su clasificación.

Para la fase I, se obtendrán los scores correspondientes a las features con las que se ha evaluado el sistema, y para la fase II, se obtendrán otros scores correspondientes únicamente a la feature de forma del cuerpo.

Un ejemplo del resultado dado por el sistema con estos scores se puede observar en la figura 4.11, en la que se puede ver a una mujer que a la vista de los resultados se clasificaría como Alta, con Hombros estrechos, Cadera normal, Largo de los brazos normales, Delgada y Rubia.

Una vez que hemos obtenido todos los scores (y por tanto la clasificación) para todos los atributos de los usuarios, se compara con la clasificación realizada para los mismos usuarios en el ground-truth.

El porcentaje de aciertos se ha calculado de la siguiente manera:

$$\%Acierto = 100 - 100 * \frac{NumMalClasificadas}{NumTotal}$$

donde NumTotal = 444 imágenes.

En la tabla 4.1 se muestran los errores cometidos por el Sistema Real con las features de la Fase I, comparándolos con la clasificación realizada en el ground-truth.

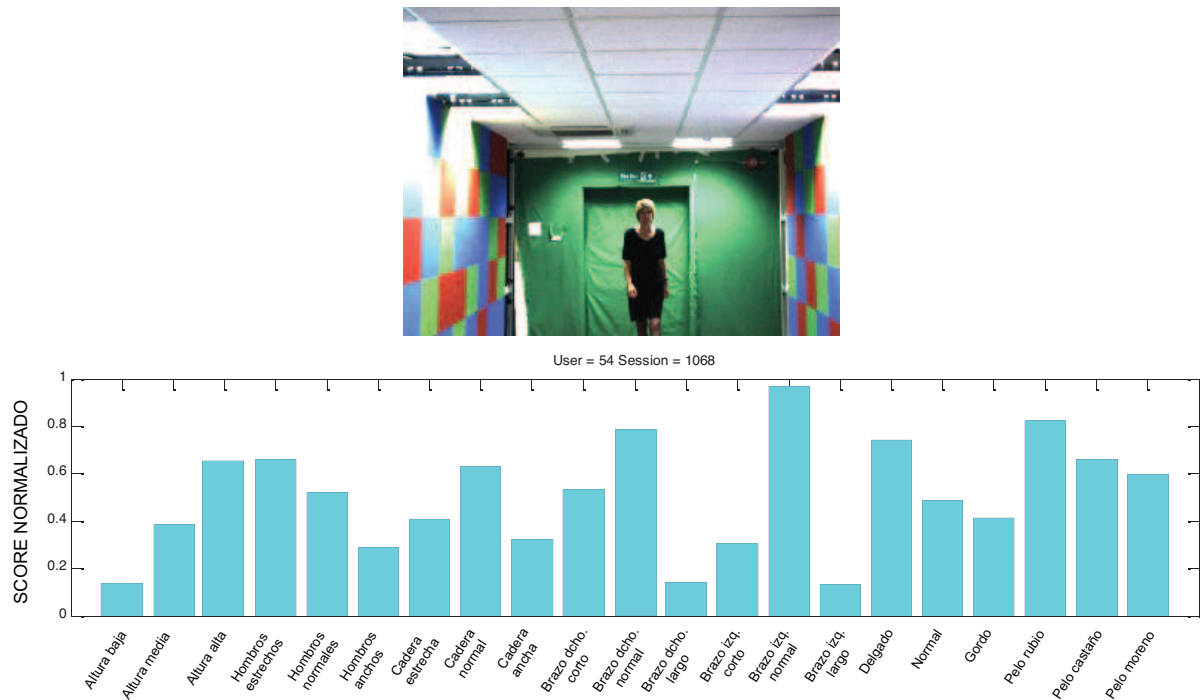


Figura 4.11: Sistema Real. Resultados del Estimador de atributos corporales.

	Imágenes Mal Clasificadas	% Acierto
<b>Altura</b>	24	94,6
<b>Hombros</b>	215	51,6
<b>Caderas</b>	200	54,9
<b>Largo Brazo derecho</b>	179	59,7
<b>Largo Brazo izquierdo</b>	147	66,9
<b>Envergadura</b>	198	55,4
<b>Color de Pelo</b>	143	67,7

Tabla 4.1: Error cometido por el Sistema Real para las features de medidas entre puntos, color de pelo y envergadura. Número de imágenes mal clasificadas y % Acierto.

Por otro lado, en la tabla 4.2 se muestran los errores cometidos por el Sistema Real con la feature de forma del cuerpo de la Fase II, comparándolos con la clasificación realizada en el ground-truth.

	Imágenes Mal Clasificadas	% Acierto
<b>Altura</b>	194	56,3
<b>Hombros</b>	213	52,0
<b>Caderas</b>	136	69,3
<b>Largo Brazo derecho</b>	193	56,5
<b>Largo Brazo izquierdo</b>	219	50,6
<b>Envergadura</b>	200	54,9
<b>Color de Pelo</b>	N/A	N/A

Tabla 4.2: Error cometido por el Sistema Real para la feature de forma del cuerpo de la Fase II. Número de imágenes mal clasificadas y % Acierto.

A la vista de los resultados no se debe emitir un juicio precipitado sobre cómo funcionan

los clasificadores, ya que existen otros factores que pueden determinarlo de una mejor forma. Simplemente se trata de una primera visión.

La comparación de estas dos fases de evaluación se realiza al final de esta sección, tras extraer **las curvas ROC**, ya que es la mejor forma de poder comparar lo bien o mal que clasifica el estimador en función de la feature que se utilice, es decir, su rendimiento.

Para evaluar nuestro sistema punto por punto se ha optado por calcular la **sensibilidad**, la **especificidad** y la **precisión** de cada uno de los clasificadores SVMs entrenados, utilizando la **matriz de confusión**, la cual se muestra en la figura 4.12.

		VALOR GROUND-TRUTH	
		True Positives	False Positives
PREDICCIÓN DEL SISTEMA		TP	FP
		False Negatives	True Negatives
		FN	TN
		Total Positives	Total Negatives

Figura 4.12: Matriz de confusión de un clasificador binario.

- La **sensibilidad** nos indica la capacidad de nuestro estimador para dar como casos positivos los casos realmente positivos (*True Positive*, TP); proporción de aciertos correctamente identificados. Es decir, la sensibilidad caracteriza la capacidad del sistema para detectar que un sujeto posee dicha característica.

$$Sensibilidad = \frac{TP}{TP+FN}$$

- La **especificidad** nos indica la capacidad de nuestro estimador para dar como casos negativos los casos realmente negativos (*True Negative*, TN); proporción de fallos correctamente identificados. Es decir, la especificidad caracteriza la capacidad del sistema para detectar la ausencia de una característica en los sujetos analizados.

$$Especificidad = \frac{TN}{TN+FP}$$

- Finalmente el rendimiento o **precisión** (*accuracy*) del sistema nos mostrará la tasa real de acierto del sistema desarrollado.

$$Precision = \frac{TN+TP}{TN+TP+FN+FP}$$

En la figura 4.13 se muestra la sensibilidad, especificidad y rendimiento calculados sobre el sistema desarrollado para cada una de las clases de los atributos para las features de la fase I. Además, en la tabla 4.3 se especifican estos valores.

Por ejemplo, para la Altura baja, los '*True positives*' (sensibilidad) se corresponderían con los casos en los que el sujeto es Bajo y se le clasifica como Bajo; y los '*True negatives*' (especificidad) se corresponderían con los casos en los que el sujeto no es Bajo y se le clasifica como Alto o Medio. Partiendo de ello, se realizan los cálculos indicados y se obtienen los valores para la sensibilidad, especificidad y precisión.

Por otra parte, en la figura 4.14 se muestra la sensibilidad, especificidad y rendimiento calculados sobre el sistema desarrollado para cada una de las clases de los atributos para las features de la fase II, adjuntando en la tabla 4.4 sus valores.

		Sensibilidad	Especificidad	Precisión
Altura	Bajo	0,93	0,99	0,99
	Medio	0,95	0,93	0,94
	Alto	0,93	0,96	0,95
Hombros	Estrechos	0,33	0,90	0,67
	Medios	0,69	0,57	0,61
	Anchos	0,55	0,79	0,73
Cadera	Estrecha	0,30	0,96	0,87
	Normal	0,82	0,38	0,57
	Ancha	0,36	0,87	0,64
Brazo derecho	Corto	0,33	0,96	0,82
	Normal	0,72	0,56	0,62
	Largo	0,61	0,80	0,73
Brazo izquierdo	Corto	0,41	0,96	0,86
	Normal	0,75	0,64	0,68
	Largo	0,70	0,84	0,78
Envergadura	Delgado	0,68	0,60	0,64
	Normal	0,42	0,74	0,63
	Grueso	0,36	0,89	0,81
Color de pelo	Rubio	0,71	0,85	0,81
	Castaño	0,36	0,85	0,71
	Moreno	0,85	0,79	0,82

Tabla 4.3: Valores de la Sensibilidad, Especificidad y Precisión para cada clase de los atributos del sistema para las features de la fase I.

		Sensibilidad	Especificidad	Precisión
Altura	Bajo	NaN	0,90	0,90
	Medio	0,58	0,55	0,56
	Alto	0,54	0,81	0,65
Hombros	Estrechos	NaN	0,79	0,79
	Medios	0,52	NaN	0,52
	Anchos	NaN	0,72	0,72
Cadera	Estrecha	NaN	0,90	0,90
	Normal	0,69	NaN	0,69
	Ancha	NaN	0,78	0,78
Brazo derecho	Corto	NaN	0,90	0,90
	Normal	0,56	NaN	0,56
	Largo	NaN	0,65	0,65
Brazo izquierdo	Corto	NaN	0,89	0,89
	Normal	0,51	0,49	0,51
	Largo	0,47	0,63	0,60
Envergadura	Delgado	0,54	NaN	0,54
	Normal	NaN	0,69	0,69
	Grueso	NaN	0,85	0,85

Tabla 4.4: Valores de la Sensibilidad, Especificidad y Precisión para cada clase de los atributos del sistema para las features de la fase II.

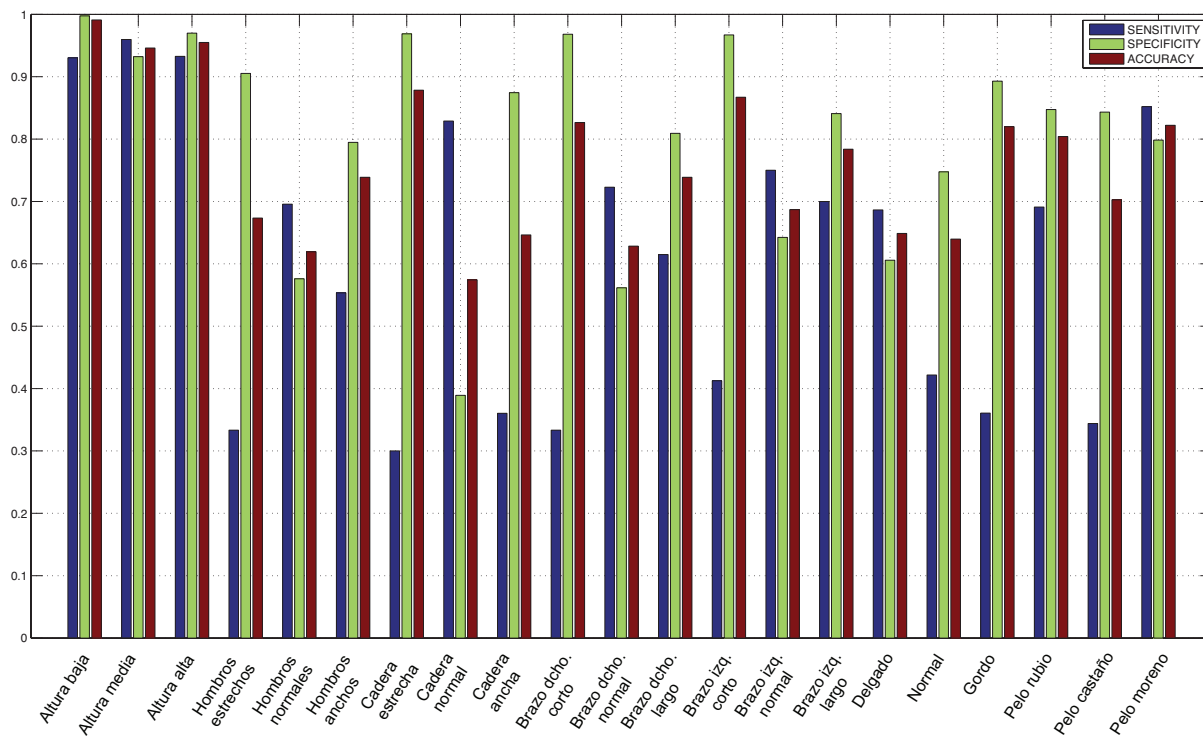


Figura 4.13: Sensibilidad, Especificidad y Precisión del Sistema Real para las features de la fase I.

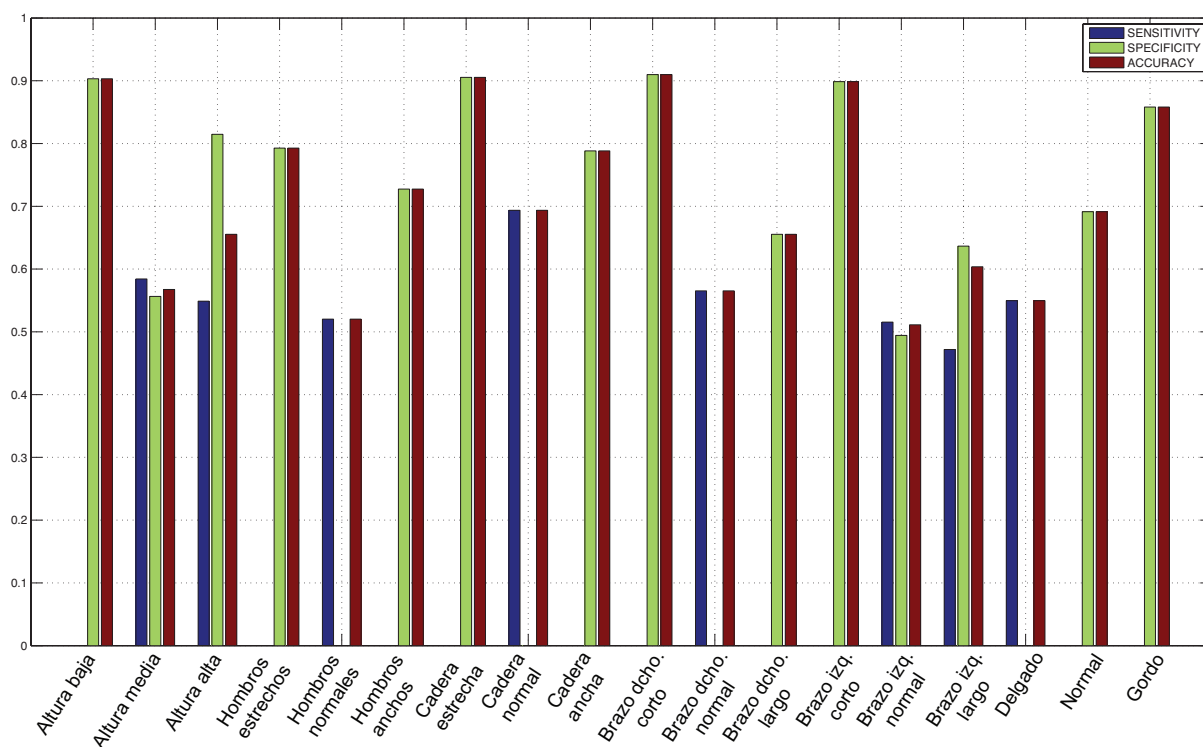


Figura 4.14: Sensibilidad, Especificidad y Precisión del Sistema Real para las features de la fase II.

Los valores indeterminados 'NaN' de la tabla 4.4 se deben a la aplicación de los valores obtenidos en la matriz de confusión sobre las ecuaciones indicadas previamente para la sensibilidad y especificidad.



Por ejemplo, para los casos en los que la Sensibilidad es indeterminada,  $True\ Positives = 0$  y  $False\ Negatives = 0$ . Por el contrario, para los casos en los que la Especificidad es indeterminada,  $True\ Negatives = 0$  y  $False\ Positives = 0$ . Estos casos no se han dado en la matriz de confusión de la tabla 4.3.

Por otro lado, para más detalle también se ha extraído la **curva ROC** del sistema desarrollado.

Una **curva ROC** (*Receiver Operating Characteristic*) es una representación gráfica de la sensibilidad frente a la especificidad para un sistema clasificador binario según se varía el umbral de discriminación.

Otra interpretación de este gráfico es la representación de la razón o ratio de verdaderos positivos (**VPR** = Razón de Verdaderos Positivos) frente a la razón o ratio de falsos positivos (**FPR** = Razón de Falsos Positivos) también según se varía el umbral de discriminación (valor a partir del cual decidimos que un caso es un positivo).

$$VPR = Sensibilidad$$

$$FPR = 1 - Especificidad$$

$$ROC = \frac{VPR}{FPR}$$

La diagonal resultante divide el espacio ROC. Los puntos por encima de la diagonal representan los buenos resultados de clasificación y los puntos por debajo de la línea son los resultados pobres. Por el contrario, una clasificación totalmente aleatoria daría un punto a lo largo de la línea diagonal, que se llama también línea de no discriminación. Una clasificación perfecta sería la que estuviera en la esquina superior izquierda.

Mostrando los resultados de la ROC para cada uno de los atributos del sistema con las features de la fase I, obtenemos las gráficas mostradas en la figura 4.15.

Si mostramos los resultados de la ROC para todos los atributos con la feature de la fase II, las gráficas que se obtienen son las que se muestran en la figura 4.16.

Como **conclusión**, a la vista de los resultados mostrados en las curvas ROC para los dos tipos de features utilizadas en la fase I y fase II, podemos realizar una **comparación** sobre cómo clasifica el sistema a partir de ellas:

- **Altura.**

El estimador funciona mucho mejor para la distancia entre los puntos etiquetados que para el vector de forma del cuerpo, ya que en este caso se aleja mucho más de la diagonal del espacio ROC. De hecho, la Altura es el atributo que mejor se clasifica de todo el sistema, y lo es utilizando las features de la Fase I.

- **Hombros.**

El estimador funciona mejor para la distancia entre los puntos etiquetados que para el vector de forma del cuerpo, ya que en este caso las curvas se encuentran tan cerca de la pendiente que es prácticamente una decisión aleatoria.

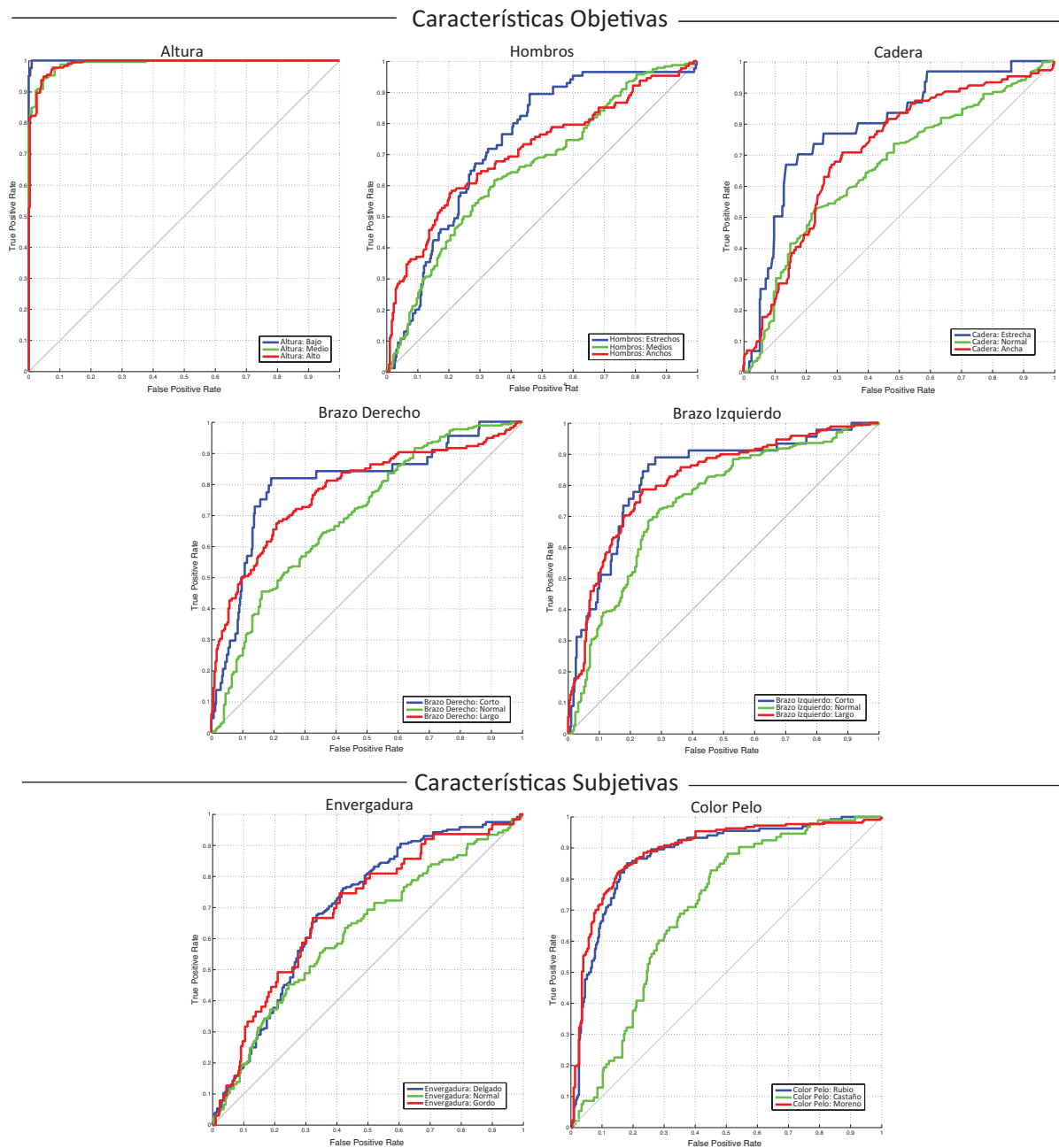


Figura 4.15: Curva ROC de todos los atributos del Sistema, utilizando las features de la fase I.

■ **Cadera.**

El estimador funciona de forma parecida para la detección de caderas normales y estrechas. Sin embargo, para la clasificación de caderas anchas, la feature de distancia entre puntos funciona mejor que la basada en el vector de Body Shape.

■ **Brazos.**

Para la feature de forma del cuerpo, la clasificación de los brazos es prácticamente aleatoria, únicamente clasifica mejor el brazo izquierdo cuando es corto.

Sin embargo, una vez más las features de distancia entre puntos permiten una mejor clasificación de ambos brazos.

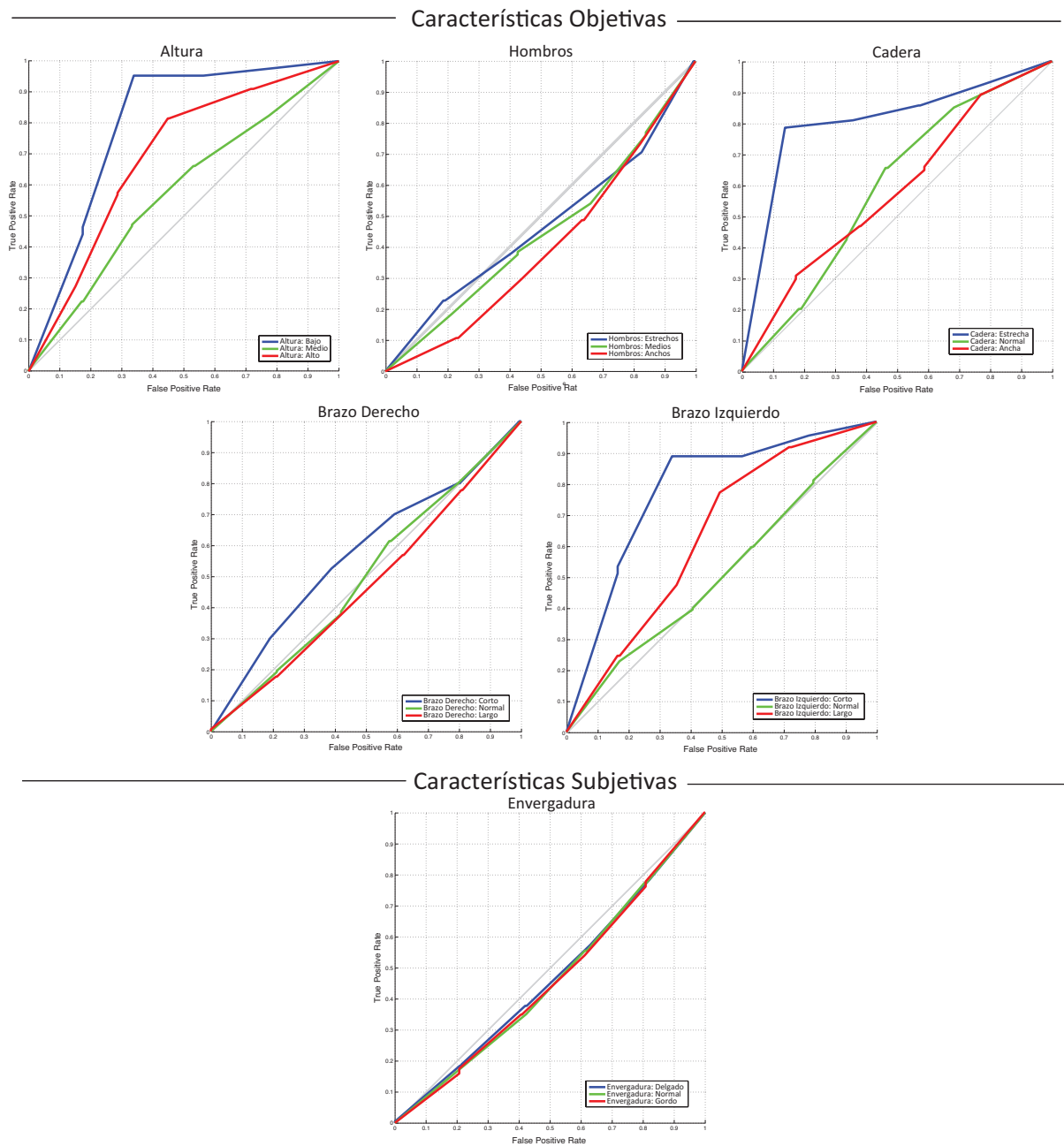


Figura 4.16: Curva ROC de todos los atributos del Sistema, utilizando las features de la fase II.

■ **Envergadura.**

La envergadura se clasifica mejor con la feature combinación del porcentaje frente-tamaño de imagen, altura, ancho de caderas y ancho de hombros, es decir, en la fase I. En la fase II se puede observar que la clasificación es completamente aleatoria, la curva ROC está totalmente pegada a la pendiente.

■ **Color del pelo.**

El color de pelo se clasifica bastante bien para la feature que contiene la información de color RGB, siendo el pelo castaño el que peor se clasifica, probablemente por ser el grupo intermedio entre los otros dos colores.

La feature de forma del cuerpo no se puede utilizar para clasificar correctamente el color del pelo, por tanto en este caso no se da lugar a la comparativa.

Como se puede ver, una propuesta de fusión entre los resultados obtenidos por ambos tipos de features mediante la suma de sus respectivos scores no conseguiría mejorar ninguna curva ROC. La clasificación del estimador no se vería mejorada para ningún atributo, ya que para todos ellos las features que permiten una mejor estimación son las de la Fase I.

Por tanto, se puede concluir que en términos generales la mejor clasificación para los atributos se obtiene con las features seleccionadas en este trabajo, siendo peor con el vector de concatenación de suma de píxeles de la figura humana.

Por último, a continuación de muestran algunos **resultados gráficos obtenidos por el Estimador de Atributos corporales** desarrollado en este proyecto, tanto para las features de la Fase I como para las de la Fase II, de forma que se puedan comparar los scores calculados para cada uno de ellos.

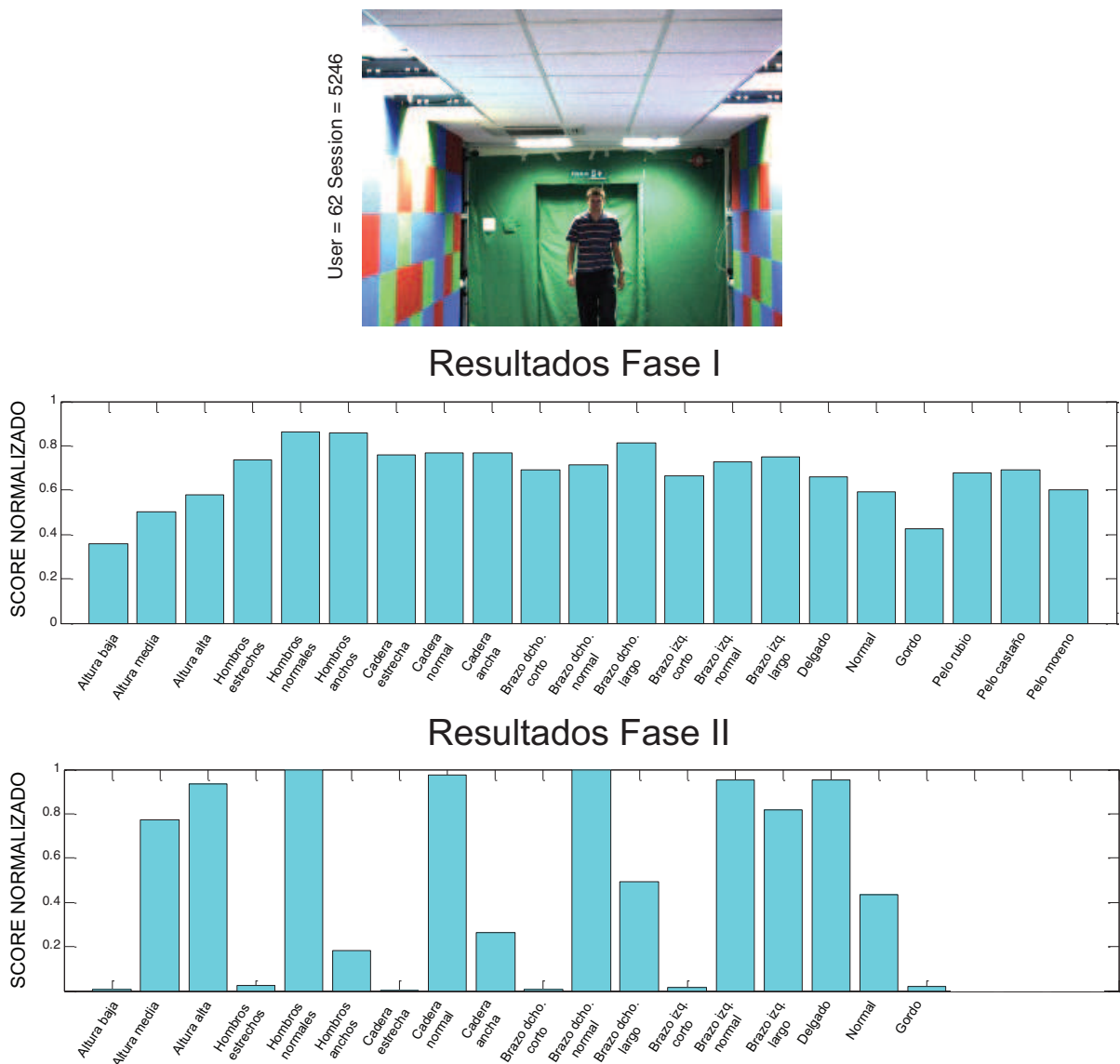


Figura 4.17: Ejemplo 1. Resultado obtenido por el Estimador de Atributos Corporales con las features de la Fase I y Fase II.



Resultados Fase I

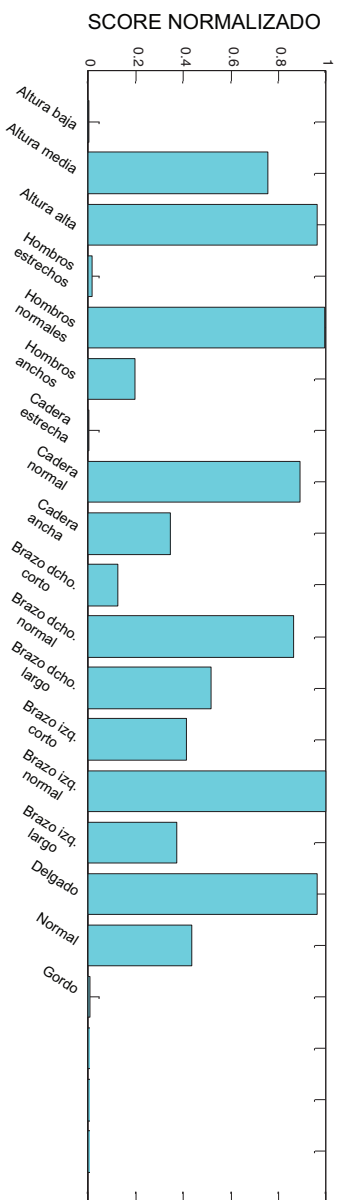
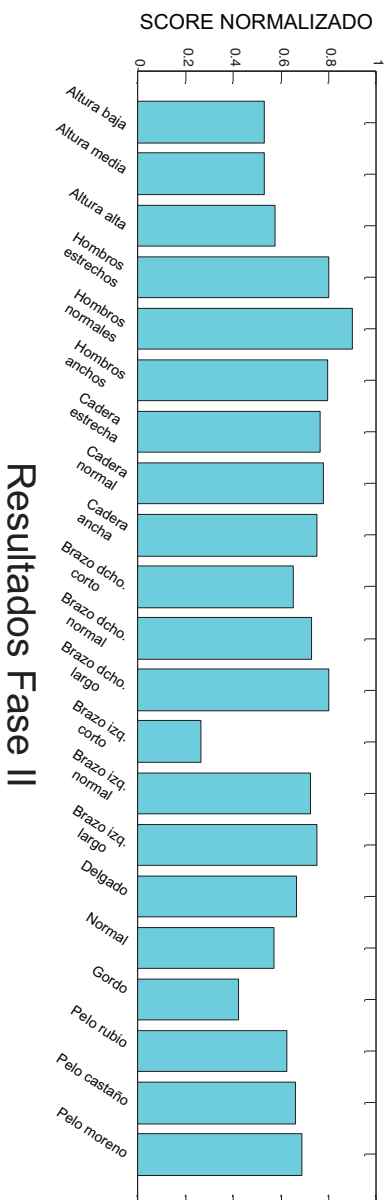


Figura 4.18: Ejemplo 2. Resultado obtenido por el Estimador de Atributos Corporales con las features de la Fase I y Fase II.

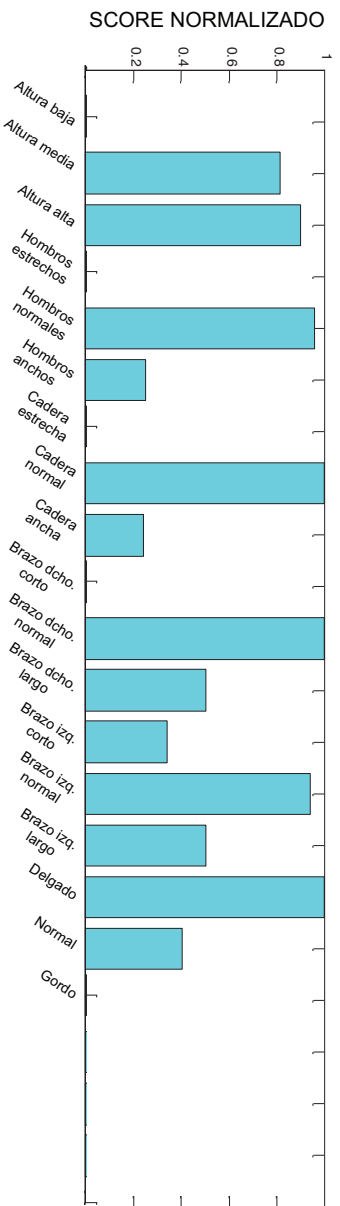
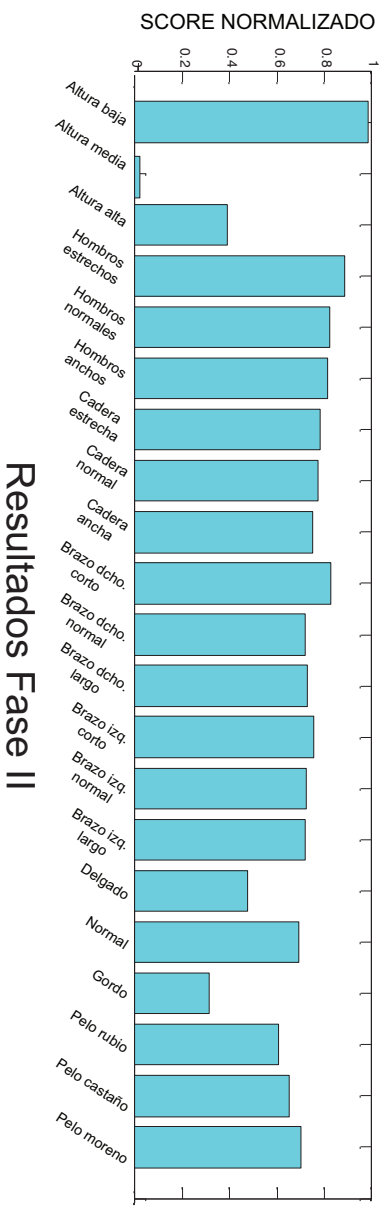


Figura 4.19: Ejemplo 3. Resultado obtenido por el Estimador de Atributos Corporales con las features de la Fase I y Fase II.

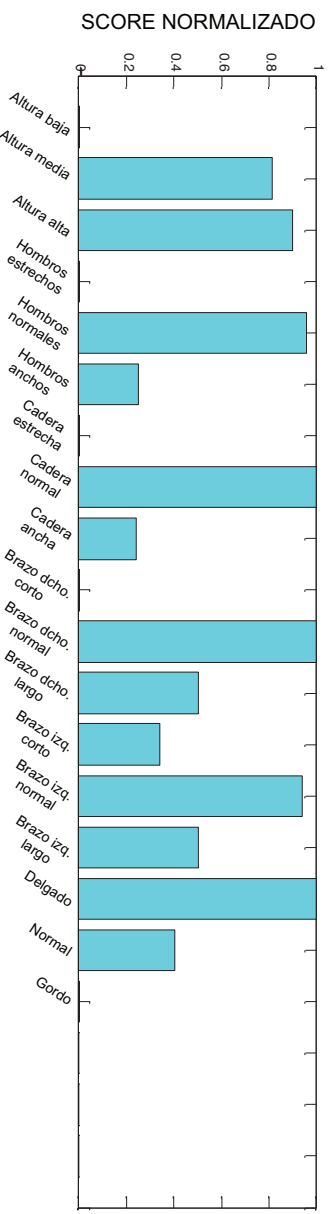
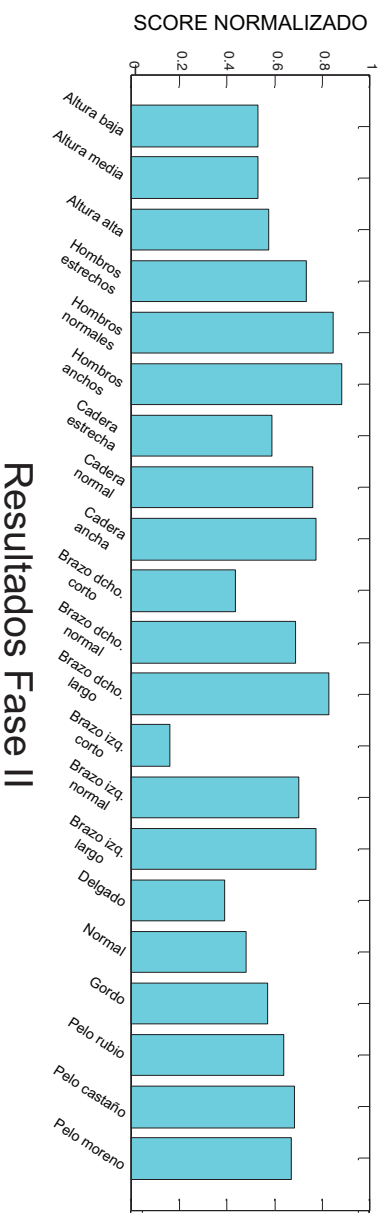


Figura 4.20: Ejemplo 4. Resultado obtenido por el Estimador de Atributos Corporales con las features de la Fase I y Fase II.







## 5 Conclusiones y Trabajo futuro

### 5.1. Conclusiones

---

El objetivo de este trabajo era desarrollar un **estimador automático de Soft biometrics corporales** a partir de imágenes de personas situadas a una determinada distancia. Esto es, atributos físicos que pueden ser inferidos de un sujeto al margen de la biometría clásica. De hecho, la consecución de esta estimación automática de características corporales puede suponer un incremento de la robustez de los sistemas biométricos clásicos como medidas complementarias.

De forma concreta, los **atributos Soft Biometrics** que se han seleccionado para ser estimados en este proyecto han sido: Altura, Ancho de Hombros, Ancho de Caderas, Largo de los brazos, Envergadura y Color de Pelo del sujeto.

Específicamente, contamos con dos tipos de características: **Objetivas** (aquellas que se pueden medir, como son la Altura o Ancho de Hombros) y **Subjetivas** (las que dependen de la percepción del usuario que clasifica, en concreto la Envergadura y el Color de pelo).

Durante el desarrollo del sistema implementado, se han abordado **diferentes formas de obtener features** que se puedan utilizar para clasificar a las personas según sus atributos corporales.

Por un lado, la primera selección de features para entrenar el estimador ha sido la siguiente: Para los atributos objetivos se han extraído automáticamente puntos clave en la imagen para después calcular la distancia entre ellos y utilizarla como feature. Para la envergadura se ha utilizado como feature específica el porcentaje de relación entre la figura del individuo y el tamaño total de la imagen, y para el color del pelo la feature escogida ha sido el contenido de color RGB.

Por otro lado y de forma separada, se han evaluado los resultados del estimador con otra feature común a todos los atributos (excepto el color del pelo): El vector de forma del cuerpo, concatenación de la suma de los píxeles de la figura en horizontal y vertical, el cual ha sido utilizado en otro trabajo relacionado con las Soft Biometrics [3].

Para cada atributo se han entrenado **tres clasificadores SVM**, uno para cada posible resultado. Es decir, existen tres grupos para cada atributo corporal. Para todos ellos se ha utilizado un **kernel gaussiano**, dada la limitación de utilizar un kernel lineal en features unidimensionales (medida de la distancia entre puntos).

El protocolo experimental ha seguido un *5-fold-cross validation*, de tal forma que la BBDD se ha dividido en 5 grupos de sujetos y para cada persona se han obtenido los resultados de test

a partir del entrenamiento realizado en el sistema con las restantes 4 partes de la BBDD.

La **evaluación de resultados** se ha llevado a cabo en **dos fases**, ejecutando de forma separada los resultados del estimador para la primera selección de features y los resultados para la feature de forma del cuerpo. Tras extraer las curvas ROC de ambos resultados del estimador, se ha concluido lo siguiente.

Los resultados de clasificación del estimador utilizando como feature el vector de forma del cuerpo no aportan un mejor rendimiento del sistema para la clasificación de ninguno de los atributos corporales. Para todos ellos, los mejores resultados de clasificación se obtienen con las features seleccionadas en este proyecto.

En el caso de los **atributos objetivos**, los resultados de clasificación del estimador utilizando las features de medidas entre puntos etiquetados automáticamente demuestran una mejor clasificación para todos ellos: Altura, Ancho de hombros, Ancho de cadera y Largo de brazos.

La **Envergadura** se clasifica mejor utilizando como feature la combinación del porcentaje frente-tamaño de la imagen, la altura, ancho de hombros y de cadera. Empeora su clasificación utilizando como feature el vector de forma del cuerpo, haciendo que sea prácticamente una decisión aleatoria.

Por último, el **Color de Pelo** únicamente se ha clasificado según la feature de color RGB, con una alta tasa de aciertos en la clasificación del estimador.

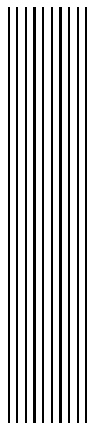
En conclusión, se ha desarrollado una primera aproximación al problema de estimación automática de atributos corporales a distancia, que necesitará ganar robustez con trabajos futuros que se realicen sobre el mismo.

## 5.2. Trabajo futuro

---

Como líneas de trabajo futuro para este proyecto se plantean las siguientes:

- Mejora de la clasificación del Ancho de Caderas y Largo de los brazos, dado que son los atributos para los que peores resultados de clasificación se han obtenido en este proyecto. Para ello algunas líneas de mejora son:
  - Extracción de los puntos de las caderas mejorando la detección de los bordes, evitando que se consideren como bordes el cambio de color de cualquier prenda que se lleve en el torso, lo cual conlleva a etiquetar los puntos de la cadera erróneamente.
  - Extracción de los puntos de los codos (los cuales sí han sido etiquetados de forma manual), ya que van a permitir una mejor aproximación de la medida del largo de los brazos.
- Introducir nuevas features que puedan mejorar la clasificación de los atributos corporales seleccionados en este trabajo.
- Añadir al estimador la extracción automática de nuevos atributos corporales subjetivos, como pueden ser la longitud del pelo y la etnia.
- Realizar una fusión completa de las features contempladas en este proyecto, de tal forma que para cada grupo de cada atributo se seleccione la opción que mejor funciona en la clasificación, y así globalmente los resultados sean mejores.
- Incluir estos atributos en sistemas de reconocimiento biométrico con aplicación en vídeo-vigilancia como características complementarias.



## Bibliografía

- [1] Daniel A Reid, Mark S Nixon, and Sarah V Stevenage. Soft biometrics; human identification using comparative descriptions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(6):1216–1228, 2014.
- [2] Pouya Samangouei, Vishal M Patel, and Rama Chellappa. Attribute-based continuous user authentication on mobile devices. In *Biometrics Theory, Applications and Systems (BTAS), 2015 IEEE 7th International Conference on*, pages 1–8. IEEE, 2015.
- [3] Ester Gonzalez-Sosa, Antitza Dantcheva, Ruben Vera-Rodriguez, Jean-Luc Dugelay, Francois Bremond, and Julian Fierrez. Image-based gender estimation from body and face across distances. In *23rd International Conference on Pattern Recognition (ICPR 2016): Image analysis and machine learning for scene understanding*, 2016.
- [4] Yining Li, Chen Huang, Chen Change Loy, and Xiaoou Tang. Human attribute recognition by deep hierarchical contexts. In *European Conference on Computer Vision*, pages 684–700. Springer, 2016.
- [5] Pedro Tome, Julian Fierrez, Ruben Vera-Rodriguez, and Mark S Nixon. Soft biometrics and their application in person recognition at a distance. *IEEE Transactions on Information Forensics and Security*, 9(3):464–475, 2014.
- [6] Google cloud vision api. <https://cloud.google.com/vision/>.
- [7] Microsoft computer vision api. <https://www.microsoft.com/cognitive-services/en-us/computer-vision-api>.
- [8] Andrew Loomis. *Figure drawing for all it's worth*. Viking Pr, 1943.
- [9] Microsoft computer vision api. <https://www.unioviedo.es>.
- [10] Sina Samangooei. *Semantic Biometrics*. PhD thesis, School of Electronics and Computer Science, University of Southampton, UK, May 2010.
- [11] Antitza Dantcheva, Petros Elia, and Arun Ross. What else does your biometric data reveal? a survey on soft biometrics. *IEEE Transactions on Information Forensics and Security*, 11(3):441–467, 2016.
- [12] Mark S Nixon, Paulo L Correia, Kamal Nasrollahi, Thomas B Moeslund, Abdenour Hadid, and Massimo Tistarelli. On soft biometrics. *Pattern Recognition Letters*, 68:218–230, 2015.

- [13] Daniel Reid, Sina Samangooei, Cunjian Chen, Mark Nixon, and Arun Ross. Soft biometrics for surveillance: an overview. *Machine learning: theory and applications*. Elsevier, pages 327–352, 2013.
- [14] Athira Nambiar, Alexandre Bernardino, and Jacinto Nascimento. Shape context for soft biometrics in person re-identification and database retrieval. *Pattern Recognition Letters*, 68:297–305, 2015.
- [15] Anil K. Jain, Sarat C. Dass, Karthik Nandakumar, and Karthik N. Soft biometric traits for personal recognition systems. In *Proceedings of International Conference on Biometric Authentication, Hong Kong*, pages 731–738, 2004.
- [16] Alphonse Bertillon. *Signaletic Instructions including the theory and practice of Anthropometrical Identification*. R.W. McClaughry Translation, The Werner Company, 1896.
- [17] Anil K. Jain, Karthik N, Xiaoguang Lu, and Unsang Park. Integrating faces, fingerprints and soft biometric traits for user recognition. In *In Proceedings of Biometric Authentication Workshop, LNCS 3087*, pages 259–269. Springer, 2004.
- [18] D.D. Heckathorn, R.S. Broadhead, and B. Sergeyev. A methodology for reducing respondent duplication and impersonation in samples of hidden populations. In *Proc. of Annual Meeting of the American Sociological Association*, pages 543–564, Toronto, Canada, 1997.
- [19] Unsang Park and Anil K. Jain. Face matching and retrieval using soft biometrics. *Trans. Info. For. Sec.*, 5(3):406–415, September 2010.
- [20] A. K. Jain and U. Park. Facial marks: Soft biometric for face recognition. In *IEEE International Conference on Image Processing*, pages 37–40, nov 2009.
- [21] Jung-Eun Lee, Anil K. Jain, and Rong Jin. Scars, marks and tattoos (smt): Soft biometric for suspect and victim identification. In *In Proc. Biometric Symposium, Biometric Consortium Conference*, pages 1–8, 2008.
- [22] S. Denman, C. Fookes, A. Bialkowski, and S. Sridharan. Soft-biometrics: Unconstrained authentication in a surveillance environment. In *Digital Image Computing: Techniques and Applications, 2009. DICTA '09.*, pages 196–203, 2009.
- [23] Pouya Samangooei and Rama Chellappa. Convolutional neural networks for attribute-based active authentication on mobile devices. In *Biometrics Theory, Applications and Systems (BTAS), 2016 IEEE 8th International Conference on*, pages 1–8. IEEE, 2016.
- [24] Joshua C Klontz, Brendan F Klare, Scott Klum, Anil K Jain, and Mark J Burge. Open source biometric recognition. In *Biometrics: Theory, Applications and Systems (BTAS), 2013 IEEE Sixth International Conference on*, pages 1–8. IEEE, 2013.
- [25] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Comput. Surv.*, 35(4):399–458, 2003.
- [26] M. S. Nixon and J. N. Carter. Automatic Recognition by Gait. *Proceedings of the IEEE*, 94(11):2013–2024, 2006.
- [27] Richard David Seely, Sina Samangooei, Lee Middleton, John Carter, and Mark Nixon. The university of southampton multi-biometric tunnel and introducing a novel 3d gait dataset. In *Proc. of IEEE Biometrics: Theory, Applications and Systems (BTAS'08)*, pages 1–6, September 2008.

- [28] R. D. Seely. *On a three-dimensional gait recognition system*. PhD thesis, School of Electronics and Computer Science, University of Southampton, UK, 2010.
- [29] C. N. Macrae and G. V. Bodenhausen. Social cognition: Thinking categorically about others. *Annual Review of Psychology*, 51(1):93–120, 2000.
- [30] J. Hewig, R. H. Trippe, H. Hecht, T. Straube, and W. H. R. Miltner. Gender differences for specific body regions when looking at men and women. *Journal of Nonverbal Behavior*, 32(2):67–78, 2008. Hewig, Johannes Trippe, Ralf H. Hecht, Holger Straube, Thomas Miltner, Wolfgang H. R.
- [31] Sergios Theodoridis and Konstantinos Koutroumbas. *Pattern Recognition, Fourth Edition*. Academic Press, 4th edition, 2008.
- [32] W. M. Campbell, J. P. Campbell, D. A. Reynolds, E. Singer, and P. A. Torres-carrasquillo. Support vector machines for speaker and language recognition. *Computer Speech and Language*, 20:210–229, 2006.
- [33] Carl Edward Rasmussen. *Gaussian processes for machine learning*. 2006.