



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
POSGRADO EN CIENCIA E INGENIERÍA DE LA COMPUTACIÓN

APRENDIZAJE E IDENTIFICACIÓN DE PERSONAS EN
AMBIENTES DINÁMICOS PARA ROBOTS DE SERVICIO

TESIS

QUE PARA OPTAR POR EL GRADO DE:
DOCTOR EN CIENCIAS (COMPUTACIÓN)

PRESENTA:

M. EN C. ARTURO RODRÍGUEZ GARCÍA

TUTOR:

DR. LUIS ALBERTO PINEDA CORTÉS
INSTITUTO DE INVESTIGACIONES EN MATEMÁTICAS APLICADAS Y EN SISTEMAS

MIEMBROS DEL COMITÉ TUTOR:

DR. JESÚS SAVAGE CARMONA
FACULTAD DE INGENIERÍA, UNAM

DR. RAÚL ROJAS GONZÁLEZ
UNIVERSIDAD LIBRE DE BERLÍN

DR. LUIS ENRIQUE SUCAR SUCCAR
INSTITUTO NACIONAL DE ASTROFÍSICA, ÓPTICA Y ELECTRÓNICA

DR. GIBRAN FUENTES PINEDA
INSTITUTO DE INVESTIGACIONES EN MATEMÁTICAS APLICADAS Y EN SISTEMAS

CIUDAD UNIVERSITARIA, CD. MX., JUNIO 2018



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

A mis padres María y Arturo, por su apoyo incondicional para cumplir este sueño.

A mis hermanas Alcis y Sandy, por su aliento y consejos para seguir en el mundo académico.

A mis sobrinos Iván, Ilse, Kiara y Alexa, por compartir conmigo sus dulces y risas en los momentos de mayor estrés.

A Isaac, Pablo, Nayeli y Chris, por haber tenido la dicha de conocerlos en el camino.

A Samantha y Alejandro, por enseñarme que los amigos verdaderos resisten el tiempo y la distancia.

Agradecimientos

El desarrollo de este trabajo fue posible gracias a la red de apoyo personal y profesional que tuve durante estos años de trabajo. Quiero empezar por agradecer al Grupo Golem por abrirme las puertas al mundo de investigación en robótica de servicio, en particular a mi asesor y líder del proyecto, el Dr. Luis Alberto Pineda Cortés, por darme la oportunidad de integrarme cuando apenas iniciaba mi maestría y por orientarme siempre de la mejor manera en aras de mi desarrollo académico.

También quiero dar las gracias al Dr. Raúl Rojas González, al Dr. Jesús Savage Carmona, al Dr. Luis Enrique Sucar Succar y al Dr. Gibran Fuentes Pineda por haber aceptado formar parte del comité de revisión de este proyecto. Sus observaciones e ideas han sido de gran ayuda para el desarrollo del mismo.

Quiero manifestar mi reconocimiento a todas las personas que han formado parte o que integran actualmente al Grupo Golem, pues sus aportaciones están presentes de una u otra forma en el robot actual y en el marco de trabajo utilizado en esta investigación. Quiero expresar mi agradecimiento y admiración al Dr. Gibran Fuentes, al Dr. Ivan Meza y al Dr. Caleb Rascón, quienes estuvieron siempre dispuestos a compartir sus conocimientos en los momentos en que surgían problemas técnicos en la implementación de mis ideas sobre el robot. También quiero mencionar a algunos miembros del equipo a quienes estimo profundamente por los momentos que hemos convivido y los viajes que hemos compartido: Lisset Salinas, Cinthya Ceja, Mauricio Reyes, Ricardo Cruz, Varinia Estrada, Esther Venegas, Saúl Martínez, Montserrat Alvarado, Iván Sánchez, Hernando Ortega, Ivette Vélez, Noé Hernández y Ángel Lee.

Quiero destacar a los alumnos de licenciatura, que como parte de estancias de investigación o de servicios sociales en el Grupo Golem, me ayudaron en la implementación y experimentación de algunos módulos del sistema: Sebastián

Chimal (identificación de gestos de la persona), Edgar Huerta (integración de *Microsoft Cognitive Services*) y Hugo Fernández (descripción de color y textura de la vestimenta).

Finalmente, quiero agradecer a la UNAM y a su Posgrado en Ciencia e Ingeniería de la Computación por la experiencia maravillosa de realizar un doctorado en uno de sus programas. Quiero hacer una mención especial al Consejo Nacional de Ciencia y Tecnología (CONACYT) que apoyó mi estudios de doctorado mediante una beca, y a los proyectos CONACYT 178673 y PAPIIT-UNAM IN-109816 por apoyar con recursos a esta tesis.

Resumen

En este trabajo de investigación se propone un esquema de comportamiento para aprendizaje en línea e identificación de personas en robots de servicio que operan en ambientes dinámicos en los que pueden ocurrir cambios en la configuración del escenario durante el desarrollo de la tarea, la presencia de múltiples personas y oclusiones parciales o totales de la persona que se desea reconocer. Este esquema combina rasgos biométricos, rasgos biométricos suaves y elementos contextuales, y aprovecha sus ventajas particulares al utilizarlos en el instante adecuado durante la ejecución del comportamiento de búsqueda de personas. Además, este esquema integra las habilidades motoras y verbales de un robot de servicio para mejorar la identificación correcta de una persona. El enfoque de biométricos suaves está basado en un modelo articulado 3D de la persona y los elementos contextuales se obtienen de la estructura de la tarea. El aprendizaje de la persona a partir de biométricos suaves genera una descripción que se almacena en la base de conocimiento del robot, la cual se puede usar con fines de identificación o para que el robot genere un retrato hablado. El esquema propuesto está embebido en un conjunto jerárquico y composicional de conductas básicas y su generalidad se evalúa resolviendo tareas de la competencia *RoboCup@Home* que involucran reconocimiento de personas. La estrategia resultante permite aprender personas en línea de manera rápida y natural, así como buscar e identificar a personas de manera correcta en escenarios con iluminación homogénea, funcionando incluso en situaciones en las que la persona no está encarando al robot.

Índice general

Agradecimientos	II
Resumen	IV
Lista de figuras	VIII
Lista de tablas	X
1. Introducción	1
1.1. Antecedentes	2
1.2. Delimitación del trabajo	4
1.3. Definición del problema	5
1.4. Objetivos	7
1.5. Metodología	7
1.6. Relevancia y contribución	8
1.7. Organización de la tesis	9
2. Aprendizaje e identificación de personas	10
2.1. Rasgos biométricos	10
2.2. Rasgos biométricos suaves	12
2.3. Elementos contextuales	16
2.4. Aprendizaje e identificación de personas en robots de servicio .	18
2.5. Recapitulación	25
3. Algoritmos de análisis para biométricos suaves	27
3.1. Detección y localización de biométricos suaves	27
3.2. Aprendizaje e identificación de biométricos suaves	35
3.3. Recapitulación	41
4. Sistema de aprendizaje en línea e identificación de personas basado en biométricos suaves	43
4.1. Detección, localización y rastreo de personas	43

<i>ÍNDICE GENERAL</i>	VI
4.2. Descripción de personas	47
4.3. Aprendizaje de personas	49
4.4. Identificación de personas	50
4.4.1. Identificación gruesa	50
4.4.2. Identificación fina	53
4.4.3. Identificación a partir de la descripción	54
4.5. Recapitulación	58
5. Integración del sistema en un robot de servicio	60
5.1. Conductas de reconocimiento de personas	61
5.2. Integración de habilidades motoras en las conductas de reconocimiento	63
5.3. SitLog, Modelos de Diálogo y Arquitectura Cognitiva Orientada a la Interacción	69
5.4. Recapitulación	71
6. Uso de las conductas de aprendizaje e identificación de personas basadas en biométricos suaves en tareas de robots de servicio	73
6.1. Integración en el robot Golem-III	73
6.2. Prueba <i>First & Second</i>	75
6.2.1. Escenario 1	77
6.2.2. Escenario 2	79
6.3. Prueba <i>Person Recognition Test 2016</i>	81
6.3.1. Escenario 1	82
6.3.2. Escenario 2	83
6.3.3. Escenario 3	83
6.3.4. Escenario 4	84
6.3.5. Ejecuciones de la prueba en competencias	86
6.4. Prueba <i>General Purpose Service Robot 2015</i>	89
6.4.1. Escenario 1, Versión 1	89
6.4.2. Escenario 1, Versión 2	89
6.5. Recapitulación	90
7. Conclusiones	92
7.1. Recapitulación y aportaciones	92
7.2. Resumen del algoritmo de aprendizaje e identificación basado en biométricos suaves y experimentos	94
7.3. Trabajo futuro	96
A. Clasificación de pose de brazos y cuerpo	97

ÍNDICE GENERAL

VII

Referencias

100

Índice de figuras

3.1. Tres formas diferentes de localizar un rasgo	29
4.1. Diagrama de bloques del Sistema de Reconocimiento de Personas Basado en Biométricos Suaves	44
4.2. Sistema de referencia del sensor de profundidad de Kinect 2.0	45
4.3. Puntos de interés del esqueleto generado por el SDK de Microsoft para Kinect 2.0.	46
4.4. Ocho observaciones del mismo usuario almacenadas en la BDUBBS	50
4.5. Comparación gruesa de dos personas	52
4.6. Conjunto de parches para comparación fina en dos vistas del mismo usuario	54
4.7. Comparación fina de dos personas	55
5.1. Biblioteca de conductas del robot Golem (Pineda <i>et al.</i> , 2013a) extendida con las conductas de biométricos suaves (marcadas con un asterisco)	62
5.2. Conducta <i>see_body</i>	63
5.3. Conducta <i>see_face</i>	64
5.4. Conducta <i>approach_person</i>	65
5.5. Modos basados en biométricos suaves de la conducta <i>see_person</i>	66
5.6. Modos basados en biométricos duros de la conducta <i>see_person</i>	67
5.7. Modos basados en la combinación de biométricos suaves y biométricos duros de la conducta <i>see_person</i>	68
5.8. Arquitectura Cognitiva Orientada a la Interacción (Pineda <i>et</i> <i>al.</i> , 2017b)	70
6.1. El robot de servicio Golem-III (Pineda <i>et al.</i> , 2017a)	75
6.2. Escenario 1 (Laboratorio de Golem)	78
6.3. Fases de localización de parches en el Escenario 1	79
6.4. Escenario 2 (Explanada del museo Universum)	81

6.5. Fases de localización de parches en el Escenario 2	81
6.6. Escenario 3 (Zona de trabajo del Torneo Mexicano de Robótica 2016)	84
6.7. Fases de localización de parches en el Escenario 3	84
6.8. Escenario 4 (Zona de trabajo de <i>RoboCup@Home</i> 2016)	85
6.9. Fases de localización de parches en el Escenario 4	86
6.10. Memorización de la persona en la competencia <i>RoboCup@Home</i> 2016	87
6.11. Fases de localización de parches en la competencia <i>RoboCup@Home</i> 2016	88
6.12. Identificación de la persona en la competencia <i>RoboCup@Home</i> 2016	88
A.1. Posiciones de brazos de una persona	98
A.2. Posiciones de cuerpo de una persona	99

Índice de cuadros

2.1. Comparación de rasgos biométricos de Jain, Ross y Prabhakar (2004:11)	11
2.2. Comparación de rasgos biométricos suaves de Dantcheva <i>et al.</i> (2010:742)	13
2.3. Dimensiones de análisis de rasgos de identificación	15
2.4. Lista de rasgos biométricos suaves	17
2.5. Ejemplos de instancias de clases de rasgos biométricos suaves .	18
2.6. Lista de clases de rasgos biométricos suaves	19
2.7. Estrategias de aprendizaje e identificación de personas usadas por robots de servicio (Parte 1)	22
2.8. Estrategias de aprendizaje e identificación de personas usadas por robots de servicio (Parte 2)	23
2.9. Estrategias de aprendizaje e identificación de personas usadas por robots de servicio (Parte 3)	24
2.10. Estrategias de aprendizaje e identificación de personas usadas por robots de servicio (Parte 4)	25
3.1. Revisión de detección basada en contexto de rasgos biométricos suaves en dos etapas.	36
3.2. Revisión de detección basada en contexto de rasgos biométricos suaves en tres o más etapas.	37
3.3. Representación computacional de propiedades de rasgos biométricos suaves localizados.	40
4.1. Vector de descripción después de aplicar las funciones de categorización.	57
6.1. Escenario 1 (Laboratorio de Golem).	78
6.2. Escenario 2 (Explanada del museo Universum).	80
6.3. Escenario 3 (Zona de trabajo del Torneo Mexicano de Robótica 2016).	83

ÍNDICE DE CUADROS

XI

6.4. Escenario 4 (Zona de trabajo de <i>RoboCup@Home</i> 2016). . . .	85
6.5. Resumen de los experimentos realizados en el Capítulo 6. . . .	90

Capítulo 1

Introducción

Con el objetivo de simplificar las tareas cotidianas de los seres humanos, desde hace varios años tanto la comunidad científica como el sector industrial han apostado por la investigación en torno al diseño y construcción de robots de servicio. A pesar de que el estado actual en el desarrollo de estas máquinas dista mucho del arquetipo de robot presente en la ciencia ficción, se han dado grandes avances en habilidades específicas (por ejemplo, en reconocimiento automático del habla, síntesis de voz, reconocimiento y manipulación de objetos, navegación en ambientes dinámicos, etc.) y en su integración para realizar tareas sencillas que son propias de los humanos (como los robots que puedan aspirar una habitación o preparar un tazón de cereal con leche).

Los robots de servicio están orientados a asistir a los humanos en tareas cotidianas del hogar, de la oficina o de espacios públicos como supermercados o museos. En la mayoría de las actividades que debe realizar un robot de servicio es necesaria la interacción con seres humanos. Incorporar el aprendizaje y la identificación de personas es indispensable para ejecutarlas de manera más realista y efectiva. Por ejemplo, sería indispensable que un robot que trabaja como mesero en un restaurante contara con las habilidades de memorizar a los clientes que le hacen un pedido y la capacidad de identificarlos posteriormente al momento de servirles.

La identificación automática de personas ha sido un tema de gran interés en el área de Visión por Computadora con resultados satisfactorios en diversas aplicaciones, incluyendo sistemas de seguridad y control de acceso, redes sociales y sistemas de videovigilancia. Este problema ha sido ampliamente abordado utilizando enfoques basados en rasgos biométricos y ha mostrado un buen funcionamiento en aplicaciones cuya información de entrada se obtiene mediante procesos controlados y con cooperación del

usuario. En años recientes, un tipo especial de rasgos llamados biométricos suaves, tal como el color de la ropa o una estimación aproximada del peso, han atraído la atención de distintas investigaciones debido a su utilidad en escenarios no controlados, en los cuales la identificación mediante el rostro no resulta apropiada. Otra estrategia para la identificación de personas son los elementos contextuales, que son factores circundantes como el lugar, el tiempo o la compañía de otras personas, que se pueden usar como pistas para inferir la identidad de una persona.

Los robots de servicio, al ser sistemas móviles que operan en tiempo real en ambientes dinámicos, agregan un nivel extra de dificultad al problema de identificación de personas. En este trabajo de investigación se propone un esquema de comportamiento para aprendizaje en línea e identificación de personas en robots de servicio que combina rasgos biométricos, rasgos biométricos suaves y elementos contextuales.

1.1. Antecedentes

Debido a la complejidad conceptual y técnica del problema de reconocimiento de personas, resulta conveniente disecarlo en un conjunto de subproblemas. Cielniak y Duckett (2004) descomponen el problema de reconocimiento de personas en tres subproblemas: detección (¿cuántas personas hay?), localización (¿en dónde están?) e identificación (¿quiénes son?). En este trabajo de investigación se agregan el aprendizaje, el rastreo, la descripción y la recuperación a la lista de subproblemas de reconocimiento de personas:

- Detección.- Consiste en determinar si un objeto presente en la escena es o no una persona.
- Localización.- Consiste en calcular la posición de una persona en determinado sistema de referencia espacial.
- Aprendizaje. - Es la capacidad de adquirir y almacenar las características de una persona que servirán para identificarla posteriormente. Este conjunto de características se asocia con una etiqueta que funciona como identificador (como el nombre de la persona). Se puede dar fuera de línea (por ejemplo, suministrando previamente un conjunto de fotos de la persona) o en línea (haciendo que el sistema obtenga las características de la persona durante su funcionamiento). El aprendizaje es a corto plazo si el sistema

necesita la información sólo durante un lapso breve después del cual puede desechar las características obtenidas. Por el contrario, en el aprendizaje a largo plazo es necesario almacenar las características con fines no inmediatos, por lo que es importante enfocarse en las características persistentes y desechar las que son inestables a través del tiempo, como la ropa o el corte de cabello.

- **Identificación.**- Consiste en decidir si una persona que se detecta en la escena es la misma que la detectada y aprendida anteriormente.
- **Rastreo.**- Es determinar y/o predecir la trayectoria que sigue una persona al moverse en la escena.
- **Descripción.** - Consiste en generar una expresión en lenguaje natural que permita la identificación única de la persona.
- **Recuperación.** - Dada una descripción de una persona, consiste en determinar si alguien en la escena la satisface.

Con el objetivo de resolver diferentes subproblemas del reconocimiento de personas han sido propuestas distintas estrategias. Cada estrategia consiste en usar un sensor (por ejemplo, cámara monocular, cámara estereoscópica, láser, cámara térmica, sensor infrarrojo, sonar, etc.) y enfocarse en una parte específica del cuerpo humano (como el rostro, las piernas, la silueta del cuerpo, etc.). Posteriormente, se aplica un algoritmo específico para procesar la información obtenida. Por ejemplo, la detección de personas se puede realizar con una cámara monocular y un clasificador de caras usando el algoritmo de Viola y Jones (2001). El aprendizaje y la identificación se pueden realizar utilizando el mismo sensor y enfocándose al rostro, pero con un algoritmo diferente, por ejemplo, el algoritmo Eigenfaces (Eigenrostros) descrito por Turk y Pentland (1991).

La identificación de personas se realiza normalmente por medio de rasgos biométricos, los cuales son características fisiológicas o de comportamiento usadas para confirmar la identidad de un individuo, por ejemplo, el ADN, la voz, la huella dactilar, la huella palmar, la geometría de la mano, el iris, la retina, el rostro, la forma de la oreja, el olor, un termograma facial, la firma, el modo de teclear al escribir o la forma de caminar (Jain, Ross y Prabhakar, 2004).

En años recientes, los rasgos llamados biométricos suaves han sido objeto de interés en algunos grupos de investigación. Los biométricos suaves son

rasgos que proveen información acerca del individuo, pero que no compiten con los rasgos biométricos en términos de su distintividad y permanencia (Jain, Dass y Nandakumar, 2004); por ejemplo, el color del cabello, el color de la ropa o el género. Los biométricos suaves pueden ser descritos verbalmente (Reid *et al.*, 2013) y pueden ser clasificados en categorías predefinidas por humanos (Dantcheva *et al.*, 2010); por ejemplo, el largo del cabello se puede clasificar como corto, medio o largo. Un mismo rasgo se puede utilizar como biométrico o como biométrico suave. Por ejemplo, el rostro humano se puede analizar como biométrico en términos de las distancias entre un conjunto de características o puntos salientes del rostro. Pero al mismo tiempo, se puede describir como un biométrico suave si su forma es cuadrada, ovalada, redonda, oblonga, triangular, etc.

Los elementos contextuales también son importantes en la identificación de una persona. Por ejemplo, si sabemos de antemano cuándo y dónde fue tomada una foto, podemos utilizar esa información como contexto temporal y espacial para inferir quién está en la foto (Choi *et al.*, 2010). El contexto social también puede ser usado; por ejemplo, es muy común que una persona en una imagen tenga algún tipo de relación con otras que lo acompañan en la foto o con el fotógrafo (Gallagher y Chen, 2009b).

En el capítulo 2, se presenta una revisión bibliográfica sobre los conceptos de biométricos, biométricos suaves y elementos contextuales en el marco de la identificación visual de personas. En el capítulo 3 se profundiza en los algoritmos y estrategias para el reconocimiento mediante biométricos suaves.

1.2. Delimitación del trabajo

El problema de aprendizaje e identificación de personas es amplio y desafiante, y se puede abordar desde distintas disciplinas, incluyendo Biometría, Antropología, Neurociencias, Psicología, Derecho, Filosofía, etc. Este trabajo de investigación se centra específicamente en el estudio del aprendizaje e identificación automático de personas mediante modelos computacionales.

Bajo este enfoque es posible encontrar una gran cantidad de literatura, la gran mayoría asociada a implementaciones realizadas en sistemas fijos de cómputo estándar. Sin embargo, cuando el estudio se enmarca en el contexto de robots de servicio la literatura se reduce considerablemente. Estudiar los procesos de aprendizaje e identificación de personas en un robot de servicio

implica:

1. Que los procesos se realizan en un sistema móvil.
2. Que los procesos deben generar soluciones en tiempo real.
3. Que los procesos se dan en un ambiente dinámico.

Además, en esta investigación se tomarán las siguientes restricciones:

1. El aprendizaje será en línea, lo que incrementa el nivel de dificultad ya que el robot debe crear en forma dinámica la base de conocimiento de personas.
2. El aprendizaje será con fines de identificación a corto plazo, con el objetivo de dar espacio al uso de biométricos suaves como el color de la ropa.
3. La identificación se hará exclusivamente mediante sistemas de visión computacional.

En el capítulo 2 se incluye una discusión sobre la manera en que los tres grupos de elementos de identificación (biométricos, biométricos suaves y elementos contextuales) se pueden utilizar en aplicaciones de robots de servicio.

1.3. Definición del problema

Muchas de las estrategias para aprendizaje e identificación de personas funcionan adecuadamente cuando se utilizan desde un sistema fijo y dentro de un ambiente controlado. Sin embargo, los robots de servicio se enfrentan a distintos retos:

- Los cambios en la apariencia de la persona
 - La postura.- ya que la figura humana es articulada puede tomar distintas posiciones (sentado, agachado, con los brazos cruzados, haciendo un gesto con los brazos, etc.).
 - La orientación con respecto al dispositivo sensor.- es decir, puede estar de frente, de costado, dando la espalda, etc. Algunos detalles específicos de la persona como el rostro o el estampado de la playera sólo son apropiados si la persona está orientada de frente al sensor.

- La distancia.- si está muy lejos del dispositivo sensor, reconocer detalles específicos como el rostro resulta muy difícil; por otra parte, si está demasiado cerca se dificulta obtener una vista del cuerpo completo.
- El movimiento.- el cual dificulta que se obtengan imágenes nítidas y enfocadas.
- Las características del ambiente
 - Las modificaciones.- un ambiente dinámico posibilita la aparición de obstáculos temporales o permanentes entre el sensor y la persona, quien se bloquea en forma parcial o total.
 - La iluminación no controlada.- la presencia de múltiples fuentes de iluminación es un gran reto, pues el color o la textura de un objeto pueden parecer diferentes dependiendo de la cantidad de luz y del modo de incidencia.
 - La presencia de múltiples personas.- que exige estrategias de identificación más robustas para discernir entre varios candidatos.
 - La contaminación visual.- algunos ambientes tienen una presencia excesiva y desordenada de objetos, lo cual complica el análisis de la escena.
- La necesidad de una interacción humano-robot ágil
 - El costo de aprender a la persona.- si el aprendizaje de la persona es en línea es necesario adquirir de manera ágil información sobre ella, evitando someter al sujeto desconocido a procedimientos que pueden resultar incómodos, como pedirle que se mantenga inmóvil mientras se le toman varias fotos.
 - El costo computacional.- en general, los algoritmos para el reconocimiento de personas son demasiado complejos. Incrementar la resolución de las lecturas, disminuir el periodo de muestreo o combinar múltiples estrategias puede brindar información más precisa de los eventos que ocurren en el mundo, pero con un costo en el tiempo de procesamiento.
- Las limitaciones de los dispositivos sensores.- el éxito en el reconocimiento de personas está relacionado con la confiabilidad de las lecturas del mundo que se realizan mediante los dispositivos sensores. Se debe tener en cuenta que todos los sensores funcionan adecuadamente

sólo dentro de cierto rango de operación (distancia, condiciones de iluminación, etc.), así como la presencia de ruido en la señal.

Como resultado de estos problemas es necesario que el reconocimiento de personas utilice un marco de trabajo que, sin perder la robustez necesaria para funcionar en ambientes complejos, sea flexible y permita reconocer al individuo en una amplia gama de situaciones.

1.4. Objetivos

El objetivo general de este trabajo de investigación es diseñar un sistema de visión computacional para sistemas móviles que realice el aprendizaje en línea y la identificación de personas en ambientes dinámicos.

Los objetivos específicos son:

- Construir un sistema para el aprendizaje e identificación de personas a partir de biométricos suaves.
- Modelar las conductas de un robot de servicio que realicen el reconocimiento de personas a partir de biométricos suaves.
- Generar una estrategia completa de identificación de personas que combine el uso de biométricos, biométricos suaves y elementos contextuales, así como las conductas verbales y motrices del robot.
- Integrar la estrategia desarrollada en tareas de robots de servicio.

1.5. Metodología

El desarrollo del sistema de aprendizaje e identificación de personas a partir de biométricos suaves está basado en un modelo articulado 3D de la persona que se utiliza como contexto para extraer parches en zonas específicas del cuerpo. Se desarrollaron tres modos de identificación, el modo grueso que funciona a partir de la comparación del color de la playera, el modo fino que se basa en la comparación de múltiples parches pequeños etiquetados de acuerdo a la región del cuerpo de la que se obtuvieron, y el modo semántico que se basa en la descripción de la persona almacenada en la base de conocimiento del robot (Pineda *et al.*, 2017b). Este sistema se explica a detalle en el Capítulo 4.

Siguiendo el concepto y la estructura funcional de un robot de servicio (Pineda *et al.*, 2013a), se definieron las conductas de aprendizaje e identificación que utilizan el sistema anterior. La biblioteca de conductas descrita está implementada en Sitlog (Pineda *et al.*, 2013b), un lenguaje de programación para tareas de robots de servicio. Cada conducta corresponde a un Modelo de Diálogo, el cual es un protocolo de interacción compuesto por un conjunto de situaciones, que a su vez son estados de información definidos por un conjunto de expectativas y acciones potenciales (*Ibidem*). La relación entre percepción, estructura de tarea y acción sigue una Arquitectura Cognitiva Orientada a la Interacción (Pineda *et al.*, 2011). A nivel de conducta, se desarrolló una estrategia que combina la identificación mediante elementos biométricos, elementos biométricos suaves y elementos contextuales, y que además integra las habilidades motoras y verbales de un robot para ayudarlo en el proceso de búsqueda de la persona. El Capítulo 5 aborda todos los aspectos de la integración del sistema de aprendizaje e identificación de personas al robot de servicio.

Estas conductas fueron integradas y probadas en el robot de servicio Golem III y utilizadas para resolver distintas pruebas de la competencia *RoboCup@Home*.¹ La descripción y resultados de estas aplicaciones se describen en el Capítulo 6.

1.6. Relevancia y contribución

El aprendizaje en línea y la identificación automática de personas en robots de servicio son problemas abiertos y de interés para la comunidad científica. Por ejemplo, en la competencia internacional *RoboCup@Home* estas habilidades forman parte de las pruebas básicas. Las soluciones que se han brindado hasta el momento en este torneo funcionan bajo restricciones fuertes. En la prueba *Person Recognition* de *RoboCup@Home 2016* la identificación se realizó forzando a que la persona estuviera encarando al robot y a una distancia específica en un escenario con pocas personas (en promedio seis) y con iluminación homogénea.

Muchas preguntas de investigación siguen abiertas en relación a los procesos de aprendizaje e identificación: ¿qué estrategia se debe usar para resolver todos los retos que enfrenta un robot de servicio?, ¿qué algoritmos serían adecuados para escenarios más hostiles?, ¿cómo integrar los procesos

¹<http://www.robocupathome.org/>

en la arquitectura cognitiva del robot?, ¿cómo integrar estos procesos dentro de tareas complejas que requieran otras habilidades del robot?, etc.

1.7. Organización de la tesis

El presente trabajo de investigación se encuentra organizado en los siguientes capítulos:

- Capítulo 2.- Se abordan a detalle el aprendizaje y la identificación de personas, explicando de manera más profunda las tres estrategias introducidas en la sección 1.1 (los rasgos biométricos, los rasgos biométricos suaves y los elementos contextuales) y se analiza su utilidad en el contexto de robots de servicio.
- Capítulo 3.- Se presenta una revisión de la literatura sobre los algoritmos para analizar rasgos biométricos suaves.
- Capítulo 4.- Se describe el sistema de aprendizaje en línea e identificación de personas basado en biométricos suaves que se propuso e implementó en este trabajo de investigación.
- Capítulo 5.- Se explica la forma en que el sistema del capítulo 4 se puede integrar en un robot de servicio, usando como marco de trabajo los siguientes elementos: SitLog, Modelos de Diálogo y la Arquitectura Cognitiva Orientada a la Interacción.
- Capítulo 6.- Se describe cómo el sistema se utilizó en el robot Golem-III para resolver tareas que involucran reconocimiento de personas y se reportan los resultados obtenidos.
- Capítulo 7.- Se presentan las conclusiones y el trabajo futuro.

Capítulo 2

Aprendizaje e identificación de personas

En el capítulo anterior se definieron el aprendizaje y la identificación como subproblemas del reconocimiento de personas. Además, se presentaron de manera introductoria tres grupos de elementos que distinguen a una persona del resto: rasgos biométricos, rasgos biométricos suaves y elementos contextuales. En este capítulo se profundizan los conceptos anteriores y se realiza un análisis sobre su utilidad en robots de servicio tomando en consideración los distintos retos que fueron presentados en la sección 1.3. Además, se presenta una revisión del estado del arte sobre las técnicas utilizadas por robots de servicio para aprender e identificar personas.

2.1. Rasgos biométricos

De acuerdo con Jain, Ross y Prabhakar (2004) la identificación mediante biométricos se enfoca en las características fisiológicas y/o de comportamiento de la persona, y no en sus pertenencias (por ejemplo, una tarjeta de identificación) o lo que recuerda (como una contraseña). Estos autores identifican siete factores para determinar qué tan adecuados son los distintos rasgos biométricos al ser usados en una aplicación:

- Universalidad.- se refiere a si todas las personas poseen ese rasgo.
- Distintividad.- el rasgo debe diferenciar de manera suficiente a dos personas cualquiera.
- Permanencia.- el rasgo debe ser suficientemente invariante durante un periodo considerable de tiempo.

- Colectabilidad.- la característica se debe poder medir en forma cuantitativa.
- Desempeño.- se refiere a la precisión y velocidad de los sistemas al utilizar el rasgo.
- Aceptabilidad.- indica el grado en el que las personas están dispuestas a permitir la medición del rasgo en sus vidas diarias.
- Elusión.- es la facilidad para burlar un sistema basado en ese rasgo mediante métodos fraudulentos.

En el Cuadro 2.1 se presenta una comparación de rasgos biométricos propuesta por Jain, Ross y Prabhakar (2004) de acuerdo a estos factores.

Cuadro 2.1: Comparación de rasgos biométricos de Jain, Ross y Prabhakar (2004:11) de acuerdo a su universalidad (Un), distintividad (Di), permanencia (Pe), colectabilidad (Co), desempeño (De), aceptabilidad (Ac) y elusión (El) usando una escala discreta de valores: alto (A), medio (M) y bajo (B).

Rasgo biométrico	Un	Di	Pe	Co	De	Ac	El
ADN	A	A	A	B	A	B	B
Oreja	M	M	A	M	M	A	M
Rostro	A	B	M	A	B	A	A
Termograma facial	A	A	B	A	M	A	B
Huella dactilar	M	A	A	M	A	M	M
Modo de caminar	M	B	B	A	B	A	M
Geometría de la mano	M	M	M	A	M	M	M
Venas de la mano	M	M	M	M	M	M	B
Iris	A	A	A	M	A	B	B
Modo de teclear	B	B	B	M	B	M	M
Olor	A	A	A	B	B	M	B
Huella palmar	M	A	A	M	A	M	M
Retina	A	A	M	B	A	B	B
Firma	B	B	B	A	B	A	A
Voz	M	B	B	M	B	A	A

2.2. Rasgos biométricos suaves

En años recientes, el concepto de biométrico suave se ha hecho presente para referirse a rasgos que no forman parte de la literatura estándar de Biometría y de sus ahora también llamados rasgos biométricos duros. Los biométricos suaves surgen como propuestas alternativas para realizar la identificación de personas. Sin embargo, la definición del concepto todavía no es precisa y se presentan discrepancias entre los autores que lo abordan.

La primera definición de este concepto se propuso por Jain, Dass y Nandakumar (2004:732), quienes señalan que los rasgos biométricos suaves son “características que proveen cierta información del individuo, pero que carecen de distintividad y permanencia para diferenciar suficientemente a cualquier par de individuos”.¹ De acuerdo con estos autores, los biométricos suaves (como el género, la edad, la etnicidad, el peso, etc.) son auxiliares a los identificadores biométricos primarios (como la huella dactilar o el rostro) y su papel dentro de la identificación de personas es secundario. Esta noción de descriptor secundario es diferente a la noción de descriptor semántico que se discute a continuación.

Dantcheva *et al.* (2010:740) definen a los rasgos biométricos suaves como “características humanas físicas, de comportamiento o adheridas, clasificables en categorías predefinidas aceptadas por los humanos”. Por ejemplo, el color del cabello se puede clasificar en negro, castaño, rojizo, etc. En la Tabla 2.2 se presenta una comparación de rasgos biométricos suaves propuesta por estos autores.

Reid *et al.* (2013:327) establecen que los rasgos biométricos suaves “se pueden describir típicamente usando etiquetas y medidas entendibles por humanos”. De acuerdo con estos autores, los humanos usan de manera natural los biométricos suaves para identificar y describir verbalmente a otros. Por el contrario, los rasgos biométricos duros no siempre se pueden traducir a formas descriptivas.

Las definiciones de los dos párrafos anteriores comparten la idea de describir al rasgo de manera entendible por un humano. En la definición de Dantcheva *et al.* (2010) la descripción del rasgo es mediante categorías. Por ejemplo, la estatura de una persona se puede clasificar como baja, mediana o

¹Esta definición puede causar confusión si no se lee de manera adecuada. Un parafraseo es el siguiente: los rasgos biométricos suaves son aquellos que tienen baja distintividad o poca permanencia o ambas carencias simultáneamente.

Cuadro 2.2: Comparación de rasgos biométricos suaves de Dantcheva *et al.* (2010:742) de acuerdo a su ubicación (Ub), naturaleza del valor (Nv), permanencia (Pe), distintividad (Di) y percepción subjetiva (Ps) usando una escala discreta de valores para los tres últimos: alto (A), medio (M) y bajo (B).

Rasgo biométrico suave	Ub	Nv	Pe	Di	Ps
Color de piel	Cara	Continuo	M	B	A
Color de cabello	Cara	Continuo	M	M	M
Color de ojos	Cara	Continuo	A	M	M
Barba	Cara	Binario	B/M	B	M
Bigote	Cara	Binario	B/M	B	M
Forma facial	Cara	Discreto	A	A	A
Maquillaje	Cara	Discreto	B	B	M
Etnicidad	Cara	Discreto	A	M	M
Marcas	Cara/Cuerpo	Discreto	A	M/A	B
Género	Cara/Cuerpo	Binario	A	B	B
Edad	Cara/Cuerpo	Continuo	B/M	M	M
Estatura	Cuerpo	Continuo	M/A	M	M
Peso	Cuerpo	Continuo	B/M	M	M
Forma de caminar	Cuerpo	Continuo	M	M	A
Forma del cuerpo	Cuerpo	Discreto	M	M	M
Color de ropa	Accesorio	Discreto	B	M	M
Lentes	Accesorio	Binary	B/M	B	B

alta. Por otra parte, la definición de Reid *et al.* (2013) no utiliza la noción de categorías, por lo que cualquier expresión entendible por humanos podría ser usada para describir el rasgo. Por ejemplo, se puede decir que la estatura de una persona es de aproximadamente metro y medio. En ambos casos subyace una definición de biométrico suave como descriptor semántico.

No todas las descripciones verbales entendibles por humanos son de interés para considerar a un rasgo como biométrico suave, por lo que se propone restringirlas mediante los siguientes criterios, los tres primeros obligatorios y el último opcional en función de la aplicación en la que se utilizan:

1. Naturalidad.- las descripciones deben ser similares a las usadas por los humanos de manera cotidiana. Con esta restricción se pueden filtrar descripciones forzadas como una explicación verbal minuciosa de las

líneas de una huella dactilar. Aunque es posible tomar una lupa y tratar de describir los contornos, las bifurcaciones o los cruces de las líneas, esto no es algo que hacen los humanos de manera cotidiana.

2. Utilidad.- la descripción debe ser útil para fines de identificación. Un ejemplo de descripción que no cumple esta condición es indicar que una persona tiene venas en la mano que se ramifican y transparentan.
3. Objetividad.- es importante excluir descripciones subjetivas (por ejemplo, decir que una persona tiene ojos hermosos y mirada profunda). Por esta razón, las descripciones que usan lenguaje literario no serían útiles.
4. Entendimiento común.- el vocabulario utilizado debe ser entendido por la mayoría de las personas. Con esta restricción se descartan descripciones que usan lenguaje técnico (por ejemplo, expresar que el somatotipo de una persona es ectomórfico). Si se cumple esta restricción, las descripciones verbales de los rasgos se pueden utilizar en aplicaciones que involucran interacción con personas no especializadas en biometría. Un ejemplo de aplicación sería el uso y generación de retratos hablados de personas en robots de servicio.

Existen rasgos que encajan apropiadamente como biométrico suave tanto en su definición de descriptor secundario como en la de descriptor semántico. Por ejemplo, la ropa de una persona carece de permanencia y de distintividad, y además se puede describir a través de su color o textura. Sin embargo, existen varios casos en donde queda de manifiesto que ambas definiciones no son equivalentes. Por ejemplo, las venas de la mano tienen una distintividad media (Jain, Ross y Prabhakar, 2004), por lo que sería un rasgo biométrico suave desde la perspectiva de descriptor secundario, pero no se puede generar una descripción que sea entendible por los humanos y que cumpla con los criterios establecidos en el párrafo anterior. Otro caso interesante es el rostro humano, el cual se considera biométrico primario, pero también puede entrar en la definición de biométrico suave como descriptor semántico al clasificarlo de acuerdo a su forma.

La primera revista en dedicar un volumen completo al tema de biométricos suaves fue *Pattern Recognition Letters* en 2015. En uno de sus artículos, Nixon *et al.* (2015) definen los biométricos suaves como “la estimación o uso de características personales describibles por humanos que pueden ser usadas para asistir o efectuar reconocimiento de personas”, que corresponde a la perspectiva de descriptor semántico. Sin embargo, varios de

los artículos incluidos en este volumen no son congruentes con esta visión. Por ejemplo, se incluye un artículo sobre identificación a partir de los patrones de los nudillos (Kumar y Wang, 2015) y uno sobre la detección de movimiento del ojo para evitar fraude en sistemas basados en el patrón del iris (Rigas y Komogortsev, 2015), los cuales no pueden ser descritos por humanos. Este volumen de la revista deja en claro que el concepto de biométrico suave todavía no es compartido por toda la comunidad.

Tomando en cuenta las definiciones anteriores en este trabajo de investigación se propone que existen dos dimensiones de análisis que son de interés al tratar de clasificar un rasgo biométrico como duro o suave:

1. La distintividad y permanencia del rasgo.
2. La posibilidad de generar una descripción verbal del rasgo.

En el Cuadro 2.3 se presentan cuatro grupos de rasgos de identificación a partir de estas dimensiones. En la literatura moderna el término biométrico suave se usa en forma indistinta para referirse a cualquier rasgo del grupo I, III o IV. Sin duda alguna es necesario que la comunidad de investigación de esta área proponga una especialización en los términos.

Cuadro 2.3: Dimensiones de análisis de rasgos de identificación que permiten distinguir cuatro grupos importantes: I, II, III y IV. El grupo II corresponde a los biométricos duros y el grupo III a los biométricos suaves. Sin embargo, en la literatura moderna no existe un consenso sobre el término adecuado para los grupos I y IV. Se incluye un ejemplo de cada uno de los grupos.

		¿Se puede describir verbalmente?	
		No	Sí
¿Es distintivo y permanente?	No	I ej. venas del brazo	III ej. ropa
	Sí	II ej. huella dactilar	IV ej. rostro

En este trabajo de investigación se propone la siguiente definición de biométrico suave como descriptor semántico. Un biométrico suave es una propiedad de la persona o de alguna de las entidades que la componen. Estas

entidades pueden ser físicas (como el ojo derecho o la playera) o abstractas (como la personalidad o el andar). El valor de la propiedad es una etiqueta en lenguaje natural. En el Cuadro 2.4 se presentan ejemplos de biométricos suaves siguiendo esta definición.²

Resulta conveniente agrupar en clases algunos de los objetos que constituyen a la persona. En el Cuadro 2.5 se presentan algunas clases de biométricos suaves y ejemplos de instancias, y en el Cuadro 2.6 se muestran algunas de sus propiedades.

2.3. Elementos contextuales

Si se considera el conjunto de circunstancias en las que ocurre la presencia de una persona es posible identificar elementos contextuales que mejoren su identificación. Entre estos factores circundantes podemos listar:

- Contexto temporal.- se puede aprovechar información del tiempo en que se toman las muestras. Por ejemplo, Choi *et al.* (2010) aprovechan las múltiples apariciones del mismo sujeto en fotos tomadas secuencialmente en un intervalo corto de tiempo.
- Contexto espacial.- si se conoce el lugar en el que se tomaron las fotos se puede mejorar la identificación de personas. Por ejemplo, es posible que una misma persona aparezca en múltiples fotos agrupadas de acuerdo al lugar en que se tomaron (Choi *et al.*, 2010). Wang y Yan (2013) aprovechan la coocurrencia entre personas y el lugar mediante el análisis de la apariencia visual.
- Contexto temporal y espacial.- si se conoce el evento en el que se tomó una foto se puede inferir quién aparece (Lin *et al.*, 2010).
- Contexto social.- personas del mismo grupo social (familia, amigos, compañeros de clase, etc.) suelen aparecer juntos (Lee y Grauman, 2011). En imágenes de grupos de personas es más probable que las mujeres aparezcan en el centro y los hombres en los extremos (Gallagher y Chen, 2009a). La relación entre las personas se puede inferir a partir de la coocurrencia en su aparición (Kim, El Saddik y Jung, 2012) y de la posición relativa en fotos (Gallagher y Chen, 2009b). Algunas personas tienen mayor predisposición a aparecer en colecciones de fotos dependiendo de su popularidad (Naaman *et al.*, 2005).

²La definición propuesta hace explícita la diferencia entre entidad y propiedad, una confusión frecuente en algunos artículos relacionados con biométricos suaves.

Cuadro 2.4: Lista de rasgos biométricos suaves. El cuadro refleja la definición de biométrico suave propuesta en este trabajo: un biométrico suave es una propiedad (columna 2) de un componente físico o abstracto de la persona (columna 1) cuyo valor es una etiqueta en lenguaje natural (columna 3).

Objeto	Propiedad	Ejemplos de valores de la propiedad
Persona	edad	niño, joven, adulto, ...
	género	masculino, femenino, ...
	etnicidad	latino, asiático, caucásico, ...
	estado emocional	feliz, triste, sorprendido, ...
Piel	color	blanca, morena, oscura, ...
Cabello	color	rubio, castaño, negro, ...
	tamaño	corto, medio, largo, ...
	estilo	lacio, ondulado, chino, ...
Rostro	forma	triangular, ovalado, cuadrado, ...
Cuerpo	altura	bajo, mediano, alto, ...
	complexión	delgado, normal, robusto, ...
	forma	rectangular, triangular, redondo, ...
	pose	parado, sentado, recostado, ...
Ojo	color	miel, verde, azul, ...
	forma	redondo, almendrado, rasgado, ...
Bigote	estilo	cepillo, húngaro, inglés, ...
	color	negro, café, rojizo, ...
Barba	estilo	completa, candado, tres días, ...
	color	negra, café, rojiza, ...
Andar	postura	erguido, encorvado, ...
	modo	sigiloso, trotando, corriendo, ...
	velocidad	lento, regular, veloz, ...
	dirección	recto, zig-zag, patrullando, ...
Anteojos	grosor de lentes	delgado, grueso, muy grueso ...
	color de lentes	transparente, negro, azul, ...
	forma	ovalados, rectangulares, circulares, ...
	grosor del armazón	delgado, grueso, muy grueso, ...
	color del armazón	negro, gris, rojo, ...
Personalidad	extroversión	sociable, reservado, tímido, ...
	amabilidad	amigable, frío, cooperativo, ...
	responsabilidad	eficiente, organizado, descuidado, ...
	estabilidad	calmado, inseguro, nervioso, ...
	creatividad	intelectual, imaginativo, ...

Cuadro 2.5: Ejemplos de instancias de clases de rasgos biométricos suaves.

Clase	Instancias de la clase
Marca	cicatriz, tatuaje, lunar, pecas, ...
Ropa	playera, camisa, pantalón, falda, sueter, zapato, sombrero, guante, bufanda, chamarra, ...
Accesorios	arete, collar, brazaletes, anillo, ...
Maquillaje	pestaña, párpados, labios, mejillas, ...
Equipaje	mochila, bolso, maleta, ...

- Contexto lógico.- algunas reglas sencillas se pueden aplicar para mejorar la identificación. Por ejemplo, una misma persona no puede aparecer más de una vez en la misma foto (Choi *et al.*, 2010).
- Contexto de almacenamiento.- la organización de fotos en álbumes se puede usar como pista ya que es probable que el dueño del álbum o sus amigos aparezcan en las fotos (Poppe, 2012). Los metadatos asociados con las muestras se pueden usar como pistas (Naaman *et al.*, 2005). En páginas Web el texto cercano a una foto puede contener información de las personas que aparecen (Cour, Sapp y Taskar, 2011). En videos, los subtítulos contienen nombres que pueden corresponder a las personas que aparecen en escena (Everingham, Sivic y Zisserman, 2009).

2.4. Aprendizaje e identificación de personas en robots de servicio

En la sección 1.3 se discutieron los problemas que enfrenta un robot de servicio para aprender e identificar personas, así como la necesidad de contar con un marco de trabajo robusto y flexible para este efecto. Una estrategia común es integrar algoritmos para reconocimiento de personas que se han desarrollado fuera del contexto de robótica de servicio. Sin embargo, esto no toma en cuenta que el reconocimiento en sistemas móviles que trabajan en ambientes dinámicos tiene sus propios retos y peculiaridades.

Los sistemas de seguridad y control de acceso usan siempre rasgos biométricos debido a que trabajan con enormes bases de datos de millones de personas y tienen la necesidad crítica de lograr una precisión casi perfecta. Estos sistemas obtienen información por medio de procesos controlados que

Cuadro 2.6: Lista de clases de rasgos biométricos suaves. Todas las instancias de la misma clase (ver ejemplos en el Cuadro 2.5) comparten las propiedades y el conjunto de etiquetas en lenguaje natural para representar su valor.

Clase	Propiedad	Ejemplos de valores de la propiedad
Marca	tamaño	pequeña, mediana, grande, ...
	ubicación	rostro, cuello, brazo, ...
Ropa	color	blanco, azul, morado, ...
	patrón	liso, rayado, punteado, ...
	material	algodón, mezclilla, satín, ...
	estilo	casual, deportivo, formal, ...
Accesorio	tamaño	pequeño, mediano, grande, ...
	color	plateado, dorado, esmeralda, ...
Maquillaje	cantidad	ligero, regular, fuerte, ...
	color	rosa, azul, verde, ...
	ocasión	trabajo, fiesta, disfraz, ...
Equipaje	tamaño	pequeño, mediano, grande, ...
	color	negro, azul, rojo, ...

requieren la cooperación del usuario; por ejemplo, pedirle que ponga su huella dactilar en un escáner digital, de modo que los datos tengan alta resolución y un alto aislamiento del ruido. Usar este tipo de métodos no es adecuado para robots de servicio debido a que la interacción natural con el humano se rompe si el robot le pide realizar un procedimiento rígido durante el proceso de identificación. Asimismo, los robots de servicio deben tener métodos para resolver situaciones en las cuales el usuario no está encarando directamente a la cámara o está siendo parcialmente ocluido por un objeto, por lo que métodos basados únicamente en rasgos biométricos del rostro no son la mejor opción. El ruido en la información de entrada y sus procesos de recolección no cooperativos son inherentes en las tareas de robots de servicio.

Las redes sociales y sus procesos de etiquetamiento automáticos de personas explotan el poder de los elementos contextuales y cuentan con métodos asistidos por humanos para el aprendizaje y la identificación basados en etiquetas y confirmaciones que proveen los usuarios. La información de entrada son fotos personales que tienen características específicas: personas centradas en primer plano mirando a la cámara e imágenes nítidas. Los robots de servicio no cuentan con toda la infraestructura de metadatos de una red social, de modo que las inferencias contextuales basadas en la red de amigos

de una persona no son posibles.

Los sistemas de videovigilancia tienen baja resolución en su información de entrada, por lo que usar biométricos suaves resulta una opción conveniente. Trabajan con cámaras fijas, por lo que no tienen que lidiar con los problemas de flujo óptico y del fondo no predeterminado al que se enfrentan los sistemas móviles.

Los párrafos anteriores muestran varios puntos en los cuales un robot de servicio tiene desventajas en comparación con otro tipo de aplicaciones. Sin embargo, también es posible encontrar ventajas explotando sus características particulares. En primer lugar, los robots de servicio son sistemas móviles, por lo que se pueden desplazar en el escenario para aproximarse o para encarar a la persona. Muchos robots de servicio también son articulados, por lo que pueden ajustar la posición de su cámara para obtener una mejor vista de la persona, por ejemplo, aprovechando la articulación del cuello para ajustar horizontal o verticalmente la inclinación de su cabeza. También pueden interactuar con el usuario, por ejemplo, pidiendo a una persona que está de espaldas que se gire para ver su rostro.

Los robots de servicio no se restringen a tener un único sensor, de modo que resulta posible combinar diferentes métodos para el mismo propósito. En escenarios específicos, como en un robot dentro de una casa, la interacción se da con un grupo reducido de personas, que es insignificante en comparación con un sistema de seguridad de acceso basado en biométricos que trabaja con millones de usuarios. En algunas tareas, como un robot trabajando de mesero en un restaurante, el robot puede aprender a una persona sólo para propósitos a corto plazo, así que resulta posible memorizarla por medio de rasgos biométricos suaves, como el color de la ropa que lleva puesto.

Ante la gran cantidad de retos explicados en la sección 1.3 y la diversidad de escenarios y aplicaciones en que un robot de servicio debe operar, resulta necesario que pueda explotar el uso de biométricos, biométricos suaves y elementos contextuales. Los rasgos biométricos pueden ser usados cuando se sabe de antemano que existirá una interacción a largo plazo con la persona, por ejemplo, con los habitantes de una casa. En estas situaciones un procedimiento largo y que requiera cooperación del usuario se justifica, como pedirle que se mantenga inmóvil en determinada posición mientras el robot toma fotos de su rostro y las almacena. En ciertas tareas, usar la alta precisión en la identificación de los métodos basados en biométricos debe ser obligatoria, como al verificar la identidad de una persona antes de abrirle la

puerta y así evitar que un intruso acceda al hogar.

Los biométricos suaves pueden ser usados en escenarios no críticos, en los cuales lo más importante sea mantener la naturalidad de la interacción. Los biométricos suaves permiten métodos no invasivos para aprender a una persona a distancia, sin necesidad de su cooperación, e incluso sin que se percate que está siendo memorizada. Aprender al usuario mientras se está teniendo una conversación o alguna actividad conjunta es posible utilizando este tipo de rasgos, con el costo de tener un modelo de la persona útil sólo a corto plazo.

Los elementos contextuales también pueden ser usados por un robot de servicio. Por ejemplo, si el robot sabe que se encuentra en la habitación de Andrés y que es medianoche, y detecta a una persona durmiendo en la cama, es altamente probable que esa persona sea Andrés. El robot también puede explotar el conocimiento que tiene acerca de los eventos previos de la tarea que está realizando. Por ejemplo, si un minuto antes el robot tuvo una interacción con Sara en el comedor y ella le pidió que le trajera un refresco de la cocina después de regresar del refrigerador con la lata, es altamente probable que la persona detectada en el comedor sea Sara esperando su bebida. Una forma de aprovechar la información que brindan los elementos contextuales es usándola como la probabilidad a priori en un esquema de decisión bayesiano en combinación con el análisis de la observación realizada por un sistema basado en biométricos (ya sea duros o suaves).

En los Cuadros 2.7, 2.8, 2.9 y 2.10 se presenta una revisión de las estrategias de detección e identificación utilizadas por los robots de servicio que participaron en la competencia internacional *RoboCup@Home 2016*.³ Este es el evento de robótica de servicio más grande a nivel mundial y las técnicas utilizadas corresponden al estado del arte. Una de las tareas era *Person Recognition*, en la que el robot tenía que aprender a una persona y después buscarla e identificarla entre un grupo de personas en otra zona del escenario (van Beek *et al.*, 2016). El aprendizaje y la identificación también se usaron por algunos equipos para hacer un rastreo más robusto en la tarea de seguir a una persona que camina en el escenario.

³Se realizó el análisis de la competencia de 2016 debido a que en la competencia de 2017 no se evaluaron las habilidades de aprendizaje e identificación de personas.

Cuadro 2.7: Estrategias de aprendizaje e identificación de personas usadas por robots de servicio (Parte 1).

Nombre del equipo	Nombre del robot	Autores	País	Estrategia de detección	Estrategia de identificación
b-it-bots	Jenny	Camargo <i>et al.</i> (2016)	Alemania	Detección de cuerpos a partir de la nube de puntos 3D mediante LSNs ³ seguido de clasificación mediante Bosques Aleatorios.	No especificada.
BahiaRT	BILL	de Souza <i>et al.</i> (2016)	Brasil	Detección de rostros mediante el algoritmo de Viola-Jones. Detección de esqueletos mediante el SDK ⁴ de OpenNI ⁵ para Kinect. ⁶	Identificación de rostros mediante el algoritmo Fisherfaces. ⁷
C.E.S.A.R.-VOXAR	I-Zak 1.0, I-Zack 2.0	Foresti <i>et al.</i> (2016)	Brasil	Detección de rostros mediante OpenCV ⁸ . Detección de cuerpos mediante el paquete ppl.detection ⁹ de ROS. ¹⁰	Identificación de rostros mediante OpenCV.
CIT Brains	Sin nombre	Ueda <i>et al.</i> (2016)	Japón	Detección de esqueletos mediante Kinect.	No especificada.
CPE Lyon Robot Forum	Sin nombre	Leber <i>et al.</i> (2016)	Francia	Detección de esqueletos y rostros mediante Kinect 2.0. ¹¹ Detección de rostros mediante OpenCV.	Identificación de rostros mediante OpenCV.
demura.net	Happy Mini	Demura <i>et al.</i> (2016)	Japón	Detección de rostros mediante el algoritmo de Viola-Jones.	Reconocimiento de rostros (sin precisar si es detección o identificación) mediante clusterización y DNNs. ¹²
eR@sers	HSR	Okada <i>et al.</i> (2016)	Japón	No especificada.	No especificada.

³ Siglas de *Local Surface Normals* (Normales de Superficies Locales). Hegger *et al.* (2012) describen su uso para detección de personas.

⁴ Siglas de *Software Development Kit* (Estuche de Desarrollo de Software).

⁵ Acrónimo de *Open Natural Interaction*. Esta organización desarrolla código abierto para interfaces naturales de usuario. <http://www.openni.ru/>

⁶ También conocido como *Kinect para Xbox 360*. <https://www.xbox.com/>

⁷ Algoritmo propuesto por Belhumeur, Hespánha y Kriegman (1997).

⁸ Acrónimo de *Open Computer Vision*. Provee la implementación del algoritmo de Viola-Jones para detección de rostros y tres algoritmos de identificación de rostros: Eigenfaces, Fisherfaces e Histogramas LBP (*Local Binary Patterns*, Patrones Binarios Locales). <https://docs.opencv.org/>

⁹ <http://wiki.ros.org/ppl.detection>

¹⁰ Siglas de *Robot Operating System* (Sistema Operativo para Robot).

¹¹ También conocido como *Kinect for Xbox One*. <https://www.xbox.com/>

¹² Siglas de *Deep Neural Networks* (Redes Neuronales Profundas).

Cuadro 2.8: Estrategias de aprendizaje e identificación de personas usadas por robots de servicio (Parte 2).

Nombre del equipo	Nombre del robot	Autores	País	Estrategia de detección	Estrategia de identificación
Golem	Golem-III	Pineda <i>et al.</i> (2016)	México	Detección de rostros mediante el algoritmo de Viola-Jones. Detección de cabezas mediante HOGs. ¹³ Detección de esqueletos mediante SDK de Kinect 2.0.	Identificación de rostros mediante el algoritmo Eigenfaces. Identificación de personas a partir de la ropa mediante extracción de parches etiquetados semánticamente de acuerdo a la región del cuerpo (torso, brazos, piernas, etc).
homer@UniKoblenz	Lisa	Memmesheimer <i>et al.</i> (2016)	Alemania	Detección de piernas mediante láser. Detección de rostros mediante OpenCV. Detección de siluetas humanas mediante cámara de profundidad.	Aprendizaje en línea de la apariencia de la persona usando clasificadores AdaBoost en histogramas de color con fines de seguimiento.
KameRider	Sin nombre	Tan <i>et al.</i> (2016)	Japón	Detección de esqueleto mediante Kinect.	No especificada.
Leon Robot	RB-1	Rodríguez Lera, Matellán y Martín Rico (2016)	España	No especificada.	No especificada.
Machinilog	Robo-Hat	Akhter <i>et al.</i> (2016)	Pakistán	Detección de rostros mediante el paquete <code>cob_people_detection</code> ¹⁴ de ROS.	Identificación de rostros mediante el paquete <code>cob_people_detection</code> de ROS.
Markovito	Sabina	Sucar <i>et al.</i> (2016)	México	Detección de ojos, bocas y narices basada en el algoritmo de Viola-Jones. Detección de rostros a partir de un algoritmo basado en histogramas RGB. ¹⁵ Detección de torsos basado en la biometría humana usando como contexto la posición del rostro. Detección de esqueletos mediante Kinect.	Identificación mediante extracción de características SIFT ¹⁶ de las regiones detectadas (ojo, boca y nariz) y uso de un esquema bayesiano para comparar con los modelos en una base de datos. Identificación mediante histogramas de color de la región del torso con fines de seguimiento.

¹³ Siglas de *Histogram of Oriented Gradient* (Histograma de Gradientes Orientados).¹⁴ http://wiki.ros.org/cob_people_detection¹⁵ Siglas de *Red Green Blue* (Rojo Verde Azul).¹⁶ Siglas de *Scale-Invariant Feature Transform* (Transformada de Características Invariantes a la Escala).

Cuadro 2.9: Estrategias de aprendizaje e identificación de personas usadas por robots de servicio (Parte 3).

Nombre del equipo	Nombre del robot	Autores	País	Estrategia de detección	Estrategia de identificación
Pumas	Justina	Biorobotics Laboratory (2016)	México	Detección de rostros usando el SDK VeriLook. ¹⁷	Identificación de rostros usando el SDK VeriLook.
RoboFEI	Judith	Masiero <i>et al.</i> (2016)	Brasil	Detección de rostros mediante OpenCV. Detección de esqueletos mediante el SDK de OpenNI para Kinect.	Identificación de rostros mediante LBPs. Identificación usando el color de la playera con fines de seguimiento.
Sepanta	Sepanta II	Mehrabi <i>et al.</i> (2016)	Irán	Detección de rostro mediante el algoritmo de Viola-Jones. Detección de esqueletos y rostros mediante Kinect 2.0.	Identificación mediante rostro usando el método disperso de Wright <i>et al.</i> (2009) e información del esqueleto (como altura y ancho de hombros). Aprendizaje de la persona combinando OpenTLD ¹⁸ y el esqueleto de la persona con fines de seguimiento.
SKUBA	Sin nombre	Wongsuwan <i>et al.</i> (2016)	Tailandia	No especificada	No especificada
SocRob	Sin nombre	Ventura <i>et al.</i> (2016)	Portugal	Detección de personas usando Kinect mediante un algoritmo que utiliza una SVM, ¹⁹ HOGs y cascadas de Haar.	Aprendizaje de rostros tomando fotos en distintas orientaciones y con diferentes expresiones faciales. Identificación de rostros mediante OpenCV.
Tech United Eindhoven	AMIGO, SERGIO	Lunenburg <i>et al.</i> (2016)	Países Bajos	Detección de rostro (no se especifica la técnica). Detección de personas mediante un algoritmo de emparejamiento de contornos.	Identificación de rostro (no se especifica la técnica).
Tinker	Tinker	Guo <i>et al.</i> (2016)	China	Detección de rostros con MCS. ²⁰	Aprendizaje de la persona con TLD con fines de seguimiento.

¹⁷ <http://www.neurotechnology.com/vl.sdk.html>

¹⁸ Implementación del algoritmo TLD (*Tracking-Learning-Detection*, Rastreo-Aprendizaje-Detección) de Kalal, Mikolajczyk y Matas (2012). <https://github.com/zk00006/OpenTLD>

¹⁹ Siglas de *Support Vector Machine* (Máquina de Soporte Vectorial).

²⁰ Siglas de *Microsoft Cognitive Services* (Servicios Cognitivos de Microsoft).

<https://azure.microsoft.com/en-us/services/cognitive-services/>

Cuadro 2.10: Estrategias de aprendizaje e identificación de personas usadas por robots de servicio (Parte 4).

Nombre del equipo	Nombre del robot	Autores	País	Estrategia de detección	Estrategia de identificación
ToBI	ToBI	Meyer zu Borgsen, Korthals y Wachsmuth (2016)	Alemania	Detección de personas usando <code>strands_perception_people</code> . ²¹	Identificación de personas usando <code>strands_perception_people</code> .
Walking Machine	SARA	Caron <i>et al.</i> (2016)	Canadá	Detección de rostros mediante el paquete <code>face_detector</code> ²² de ROS. Detección de esqueletos (no se especifica la técnica).	Identificación de rostros mediante el paquete <code>cob_people_detection</code> de ROS.
Wright Eagle	KeJia	Shuai <i>et al.</i> (2016)	China	Detección de rostros mediante el algoritmo de Viola-Jones. Detección de personas mediante un algoritmo basado en HODs. ²³	Identificación de rostros usando el SDK VeriLook.

²¹ Incluye detección de piernas y de parte superior del cuerpo con fines de rastreo.

https://github.com/strands-project/strands_perception_people

²² http://wiki.ros.org/face_detector

²³ Siglas de *Histogram of Oriented Depths* (Histograma de Profundidades Orientadas).

Aunque existe una gran variedad de técnicas usadas por los equipos, es claro que la estrategia principal de identificación es mediante el rostro (biométrico duro), y sólo algunos equipos aprenden la apariencia del cuerpo (biométrico suave) para usarlo en la tarea de seguimiento.

2.5. Recapitulación

En este capítulo se revisaron tres grupos de elementos que se pueden utilizar para la identificación de personas: los rasgos biométricos duros, los rasgos biométricos suaves y los elementos contextuales. Los rasgos biométricos duros son características fisiológicas y/o de comportamiento que permiten distinguir a una persona del resto. Jain, Ross y Prabhakar (2004) proponen siete factores para evaluar la utilidad de un rasgo en aplicaciones, entre ellos la distintividad y la permanencia.

El concepto de rasgo biométrico suave surgió de manera más reciente y se utiliza para referirse a dos cosas diferentes. Por un lado, algunos autores

utilizan el concepto para referirse a un descriptor secundario o auxiliar, que carece ya sea de distintividad o de permanencia, y que tiene un papel complementario en el proceso de identificación de personas. Por otra parte, otros autores utilizan el término para referirse a un descriptor semántico, de modo que el rasgo se puede describir con una expresión en lenguaje natural. Ante la discrepancia en el uso del término por parte de la comunidad, en este trabajo se propuso en el Cuadro 2.3 una forma de clasificar a los rasgos de identificación en cuatro categorías a partir de dos dimensiones: i) su distintividad y permanencia, ii) la posibilidad de describirlos verbalmente.

Entre los trabajos que utilizan el concepto de biométrico suave como descriptor semántico se presenta una confusión frecuente entre entidades y propiedades. Por ejemplo, en la lista de rasgos biométricos suaves de Dantcheva *et al.* (2010:742) se encuentran los lentes (una entidad) y el peso (una propiedad). En este capítulo se propuso una definición de biométrico suave como descriptor semántico que distingue de manera clara las entidades, las clases, las propiedades y los valores, y que sirvió para elaborar la lista de biométricos suaves de los Cuadros 2.4, 2.5 y 2.6. De igual forma, se propusieron cuatro criterios para restringir el tipo de descripciones verbales necesarias para considerar a un rasgo como biométrico suave.

También se revisaron los elementos contextuales y se presentaron ejemplos de investigaciones en donde estos elementos circundantes a la persona tuvieron relevancia en su identificación. Se propuso una clasificación de ellos en seis tipos de contexto: temporal, espacial, temporal-espacial, social, lógico y de almacenamiento.

Otro punto importante del capítulo fue el análisis de los problemas de aprendizaje e identificación de personas en el contexto de robots de servicio. En la sección 2.4 se discutió de manera amplia la forma en que cada uno de los tres grupos de elementos de identificación se pueden aprovechar en las tareas que realiza un robot de servicio. Además, se realizó una revisión de las estrategias de aprendizaje e identificación de los robots que participaron en la competencia *RoboCup@Home 2016* y se llegó a la conclusión de que la principal técnica es el uso del rostro para identificar personas (es decir, un biométrico duro). Se encontró que algunos equipos usaron el aprendizaje de la apariencia del cuerpo (como el color de la playera, que corresponde a un biométrico suave) para hacer más robusto el proceso de rastreo en la tarea de seguir a la persona en el escenario. En el siguiente capítulo se explorarán distintos algoritmos y estrategias para trabajar con biométricos suaves, algunos de los cuales se pueden utilizar en robótica de servicio.

Capítulo 3

Algoritmos de análisis para biométricos suaves

En el capítulo anterior se presentaron distintos elementos para aprender e identificar a una persona y se discutió la manera en que se pueden utilizar dentro de tareas de robots de servicio. Además, se presentó una revisión bibliográfica sobre las estrategias utilizadas por distintos robots y se llegó a la conclusión de que la mayoría de ellos utilizan biométricos duros y sólo algunos utilizan biométricos suaves. Este resultado motiva a explorar de manera profunda la incorporación de biométricos suaves dentro de robots de servicio.

En este capítulo se hace énfasis en los biométricos suaves y se presenta una revisión bibliográfica extensa sobre ellos guiada por la forma en que estos rasgos se pueden detectar, localizar, aprender e identificar. Este análisis resulta de importancia para poner sobre la mesa los distintos algoritmos y estrategias existentes. Resulta importante señalar que casi todos los algoritmos y estrategias presentados en esta sección son resultado de trabajos de investigación a partir de colecciones de imágenes, por lo que no todos ellos se pueden integrar dentro de un robot de servicio debido a los retos que se presentaron en la sección 1.3.

3.1. Detección y localización de biométricos suaves

La detección es el primer paso en el análisis de un rasgo biométrico y se puede plantear como un problema de clasificación: dada una imagen decidir si un objeto en ella corresponde o no a una parte de la persona (por ejemplo,

la cara, el cabello, la piel, etc.). Después de ser detectado, un rasgo se localiza delimitando su posición en la escena (por ejemplo, obtener la coordenada 3D del centro de masa, determinar el conjunto de píxeles que ocupa la proyección del objeto en una imagen 2D, etc.).

La detección y localización son subproblemas de reconocimiento que deben ser resueltos de manera obligatoria antes de proceder con los demás subproblemas. Si esta etapa falla no es posible trabajar con el aprendizaje, la identificación, el rastreo o la descripción del elemento. Por esta razón es que la mayoría de investigación en biométricos suaves se enfoca en esta etapa inicial del reconocimiento.

La primera cuestión que surge es si es posible detectar y localizar un elemento en particular. De acuerdo a esto los distintos rasgos biométricos suaves presentados en el Capítulo 2 se pueden clasificar de la siguiente manera:

- Rasgos localizados.- son partes o regiones concretas del cuerpo humano, ya sea naturales o adheridas. Por ejemplo, la cara, la piel, los lentes, la playera, etc.
- Rasgos de comportamiento.- aparecen en forma esporádica. Por ejemplo, el modo de caminar o el estado emocional.
- Rasgos emergentes.- algunos rasgos como la edad, el género o la etnicidad, no se pueden detectar en una parte específica del cuerpo humano sino que emergen del análisis de un conjunto de características de la persona. En este contexto la pregunta “¿Dónde está la edad?” no tiene respuesta, pero “¿Cuál es la edad?” se puede responder por la combinación de otros rasgos (por ejemplo, tener cabello blanco y caminar encorvado puede llevar a la conclusión de que se trata de una persona de edad avanzada).

En esta sección se abordarán los métodos para trabajar con rasgos localizados. Después de ser detectados, existen diferentes formas de indicar su localización en la escena: como objetos, como materiales o como objetos articulados. Forsyth *et al.* (1996) enfatizan la diferencia entre objetos y materiales. Un objeto tiene tamaño y forma específica (por ejemplo, un anillo). Por otra parte, un material carece de forma o extensión espacial distintiva o específica, y se define por un patrón homogéneo o repetitivo (por ejemplo, la piel). En la Figura 3.1 la misma imagen se analiza bajo ambas perspectivas. Cuando los rasgos son localizados como objetos se delimitan



Figura 3.1: Tres formas diferentes de localizar un rasgo: como objeto (izquierda), como material (centro) o como objeto articulado (derecha)

por un rectángulo, pero cuando son analizados como materiales la imagen es segmentada en regiones.

Algunos objetos están compuestos por otros objetos. Un objeto articulado es un sistema formado por al menos dos componentes rígidos y a lo mucho seis grados de libertad independientes entre cualquier par de sus componentes (Comport, Marchand y Chaumette, 2005). Este tipo de objetos se pueden representar por un grafo, en el cual los vértices corresponden a las articulaciones y extremos del objeto, mientras que las aristas corresponden a sus componentes rígidos como se puede ver en el Figura 3.1.

Existen métodos específicos para detectar y localizar algunos rasgos. El rostro ha sido investigado en forma extensa usando diferentes técnicas y algoritmos. Yang, Kriegman y Ahuja (2002) clasifican los métodos de detección de rostro en cuatro grupos: i) Basados en conocimiento (codifican el conocimiento humano acerca de rostros en reglas), ii) Características invariantes (encontrar características estructurales invariantes que permanezcan estables incluso con variaciones de pose, puntos de vista o iluminación), iii) Emparejamiento de patrones (encontrar correlación entre la imagen de entrada y una base de datos de patrones estándar), iv) Basados en apariencia (los patrones del rostro se obtienen por técnicas de aprendizaje de máquinas). Zhang y Zhang (2010) presentan una revisión sobre el tema y concluyen que la mayoría de los detectores de rostro modernos utilizan métodos basados en apariencia. Zafeiriou, Zhang y Zhang

(2015) categorizan las técnicas modernas de detección de rostro *in-the-wild*¹ en dos grupos: i) Plantillas rígidas (que se aprenden por métodos como redes neuronales profundas o métodos de boosting), ii) Modelos deformables (que describen al rostro mediante sus partes). Prihasto *et al.* (2017) presentan una revisión de las técnicas de detección de rostro basadas en aprendizaje profundo, las cuales tienen gran auge en la actualidad.

La detección de piel también está presente en muchas investigaciones. Las metodologías aplicables para la detección de piel en imágenes 2D están basadas en la información de color en el espectro visual. La detección de piel se puede ver como un problema de clasificación binaria a nivel de pixel, determinando si corresponde a piel o no. Se han propuesto distintos métodos usando difentes espacios de colores, clasificadores de pixeles y enfoques dinámicos adaptativos (Kakumanu, Makrogiannis y Bourbakis, 2007). Entre las técnicas más recientes de detección de piel se encuentra el uso de aprendizaje profundo, que ha mostrado mejores resultados que los métodos basados en reglas, los modelos Gaussianos y las redes neuronales prealimentadas (Hajiarbabi y Agah, 2015).

La pose del cuerpo humano es modelada generalmente como un objeto articulado y su estudio también está respaldado por una gran cantidad de investigación. La pose del cuerpo es la estructura cinemática subyacente de una persona y requiere de la detección del cuerpo completo y sus partes (Perez-Sala, Escalera y Angulo, 2012). La detección de partes del cuerpo como el torso, la zona de cabeza-hombros, el brazo, la pierna, etc., se puede realizar mediante métodos basados en HOGs, los cuales localizan los elementos como objetos (Corvee y Bremond, 2010). Otro enfoque es el uso de *poselets*, que son detectores de partes del cuerpo que codifican tanto la configuración espacial 3D parametrizada por las coordenadas de las articulaciones, como la apariencia espacial 2D de la imagen (Bourdev y Malik, 2009). El uso de aprendizaje profundo también ha conseguido buenos resultados en detección de pose (Toshev y Szegedy, 2014). Otro objeto articulado con gran cantidad de investigación es la mano (Zaman y Adnan, 2012), incluyendo el uso de aprendizaje profundo para su detección (Chen *et al.*, 2016).

La detección del caminar ha recibido gran interés en investigaciones por sus aplicaciones en sistemas de videovigilancia. En el proceso de caminar el cuerpo humano muestra un movimiento periódico regular (Zhang, Hu y

¹Este término se usa para referirse a condiciones de captura no restringidas.

Wang, 2011). A diferencia de los rasgos presentados en los párrafos anteriores que se pueden detectar a partir de una imagen, la detección de este rasgo de comportamiento implica un análisis sobre una secuencia de imágenes ordenadas en forma temporal. El uso de aprendizaje profundo ha logrado buenos resultados en la detección del caminar en tiempo real (Angelova *et al.*, 2015).

En años recientes el aprendizaje profundo se ha convertido en una de las técnicas con mejores resultados y ha permitido la detección de prendas de vestir (Liu *et al.*, 2016), cabello y barba (Qin, Kim y Manduchi, 2017), ojos (Karahan y Akgül, 2016), entre otros.

Una estrategia común en la detección de partes del cuerpo se basa en predecir su posición en la imagen a partir del uso de otro rasgo como pista contextual. La idea básica de usar el contexto semántico para entender una imagen es que las etiquetas de los objetos se deben asignar con respecto a otros objetos en la escena. Por ejemplo, en una imagen que contiene una raqueta es más probable que una bola verde corresponda a una pelota de tenis que a un limón (Rabinovich *et al.*, 2007). Si una imagen es segmentada semánticamente mediante la agrupación de píxeles de acuerdo a su significado semántico (Johnson, 2008) entonces los materiales también se pueden utilizar como parte de una estrategia de contexto semántico. Heitz y Koller (2008) dividen el uso de contexto en cuatro categorías: i) Contexto escena-objeto (permite que la información a nivel de la escena se use para determinar la ubicación de los objetos), ii) Contexto material-material (captura la relación espacial entre los materiales), iii) Contexto objeto-objeto (considera la coocurrencia de los objetos) y iv) Contexto material-objeto (permite que las regiones de textura agreguen poder predictivo a la detección de objetos).

En el análisis de imágenes no basta con listar los elementos que aparecen, sino que resulta necesario obtener la composición semántica, es decir, dónde están los objetos y los materiales y cómo están relacionados. Biederman, Mezzanotte y Rabinowitz (1982) identifican cinco clases de relaciones entre un objeto y el escenario: i) interposición (los objetos interrumpen el fondo), ii) soporte (los objetos se encuentran sobre superficies), iii) probabilidad (un objeto en particular tiende a aparecer en algunas escenas y no en otras), iv) posición (dada una escena, un objeto suele aparecer en determinadas posiciones) y v) tamaño (la proporción del tamaño de un objeto con respecto a otros se mantiene constante).

Estas ideas también se pueden utilizar en reconocimiento de personas,

sacando ventaja de la relación entre las diferentes partes del cuerpo humano. En general, el procedimiento para detectar un rasgo usando pistas contextuales empieza por la detección de un elemento del cuerpo a partir de un clasificador de alto rendimiento (por ejemplo, uno de rostros o uno de piel) cuya ubicación se utiliza como contexto inicial. A continuación se genera una hipótesis sobre la localización del rasgo que se desea detectar utilizando el conocimiento acerca de la organización espacial entre los elementos del cuerpo humano mediante un método específico (por ejemplo, un sistema basado en reglas, un modelo probabilístico, etc.). Después se realiza un procedimiento para verificar si la región seleccionada contiene el elemento que se está buscando. Finalmente, dependiendo de si el rasgo se desea utilizar como objeto o como material, se establece un recuadro delimitador o una región segmentada. El nuevo rasgo se puede utilizar como contexto para buscar otros rasgos.

Por ejemplo, Cushen y Nixon (2012) proponen un detector de playeras basado en el siguiente procedimiento. Primero se realiza la detección del rostro mediante el algoritmo de Viola-Jones. Después, la región superior del cuerpo se detecta mediante una regla que determina un recuadro delimitador proporcional al tamaño del rostro y alineado respecto a este último. Usando otra regla, se determina la región de piel del rostro y se calcula la media del color de los píxeles. A continuación se detectan ventanas con el mismo color en la parte superior del cuerpo y se elige aquella que corresponde al cuello. Finalmente, la región de la playera se segmenta como material descartando la región de piel en el cuello. En este ejemplo se utilizan reglas basadas en suposiciones espaciales acerca del cuerpo humano que funcionan sólo cuando se satisfacen una serie de presuposiciones: el procedimiento tendrá éxito sólo si la persona se encuentra de pie y encarando a la cámara, pero fallará si se encuentra recostado en el piso o parado de manos.

Para simplificar el análisis de otros trabajos que realizan detección basada en contexto se propone la siguiente notación. Sea $X = \{x_1, x_2, \dots, x_n\}$ un conjunto de rasgos de interés y sea X_i un subconjunto no vacío de X con subíndice $i \in \mathbb{N}$. Se define $X_i \rightarrow X_{i+1}$ como la detección de los rasgos X_{i+1} usando a X_i como contexto. Se asume que si un rasgo se ha detectado su información estará disponible en el resto del análisis, por lo que en un proceso de detección de n etapas $X_1 \rightarrow X_2 \rightarrow \dots \rightarrow X_i \rightarrow X_{i+1} \rightarrow \dots \rightarrow X_n$ la detección de los rasgos X_{i+1} usó $X_1 \cup X_2 \cup \dots \cup X_i$ como contexto. En la presente notación se indica si el rasgo se localizó como objeto, material u objeto articulado con el superíndice o , m o a respectivamente. Por ejemplo, el proceso de detección en etapas de

Cushen y Nixon (2012) es: $\{\text{rostro}^o\} \rightarrow \{\text{cuerpo_superior}^o, \text{piel_facial}^o\} \rightarrow \{\text{piel_en_cuerpo_superior}^o\} \rightarrow \{\text{cuello}^o\} \rightarrow \{\text{ropa_en_cuerpo_superior}^m\}$.

El planteamiento de la detección basada en contexto en fases del párrafo anterior tiene similaridad con el método de cadenas de clasificadores (Read *et al.*, 2009). Este método transforma un problema de clasificación multietiqueta² en un conjunto de problemas de clasificación binarios, uno por cada etiqueta, formando una cadena de clasificadores C_1, \dots, C_n , donde n es el número de etiquetas del problema original. Cada clasificador C_i de la cadena se encarga de aprender y predecir la etiqueta l_i utilizando el espacio de características x_i de ese clasificador junto con el resultado de todos los clasificadores que le preceden en la cadena, es decir, determinar $P(l_i|x_i, l_1, \dots, l_{i-1})$. El orden de las etiquetas en el método de cadenas de clasificadores (por ejemplo: $l_1 = \text{piel}, l_2 = \text{rostro}, l_3 = \text{ojos}$) correspondería a las fases del método de detección basado en contexto en fases (por ejemplo: $\{\text{piel}^m\} \rightarrow \{\text{rostro}^o\} \rightarrow \{\text{ojos}^o\}$).

La diferencia entre las dos propuestas de notación se manifiesta cuando se trata de representar una fase de análisis en donde dos o más clasificadores trabajan de manera paralela. Por ejemplo, en el trabajo de Cushen y Nixon (2012) las tres primeras fases son las siguientes: $\{\text{rostro}^o\} \rightarrow \{\text{cuerpo_superior}^o, \text{piel_facial}^o\} \rightarrow \{\text{piel_en_cuerpo_superior}^o\}$. Tanto el clasificador de cuerpo superior como el de piel facial requieren usar como contexto al rostro y sirven como contexto para la siguiente fase. Sin embargo, en el trabajo de los autores no existe una dependencia entre ellos. Si se trata de representar en una cadena de clasificadores siguiendo la notación de Read (2009) y suponemos que $l_1 = \text{rostro}, l_{2a} = \text{cuerpo_superior}, l_{2b} = \text{piel_facial}$ y $l_3 = \text{piel_en_cuerpo_superior}$, se tendría que elegir entre el encadenamiento C_1, C_{2a}, C_{2b}, C_3 o el C_1, C_{2b}, C_{2a}, C_3 , y ninguno de ellos coincide con la manera en que los autores están realizando el procedimiento.

Usar el contexto para la detección de objetos o materiales requiere de un modelo con conocimiento del cuerpo humano y su composición. Algunos sistemas se basan en reglas que son definidas por expertos o aprendidas automáticamente a partir de una base de datos. Las reglas espaciales codifican el conocimiento a priori del cuerpo humano y tienen como objetivo predecir la posición y tamaño de otros rasgos a partir del contexto. Por ejemplo, la búsqueda de la posición de los ojos debe estar restringida a la

²En la clasificación multietiqueta cada una de las instancias x se asocia con un subconjunto S del conjunto de etiquetas L .

parte superior del rostro (Asteriadis *et al.*, 2006). Las reglas morfológicas también pueden ayudar a la predicción; por ejemplo, un ojo humano típico consiste de una esclerótica blanca alrededor de una pupila oscura por lo que en este contexto una transición de blanco a oscuro o viceversa se puede asociar con la presencia de un ojo (Turkan, Pardas y Cetin, 2008). Las reglas morfológicas se usan de manera implícita por los métodos basados en características.

Otra alternativa es dar al sistema un patrón estándar del rasgo y proceder al análisis mediante técnicas de comparación de plantillas. Otra forma es aprender el modelo estadístico a partir de un conjunto de datos (por ejemplo, modelos basados en PCA³ o modelos bayesianos). También se pueden utilizar PGMs⁴ tales como Redes de Markov, MRFs⁵ o CRFs.⁶ Algunos clasificadores como SVMs, Bosques Aleatorios, clasificadores dispersos y métodos profundos como R-CNNs⁷ se han utilizado en algunas investigaciones para detectar rasgos. Otra estrategia es el uso de modelos composicionales que dividen al objeto de estudio (la cara o el cuerpo) en pequeños parches que posteriormente se reensamblan usando como guía un grafo que contiene la estructura semántica a priori de los parches. También se ha empleado el método de aprender una máscara probabilística personalizada a partir de varias imágenes de la misma persona o de sus partes usando como guía una máscara o modelo general. Algunos modelos son dinámicos, como los ASMs⁸ que se deforman iterativamente hasta ajustarse a un caso particular, o los AMMs⁹ que consideran tanto la forma como la textura.

Debido a la dificultad en la detección y localización de rasgos los trabajos de investigación utilizan una combinación de los métodos mencionados anteriormente. Por ejemplo, un método se utiliza para proponer candidatos y otro para elegir al mejor. En algunos casos, después de seleccionar uno de los candidatos y localizarlo como objeto, se aplica otro método para localizarlo como material mediante una segmentación fina de la región.

En los Cuadros 3.1 y 3.2 se presenta una revisión de diferentes trabajos

³Siglas de *Principal Component Analysis* (Análisis de Componentes Principales).

⁴Siglas de *Probabilistic Graphical Model* (Modelo Gráfico Probabilístico).

⁵Siglas de *Markov Random Field* (Campo Aleatorio de Markov).

⁶Siglas de *Conditional Random Field* (Campo Aleatorio Condicional).

⁷Siglas de *Regions with Convolutional Neural Network features* (Regiones con características de Redes Neuronales Convolucionales).

⁸Siglas de *Active Shape Model* (Modelo Activo de Forma).

⁹Siglas de *Active Appearance Model* (Modelo Activo de Apariencia).

de investigación que realizan la detección de rasgos biométricos suaves basándose en el contexto. Los trabajos se clasifican en categorías de acuerdo al número de etapas en el proceso de detección y la forma en que los rasgos se localizan en cada etapa (como objeto o como material). El Cuadro 3.1 contiene detecciones en dos etapas, mientras que el Cuadro 3.2 presenta trabajos con tres o más etapas. Una nota importante es que algunos de estos trabajos utilizan como entrada fotos de rostro, por lo que no se ven en la necesidad de aplicar un algoritmo para detectarlo, sino que asumen que ya se encuentra presente. Los tratamientos especiales que se dan a la imagen (como filtros o transformaciones) no se cubren en esta revisión. La mayoría de los trabajos de investigación se apoyan en reglas espaciales para restringir el espacio de búsqueda, pero sólo se mencionarán en la columna de método de clasificación cuando sean la estrategia principal del trabajo de investigación. Por inspección de la tabla se concluye que el rostro es el elemento que más se usa como elemento contextual en la detección de otras partes del cuerpo.

En este trabajo se utilizará la pose de la persona en 3D como contexto inicial para determinar la posición de la ropa y extraer parches en esas zonas, que estarán semánticamente etiquetados de acuerdo al lugar al que pertenecen (por ejemplo, el centro del torso, la nuca, la mano, etc.). El esquema de trabajo será del tipo $\{pose_corporal^a\} \rightarrow \{rostro^o, ropa^o, piel^o, cabello^o\}$. Ya que el sistema se usará en un robot de servicio se evitó localizar a los rasgos biométricos como materiales debido a que la segmentación fina implicaría un costo excesivo de recursos computacionales, lo que tendría repercusión en el tiempo de respuesta. Por ello se decidió trabajar con múltiples parches correspondientes a regiones de ropa, piel y cabello que se localizan como objetos. La descripción de esta implementación se detalla en el Capítulo 4.

3.2. Aprendizaje e identificación de biométricos suaves

Después de que un rasgo se detecta y localiza es posible extraer características para obtener una representación computacional que pueda ser almacenada y utilizada para diversos propósitos, incluyendo los procesos de aprendizaje, identificación y descripción del rasgo. En esta sección la atención se centra en las propiedades de los rasgos y en los métodos y algoritmos específicos para obtener su representación computacional.

Cuadro 3.1: Revisión de detección basada en contexto de rasgos biométricos suaves en dos etapas.

Categoría	Etapas	Método de clasificación	Autores
<i>Objeto</i> → <i>Objeto</i>	{ <i>rostro</i> ^o } → { <i>ojos</i> ^o }	PCA, comparación de plantillas	Asteriadis <i>et al.</i> (2006)
		Reglas espaciales y morfológicas	Shafi y Chung (2009)
		Modelo basado en características, comparación de plantillas	Peng y Chen (2005)
	{ <i>rostro</i> ^o } → { <i>playera</i> ^o }	Reglas espaciales y morfológicas	Anguelov <i>et al.</i> (2007) Everingham, Sivic y Zisserman (2009)
	{ <i>rostro</i> ^o } → { <i>cabello</i> ^o , <i>playera</i> ^o }	Aprendizaje de máscara	Sivic, Zitnick y Szeliski (2006)
<i>Objeto</i> → <i>Material</i>	{ <i>rostro</i> ^o , <i>ojos</i> ^o } → { <i>cabello</i> ^m }	PCA, ASM	Julian <i>et al.</i> (2010)
	{ <i>rostro</i> ^o } → { <i>rostro</i> ^m , <i>cabello</i> ^s }	Aprendizaje de máscara, MRF	Lee <i>et al.</i> (2008)
	{ <i>rostro</i> ^o } → { <i>ropa_superior</i> ^m , <i>piel</i> ^m }	Reglas espaciales y morfológicas	Hu, Yan y Lin (2008)
	{ <i>rostro</i> ^o } → { <i>ropa_superior</i> ^m }	Bosques aleatorios, modelo bayesiano, redes de Markov	Wang y Ai (2011)
	{ <i>multiples_partes_del_cuerpo</i> ^o } → { <i>ropa_superior</i> ^m , <i>ropa_inferior</i> ^m }	Aprendizaje de máscara	Weber, Bäuml y Stiefelhagen (2011)
	{ <i>rostro</i> ^o } → { <i>lentes</i> ^m }	PCA	Nguyen <i>et al.</i> (2008)
	{ <i>rostro</i> ^o } → { <i>barba</i> ^m }	PCA	Nguyen <i>et al.</i> (2008)
	{ <i>rostro</i> ^o } → { <i>barba</i> ^m , <i>bigote</i> ^m }	ASM, clasificador disperso	Le <i>et al.</i> (2012)
<i>Objeto articulado</i> → <i>Objeto</i>	{ <i>pose_corporal</i> ^a } → { <i>ropa</i> ^o }	R-CNN, SVM	Hara, Jagadeesh y Piramuthu (2016)
<i>Objeto articulado</i> → <i>Material</i>	{ <i>pose_corporal</i> ^a } → { <i>ropa</i> ^m }	AAM, MRF	Hasan y Hogg (2010)
	{ <i>pose_corporal</i> ^a } → { <i>ropa</i> ^m , <i>accessorios</i> ^m }	SVM, CRF	Chen, Gallagher y Girod (2012)
	{ <i>pose_corporal</i> ^a } → { <i>piel</i> ^m , <i>cabello</i> ^m , <i>ropa</i> ^m , <i>accessorios</i> ^m , <i>equipaje</i> ^m }	CRF	Yamaguchi <i>et al.</i> (2012)
<i>Material</i> → <i>Objeto</i>	{ <i>piel</i> ^m } → { <i>ojos</i> ^o }	Reglas espaciales y morfológicas	Sidibe, Montesinos y Janaqi (2006)
	{ <i>piel</i> ^m , <i>cabello</i> ^m } → { <i>rostro</i> ^o }	Modelo basado en características	Ghofrani, Neshat y Aghaeinia (2012)
<i>Material</i> → <i>Material</i>	{ <i>rostro</i> ^m } → { <i>barba</i> ^m }	Reglas espaciales y morfológicas	Dahri y Belaid (2013)
	{ <i>piel</i> ^m , <i>ojos</i> ^m , <i>boca</i> ^m } → { <i>cabello</i> ^m }	Reglas espaciales y morfológicas	Lipowesky, Mamo y Cohen (2008)

Cuadro 3.2: Revisión de detección basada en contexto de rasgos biométricos suaves en tres o más etapas.

Categoría	Etapas	Método de clasificación	Autores
Objeto → Objeto → Objeto	{rostro ^o } → {piel ^o } → {ropa_superior ^o }	Reglas espaciales y morfológicas	Song y Leung (2006)
Objeto → Objeto → Material	{rostro ^o , ojos ^o } → {cuerpo ^o } → {ropa ^m }	Aprendizaje de máscara	Gallagher y Chen (2008)
	{rostro ^o } → {cuerpo ^o } → {piel ^m , ropa ^m }	SVM	Yang y Yu (2011a)
	{rostro ^o } → {ojos ^o , boca ^o , nariz ^o } → {barba ^m }	Modelo basado en características	Wang y Yau (2013)
	{rostro ^o , ojos ^o } → {parches_contorno_facial ^o } → {rostro ^m , cabello ^m , playera ^m }	Modelo composicional, CRF	Wang, Ai y Lao (2010)
Objeto → Material → Material	{rostro ^o } → {piel ^m , ojos ^m , cejas ^m , nariz ^m , boca ^m , cabello ^m } → {verrugas ^m , cicatrices ^m , tatuajes ^m }	SVM	Heflin, Scheirer y Boulton (2012)
Material → Objeto → Objeto	{piel ^m } → {rostro ^o } → {ojos ^o }	Modelo basado en características, SVM	Turkan, Pardas y Cetin (2008)
Material → Material → Material	{cuerpo ^m } → {piel ^m } → {ropa ^m }	Reglas espaciales y morfológicas	Yao, Khan y Farbiz (2011)
	{piel ^m } → {cabello ^m } → {verrugas ^m }	SVM	Cho, Freeman y Tsao (2007)
	{rostro ^m } → {ojos ^m , nariz ^m , boca ^m } → {marcas_faciales ^m }	Modelo basado en características	Jain y Park (2009)
Objeto → Objeto → Objeto → Material	{rostro ^o , ojos ^o } → {regiones_piel_facial ^o } → {regiones_cabello ^o } → {cabello ^m }	Reglas espaciales y morfológicas	Yacoob y Davis (2006)
Objeto → Objeto → Objeto → Material	{rostro ^o } → {cuerpo_superior ^o , piel_facial ^o } → {piel_en_cuerpo_superior ^o } → {cuello ^o } → {ropa_en_cuerpo_superior ^m }	Reglas espaciales y morfológicas	Cushen y Nixon (2012)

La literatura de Visión por Computadora provee una gran cantidad de métodos para describir el color, la textura o la forma de regiones en una imagen. El color es una propiedad de interés para rasgos como la ropa, el cabello, la piel, etc. Para rasgos biométricos suaves se han utilizado distintos espacios de color, incluyendo RGB, HSV, LAB, HCL, CIE L*a*b*, LCC y YCbCr. Dependiendo del espacio de color cada pixel se representa por una 3-tupla o una 4-tupla de valores discretos, donde cada valor corresponde a una componente del modelo. El color de una región se puede pensar como el conjunto de todos los valores de los pixeles, pero en muchas ocasiones se requiere una representación más compacta para mejorar el desempeño en términos de tiempo de procesamiento y memoria. Existen varias opciones para tener una representación más ligera del color de una región: el valor medio, un intervalo de posibles valores, el color dominante, un histograma, una representación clusterizada de los valores de color, los parámetros de un modelo de distribución de Gauss, etc.

La textura se puede usar para identificar patrones en la ropa, el tipo de cabello, la presencia de barba, etc. Entre los métodos para caracterizar la textura se encuentran los filtros de detección de bordes, los filtros de Gabor, filtros de respuesta máxima y transformadas de Fourier (incluyendo ondículas, ondículas de Gabor y DCTs¹⁰). Otras opciones son el uso de características locales (como LBP y descriptores SIFT) y descriptores estadísticos (como varianza, entropía, suavizado y homogeneidad).

Los modelos BoF¹¹ son una opción para representar color y/o textura. Consisten en extraer un conjunto representativo de parches independientes de la imagen, cada uno de los cuales se evalúa mediante un descriptor visual. La distribución resultante se usa para caracterizar la imagen (Nowak, Jurie and Triggs, 2006).

La descripción de la forma es importante para rasgos como el rostro, los ojos, el cuerpo, el cabello, la ropa, etc. A pesar de que la Visión por Computadora provee muchas técnicas para describir la forma de objetos hasta el momento no se han utilizado en el área de biométricos suaves. Otra propiedad que no se ha trabajado es el albedo o coeficiente de reflexión, que permite clasificar el material del que está hecha la ropa. El Cuadro 3.3 presenta un resumen de las distintas representaciones utilizadas para las propiedades de biométricos suaves.

¹⁰Siglas de *Discrete Cosine Transforme* (Transformada de Coseno Discreta).

¹¹Siglas de *Bag of Features* (Bolsa de Características).

La mayoría de los trabajos de investigación del cuadro 3.3 utilizan la representación computacional de las propiedades para lograr una localización más fina del rasgo. Sólo algunos de los trabajos las utilizan para fines de identificación. Sivic, Zitnick y Szeliski (2006) encuentran todas las apariciones de una persona en una secuencia de fotografías tomadas en un periodo corto de tiempo a partir de un algoritmo que identifica ropa y cabello. Yacoob y Davis (2006) combinan atributos de cabello y PCA en un sistema de identificación de personas. Anguelov *et al.* (2007) presentan un sistema para identificación de personas en álbumes de fotos basado en MRF que utiliza el rostro, la ropa y la restricción lógica de que la persona sólo puede aparecer una vez en la imagen. Los experimentos de Gallagher y Chen (2008) demuestran que combinar el uso del rostro y la ropa mejora la precisión en la identificación en comparación con su uso por separado.

Everingham, Sivic y Zisserman (2009) crearon un sistema basado en SVM para etiquetado automático de personajes en programas de televisión a partir de subtítulos, transcripciones, detección del habla, rostro y ropa. Los experimentos de Jain y Park (2009) demuestran que el uso de marcas faciales (pecas, verrugas y cicatrices) mejora la precisión en la identificación automática de personas. Weber, Bäuml y Stiefelhagen (2011) encuentran todas las apariciones de un personaje con la misma vestimenta en una serie de televisión.

Los rasgos biométricos suaves de comportamiento también pueden ser caracterizados con fines de identificación. La clasificación del estado afectivo de una persona (como felicidad, tristeza, miedo, enojo, disgusto o sorpresa) se ha explorado ampliamente a partir de la expresión facial (Zeng *et al.*, 2007). El estado afectivo de una persona puede cambiar rápidamente; sin embargo un análisis a largo plazo puede ayudar a determinar los rasgos de personalidad. Batrinca *et al.* (2012) diseñaron un estudio para reconocer los rasgos de personalidad a partir de videos de una persona que realiza una tarea interactiva con una máquina. Sidorov y Minker (2014) crearon un sistema audiovisual para reconocer afecto y depresión a partir del análisis del rostro de personas grabadas en video.

Para caracterizar los rasgos biométricos suaves emergentes se necesita partir de rasgos biométricos suaves localizados y/o de comportamiento. Lin, Lu y Zhang (2006) utilizan un clasificador SVM para determinar el género, la edad y la etnicidad a partir del rostro de la persona. Yang *et al.* (2011b)

Cuadro 3.3: Representación computacional de propiedades de rasgos biométricos suaves localizados.

Propiedad	Rasgo	Método	Autores
Color	Cabello	Modelo Lambertiano de superficie	Yacooob y Davis (2006)
		Modelo Mixto Gaussiano en RGB	Sivic, Zitnick y Szeliski (2006)
	Piel	Valor medio de HSV	Cushen y Nixon (2012)
		Intervalo de valores HSV	Yao, Khan y Farbiz (2011)
		Histograma LAB	Yang y Yu (2011a)
		Modelo de distribución Gaussiana en HCL	Hu, Yan y Lin (2008)
	Ropa	Valor medio de HSV	Cushen y Nixon (2012)
		Histograma HSV	Yao, Khan y Farbiz (2011)
		LAB	Chen, Gallagher y Girod (2012)
		Histograma LAB	Yang y Yu (2011a)
		Modelo Mixto Gaussiano en RGB	Hu, Yan y Lin (2008)
			Hasan y Hogg (2010)
			Sivic, Zitnick y Szeliski (2006)
		Histograma RGB	Weber, Bäuml y Stiefelhagen (2011)
		Histograma RGB e Histograma CIE L*a*b*	Yamaguchi <i>et al.</i> (2012)
		Histograma LCC	Gallagher y Chen (2008)
		Histograma YCbCr	Everingham, Sivic y Zisserman (2009)
Color dominante en HSV	Fu <i>et al.</i> (2012)		
Clusterización en K-medias de valores RGB	Anguelov <i>et al.</i> (2007)		
Textura	Cabello	Ondículas de Gabor	Yacooob y Davis (2006)
	Ropa	Filtros de respuesta máxima, SIFT	Chen, Gallagher y Girod (2012)
		Histograma de respuestas de filtros de Gabor	Yamaguchi <i>et al.</i> (2012)
		HOG, BoW de características SIFT densas, respuesta de DCT	Yang y Yu (2011a)
		LBP, SIFT, HOG, filtro de Gabor	Fu <i>et al.</i> (2012)
		Histogramas de detección de bordes verticales y horizontales	Gallagher y Chen (2008)
		Clusterización en K-medias de respuestas de filtros de Gabor	Anguelov <i>et al.</i> (2007)
		Descriptores estadísticos y descriptores SIFT de las imágenes originales y las sub-bandas de las ondículas asociadas	Yang, Yuan y Tian (2011c)
	Tatuajes	LBP	Heflin, Scheirer y Boulton (2012)
	Marcas y cicatrices	HOG	Heflin, Scheirer y Boulton (2012)
Color y textura	Ropa	Histograma de parches	Song y Leung (2006)
Forma	Tatuajes	SIFT	Lee, Jin y Jain (2012)
			Heflin, Scheirer y Boulton (2012)
Longitud	Cabello	Reglas espaciales	Yacooob y Davis (2006)
		Comparación con máscara	Lee <i>et al.</i> (2008)
Volumen	Cabello	Reglas espaciales	Yacooob y Davis (2006)

presentan un trabajo similar basado en una CNN.¹² Chen, Gallagher y Girod (2012) determinan el género a partir de un CRF basado en ropa y accesorios. Li *et al.* (2008) usa el modo de caminar para inferir el género. Wang y Yau (2013) usan la detección de la barba como pista para inferir el género y la edad.

3.3. Recapitulación

En este capítulo se presentó una revisión de los algoritmos para análisis de biométricos suaves. De manera similar al reconocimiento de personas, el reconocimiento de un rasgo se puede descomponer en los subproblemas de detección, localización, aprendizaje, identificación, etc. No todos los rasgos biométricos se ubican físicamente en una parte concreta del cuerpo, por lo que se propuso clasificarlos en tres grupos: localizados, de comportamiento y emergentes. Sólo los elementos del primer grupo pueden ser delimitados en una región del cuerpo y se explicaron tres formas de señalar su ubicación: como objeto, como material y como objeto articulado.

Se procedió con una revisión de algoritmos para detectar rasgos biométricos y se encontró una gran cantidad de métodos de aprendizaje de máquinas aplicados en la detección de rasgos como el rostro, la piel, la pose, la mano, etc. Se concluyó que la tendencia de los últimos años es el uso de clasificadores profundos para la detección de rasgos. También se planteó la detección basada en contexto, que consiste en usar la localización de un rasgo como apoyo para la detección de otro rasgo, de manera similar a las cadenas de clasificadores.

Posteriormente, se presentó una revisión de algoritmos de aprendizaje e identificación. Para ello, se mostraron diferentes métodos que permiten representar la información de un rasgo, por ejemplo, un histograma RGB para representar el color de la playera de una persona. Esta representación se puede almacenar en una base de conocimiento en la fase de aprendizaje y se puede comparar con otras del mismo tipo con fines de identificación.

La revisión de la literatura de este capítulo permite mostrar que existe un gran repertorio de técnicas. Sin embargo, todos estos trabajos de investigación están realizados fuera del contexto de robótica de servicio. En la mayoría de ellos, la información de entrada tiene restricciones fuertes, por

¹²Siglas de *Convolutional Neural Network* (Red Neuronal Convolutacional).

ejemplo, funcionan a partir de fotos de rostro o de cuerpo completo con condiciones de captura similares (de frente, a la misma distancia, centradas, etc.). Este tipo de imágenes son difíciles de obtener cuando un robot de servicio tiene que buscar e identificar a una persona. En este trabajo de investigación se realizará detección de biométricos suaves basada en contexto mediante el esquema $\{pose_corporal^a\} \rightarrow \{rostro^o, ropa^o, piel^o, cabello^o\}$. Además, se implementará una estrategia de aprendizaje en línea que permite almacenar varias vistas de la misma persona en distintas orientaciones con el fin de tener un modelo más completo. Otra de las diferencias es que en este trabajo se está proponiendo no sólo un algoritmo, sino un sistema que resuelve distintos subproblemas del reconocimiento de personas y que se puede integrar en un robot de servicio. Los detalles de este sistema se explican en el siguiente capítulo.

Capítulo 4

Sistema de aprendizaje en línea e identificación de personas basado en biométricos suaves

En este capítulo se describe la arquitectura e implementación de un sistema para aprendizaje en línea e identificación de personas basado en biométricos suaves que puede ser utilizado por robots de servicio. El objetivo de este sistema es proveer un método robusto y ágil para memorizar personas por medio de la construcción de un modelo 3D con la información suficiente para usarse posteriormente en el proceso de identificación.

En la Figura 4.1 se presenta un diagrama de bloques del sistema propuesto. El sistema recibe información del ambiente en forma constante mediante una cámara RGB y un sensor de profundidad.¹ El sistema se conforma por un conjunto de módulos que se encargan de resolver distintos subproblemas del reconocimiento de personas. Una interfaz de consulta se encarga de responder solicitudes accediendo a la información y algoritmos disponibles en los distintos módulos del sistema.

4.1. Detección, localización y rastreo de personas

El Sistema de Detección, Localización y Rastreo de Personas (SDLRP) se tomó del SDK que provee Microsoft para Kinect 2.0.² Este módulo utiliza

¹Los sensores forman parte de un *Kinect 2.0*.

²Kinect para Windows, <https://dev.windows.com/en-us/kinect>

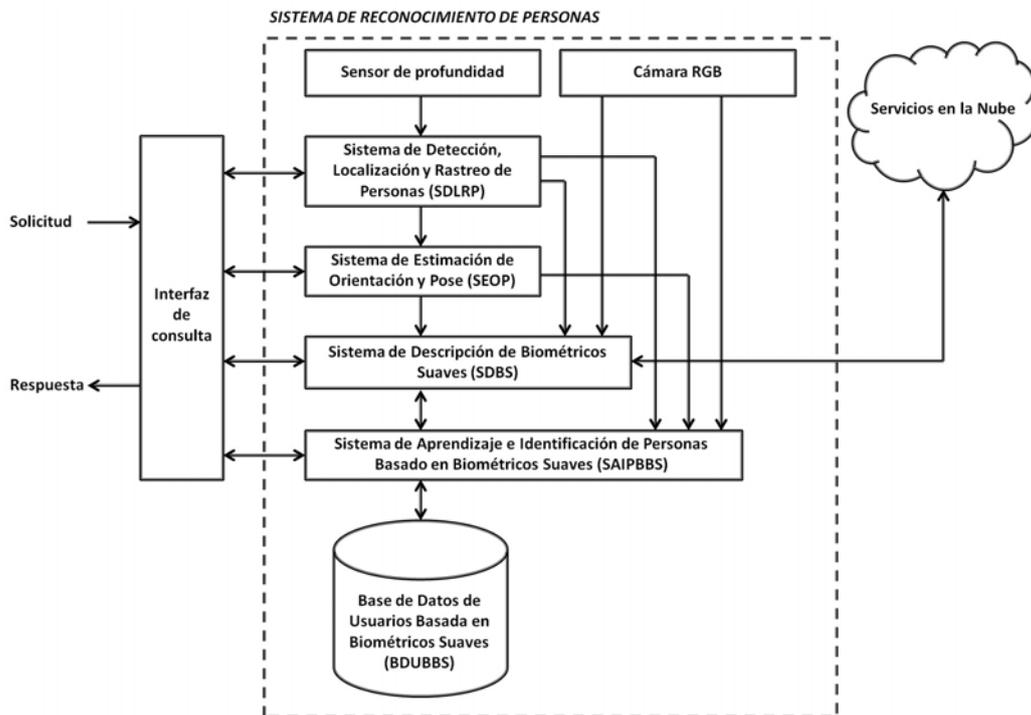


Figura 4.1: Diagrama de bloques del Sistema de Reconocimiento de Personas Basado en Biométricos Suaves. El sistema obtiene su información en forma constante mediante una cámara RGB y un sensor de profundidad. El SDLRP se encarga de la detección, localización y rastreo de personas. El SEOP y el SDBS tienen como función la descripción de la persona apoyándose de servicios en la nube. El SAIPBBS se ocupa del aprendizaje e identificación de personas, almacenando su información en la BDUBBS. La interfaz de consulta se encarga de gestionar las peticiones al sistema. En este capítulo se presenta una explicación detallada del sistema y de cada uno de sus módulos.

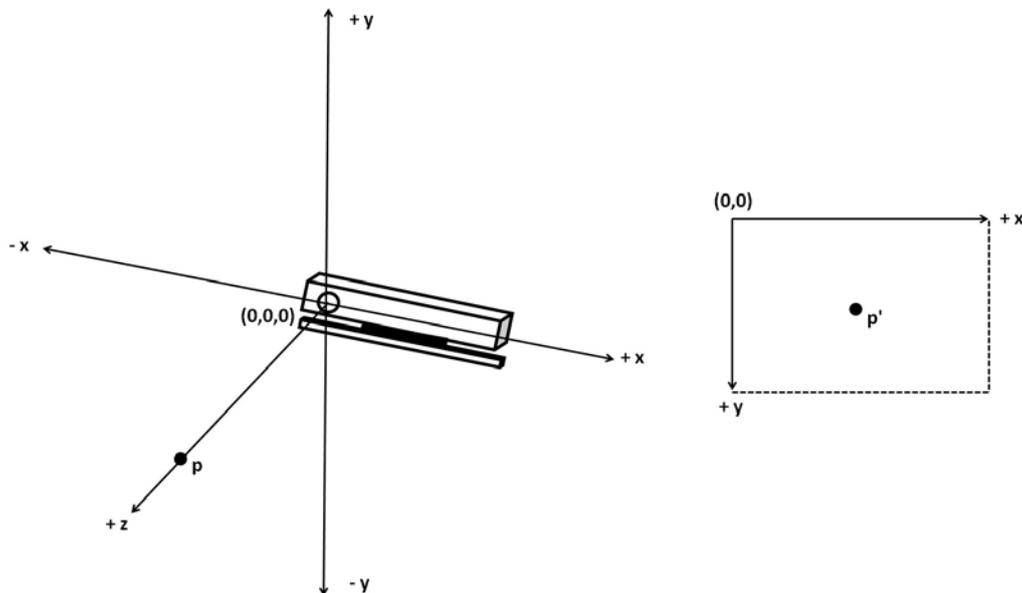


Figura 4.2: Sistema de referencia del sensor de profundidad de Kinect 2.0. En la parte izquierda de la figura se ilustra el sistema de referencia 3D cuyo origen corresponde al sensor de profundidad de Kinect. En la parte derecha de la figura se muestra el sistema de referencia 2D cuando las coordenadas tridimensionales se proyectan sobre el plano bidimensional. La proyección del punto p ubicado sobre el eje z en el sistema de referencia 3D corresponde al punto p' en el sistema de referencia 2D que se utiliza en las imágenes del sensor de profundidad, de modo que el punto p' queda en el centro de la imagen.

como entrada la información del sensor de profundidad y puede detectar hasta seis personas en la escena. La localización de una persona se da como objeto articulado mediante las coordenadas 3D de 25 puntos de interés del esqueleto (incluyendo articulaciones y extremidades). En la Figura 4.2 se presenta el sistema de referencia que utiliza el sensor de profundidad. En la Figura 4.3 se muestran los puntos de interés del esqueleto.

Cuando se detecta a una persona nueva el sistema le asigna un identificador numérico entre 0 y 5 que sirve para su rastreo dentro de la escena. Este sistema no realiza aprendizaje de la persona por lo que no es posible realizar identificación. Por ejemplo, si una persona tiene asignado el identificador 4 y se sale de la escena o es ocluida, al momento de volver a ser detectada el sistema le asigna en forma arbitraria uno de los identificadores

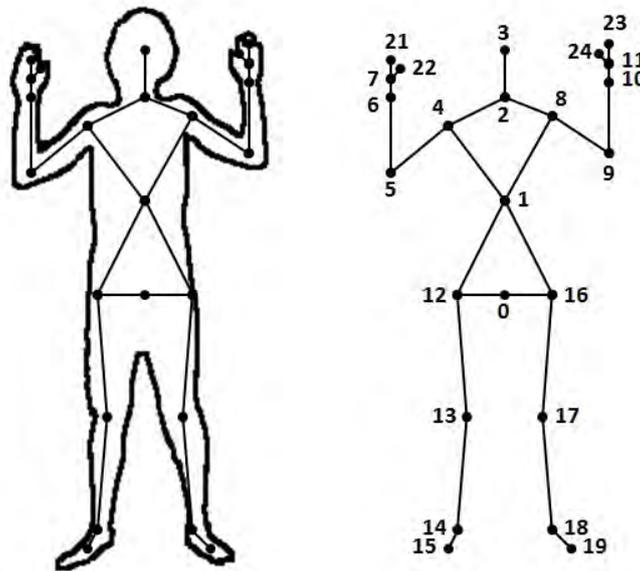


Figura 4.3: Puntos de interés del esqueleto generado por el SDK de Microsoft para Kinect 2.0. Se inicia la numeración en cero para hacer coincidir la lista de puntos de interés con su respectiva posición en el arreglo del código fuente del SDK. Los puntos de interés corresponden a las siguientes partes del cuerpo humano:

- | | |
|-----------------------------------|--------------------------------|
| 0. Base de la columna vertebral | 13. Rodilla izquierda |
| 1. Centro de la columna vertebral | 14. Tobillo izquierdo |
| 2. Cuello | 15. Pie izquierdo |
| 3. Cabeza | 16. Cadera derecha |
| 4. Hombro izquierdo | 17. Rodilla derecha |
| 5. Codo izquierdo | 18. Tobillo derecho |
| 6. Muñeca izquierda | 19. Pie derecho |
| 7. Mano izquierda | 20. Espalda |
| 8. Hombro derecho | 21. Punta de la mano izquierda |
| 9. Codo derecho | 22. Pulgar izquierdo |
| 10. Muñeca derecha | 23. Punta de la mano derecha |
| 11. Mano derecha | 24. Pulgar derecho |
| 12. Cadera izquierda | |

disponibles. Es importante señalar que este software fue diseñado para videojuegos y supone que el sensor se mantiene en la misma posición todo el tiempo, por lo que el desempeño se ve afectado cuando el sensor se encuentra en un sistema móvil.

4.2. Descripción de personas

El Sistema de Estimación de Orientación y Pose (SEOP) se implementó utilizando como entrada los datos del SDLRP y trabaja a la misma frecuencia que la toma de imágenes del sensor de profundidad. En cada muestra se analizan todas las personas detectadas por el SDLRP y se determina para cada una de ellas:

- La posición de la persona sobre el plano XZ en coordenadas polares (r, θ) .
- El ángulo de orientación de la persona con respecto al sensor (0° corresponde a una persona encarando al sensor; $\pm 90^\circ$ si está de costado y 180° si se encuentra dando la espalda al mismo).
- La pose de los brazos, la cual se clasifica en: i) ambos brazos a los costados (apuntando hacia abajo), ii) un brazo extendido (apuntando al horizonte), iii) un brazo levantado (apuntando hacia arriba), iv) ambos brazos levantados (apuntando hacia arriba).
- La pose del cuerpo, la cual se clasifica en: i) de pie, ii) sentada o iii) recostada en el piso.

Todos estos resultados se obtuvieron a partir de la aplicación de geometría analítica en 3D a los vértices que conforman el esqueleto de la persona. En el caso de la pose de brazos y cuerpo se utilizó un conjunto de reglas basadas en la posición relativa de algunos vértices de interés. Por ejemplo, si el vértice de la mano derecha se encuentra por encima del vértice de la cabeza se considera que la persona está levantando la mano. En la pose de brazo extendido hacia el horizonte se calcula también el ángulo sobre el plano XZ hacia el cual está apuntando el usuario siguiendo la misma convención que el ángulo de orientación de la persona con respecto al sensor (0° si la persona apunta al sensor, $\pm 90^\circ$ si apunta hacia la derecha o izquierda). En el Apéndice A se presenta una explicación detallada sobre la forma en que se determina la pose de los brazos y del cuerpo.

A diferencia del SDLRP y el SEOP que trabajan de manera ininterrumpida, el Sistema de Descripción de Biométricos Suaves (SDBS) funciona bajo demanda del usuario en momentos específicos. Cuando se le realiza una petición de descripción de una persona en la escena recopila la información del SDLRP, el SEOP y la cámara RGB. El SDBS envía la imagen RGB de la persona a un servicio en la nube³ para obtener la siguiente información de la persona:

- Género
- Edad
- Presencia de lentes
- Presencia de vello facial:
 - Barba
 - Bigote
 - Patilla
- Presencia de sonrisa
- Estado emocional:
 - Enojo
 - Desprecio
 - Asco
 - Miedo
 - Felicidad
 - Tristeza
 - Sorpresa
 - Neutral
- Inclinación del rostro:
 - Ángulo sobre el plano XY
 - Ángulo sobre el plano XZ
 - Ángulo sobre el plano YZ

³*Microsoft Cognitive Services.*

Dentro del SDBS se implementó un proceso que a partir del esqueleto de la persona determina la estatura de la persona y el color principal de la playera. La descripción de la persona consiste en un vector que contiene todos los elementos señalados en esta sección.

4.3. Aprendizaje de personas

El Sistema de Aprendizaje e Identificación de Personas Basado en Biométricos Suaves (SAIPBBS) trabaja bajo demanda del usuario realizando almacenamientos y consultas en la Base de Datos de Usuarios Basada en Biométricos Suaves (BDUBBS). Este sistema requiere para su funcionamiento la información proveniente de la cámara RGB, el SDLRP, el SEOP y el SDBS.

Al realizar una solicitud de aprendizaje al sistema se especifica cuál de las personas detectadas en la escena debe ser memorizada y se brinda una etiqueta ID para ser usada como identificador. El sistema crea una observación $O_\theta = \langle I, S \rangle$, en la cual el índice θ es el ángulo de orientación de la persona, I es una imagen RGB de alta resolución y $S = \langle J_0, J_1, \dots, J_{24} \rangle$ es el esqueleto del usuario, en el cual cada articulación $J_i = \langle x, y, t \rangle$ es una 3-tupla que incluye una coordenada (x, y) y una bandera lógica t que es verdadera si la articulación está siendo detectada.

La observación O_θ se almacena en la base de datos de usuarios indexada por la etiqueta y el ángulo. Si no hay una observación previa de la persona con el identificador ID se crea un nuevo registro $P_{ID} = \{O_\theta\}$ en la BDUBBS. Si ya existe un registro con el mismo identificador se recolecta la nueva observación haciendo $P_{ID} = P_{ID} \cup \{O_\theta\}$. De esta manera, el sistema puede almacenar varias muestras del mismo usuario en distintas orientaciones. Por ejemplo, en la Figura 4.4 se presenta un conjunto de observaciones del mismo usuario en distintas orientaciones que se almacenan en la base de datos indexadas con la misma etiqueta pero con diferente ángulo.

Este procedimiento permite incrementar gradualmente la información acerca de una persona añadiendo información en la base de datos en momentos estratégicos y no tiene restricciones respecto al orden en el que se obtienen las observaciones. También es posible intercalar el proceso de toma de observaciones de diferentes personas.



Figura 4.4: Ocho observaciones del mismo usuario almacenadas en la BDUBBS. Cada observación se almacena indexada por el identificador del usuario y la orientación a la que se encuentra con respecto al dispositivo sensor: (0° si la persona está encarando al sensor, $\pm 90^\circ$ si está de costado y 180° si se encuentra dando la espalda).

4.4. Identificación de personas

Cuando se le solicita al SAIPBBS realizar la identificación de la persona x de la escena se genera la observación $O_x = \langle I_x, S_x \rangle$ y se compara con las observaciones almacenadas en la base de datos. Para cada persona P_i en la base de datos se selecciona la observación más cercana O_y del usuario i en términos de su orientación calculando $\operatorname{argmin}_{O_y \in P_i} |x - y|$. Se determina la similitud $D(O_x, O_y)$ entre la persona observada y la persona i en la base de datos. El proceso de identificación tiene los modos grueso, fino y basado en descripción, que difieren en la función de similitud D y que se explican más adelante. Al finalizar el proceso, el sistema escoge al usuario más similar en la base de datos, el cual corresponde a $\operatorname{argmax}_i D(O_x, \operatorname{argmin}_{O_y \in P_i} |x - y|)$. Si la similitud entre O_x y $O_y \in P_{\max_i}$ es mayor que el umbral γ , el sistema regresa $ID = i$; de lo contrario responde con la etiqueta reservada *unknown*, la cual indica que la persona observada no se encuentra dentro la base de datos.

4.4.1. Identificación gruesa

El modo de comparación grueso está creado para situaciones en las cuales el usuario se encuentra demasiado lejos y funciona mediante el análisis de la ropa a partir de una muestra que se extrae del área del torso.

Como primer paso se extraen las regiones que corresponden a la región de

ropa del torso a_x y a_y de ambas imágenes. Los esqueletos S_x y S_y se utilizan como pistas contextuales para generar un área rectangular contenida en el polígono delimitado por los vértices que corresponden a los hombros (vértices 4 y 8) y las caderas (vértices 12 y 16) y centrado en el vértice que corresponde al punto medio de la columna vertebral (vértice 1). Debido a que no existe garantía de que los cinco vértices se rastreen simultáneamente, se genera un estimado del tamaño de la región de interés mediante un promedio ponderado que considera las distancias entre algunas parejas de los vértices que están siendo detectados.

Debido a que la imagen del sensor de profundidad y la imagen de la cámara RGB no tienen el mismo origen dentro del sistema de referencia, se aplica una transformación lineal (de traslación y escalamiento) $T(x, y) = (x', y')$ para obtener la posición (x', y') del vértice en la imagen de color a partir de su posición (x, y) en la imagen de profundidad.

Se procede con el cálculo de los histogramas H_x y H_y de las regiones seleccionadas. Se calcula el coeficiente de correlación $C(H_x, H_y)$ como medida de similitud a partir de la siguiente fórmula:

$$C(H_x, H_y) = \frac{\Sigma(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\Sigma(X_i - \bar{X})^2} \sqrt{\Sigma(Y_i - \bar{Y})^2}}$$

La elección de la correlación como medida de similitud se realizó mediante experimentos preliminares de identificación de personas. El primer experimento se realizó con fotos obtenidas de la web y con el recorte de la región de la playera realizado por un humano. El segundo experimento se realizó con fotos tomadas por el robot y con identificación automática del esqueleto. Los experimentos se realizaron con los siguientes métodos de comparación de histogramas: prueba de chi cuadrado, intersección, distancia de Bhattacharyya y correlación. Este último método mostró mejores resultados por lo que se decidió utilizarlo en la versión final del sistema.

En la Figura 4.5 se presenta un ejemplo del funcionamiento del sistema. En la parte izquierda se muestran dos imágenes de personas; en la parte central se presentan las regiones de ropa extraídas automáticamente. Es importante notar que el tamaño de las regiones difiere debido a que el área de interés está en función de las medidas del cuerpo de la persona. Finalmente, en la parte derecha se muestran los histogramas RGB⁴ y su correlación.

⁴Las gráficas fueron elaboradas con el graficador de histogramas incluido en la documentación de OpenCV: <http://docs.opencv.org>



Figura 4.5: Comparación gruesa de dos personas. En la parte izquierda se muestran dos imágenes de personas diferentes. Utilizando la información del esqueleto que provee el SDLRP se determinan cuatro puntos (en verde) que sirven como vértices de un rectángulo que delimita la región del torso. En la parte central se muestran los segmentos de imagen extraídos que contienen la ropa de la persona en la zona del torso y en la parte derecha se muestran sus respectivos histogramas RGB. El cálculo de la correlación entre ambos histogramas sirve como medida de similitud.

4.4.2. Identificación fina

El modo de comparación fino se utiliza cuando el usuario está cerca del sensor RGB-D.⁵ Las articulaciones del esqueleto se usan como vértices y se define una gráfica no dirigida mediante una matriz triangular inferior de 25×25 :

$$G_{i,j} = \begin{cases} 0 & \text{si } i < j \\ r & \text{si } i \geq j \end{cases}$$

en la cual $r = 1$ si la arista es relevante en el análisis del cuerpo humano (por ejemplo, la conexión entre el hombro izquierdo y el codo izquierdo) y $r = 0$ para una arista sin sentido (como la conexión entre la mano izquierda y la rodilla derecha). Un vértice de interés k se declara como $G_{k,k} = 1$.⁶

Para cada arista relevante $G_{i,j} = 1$ se revisan los esqueletos S_x y S_y de las observaciones. Si los vértices J_i y J_j de ambos esqueletos están siendo detectados, entonces las aristas $E_x(i, j)$ y $E_y(i, j)$ se comparan por medio de la extracción de un parche centrado en el punto medio de la arista. El tamaño de los parches es directamente proporcional a las medidas del cuerpo humano. Si la arista es un bucle, el parche está centrado en el vértice. En la Figura 4.6 se presenta un ejemplo de los parches seleccionados en dos vistas de un mismo usuario.

Los parches p_x y p_y se recortan y se calculan sus histogramas H_x y H_y . Si la correlación $C(H_x, H_y)$ es mayor que el umbral σ , la pareja de parches da un voto positivo.⁷ El algoritmo suma el número de votos de todas las comparaciones entre parejas de parches y el resultado se utiliza como medida de similitud.

En la Figura 4.7 se presenta un ejemplo de funcionamiento del sistema en

⁵Como resultado de experimentos del sistema se observó que el mejor desempeño ocurre cuando el usuario se encuentra a una distancia de 1.5 metros del sensor RGB-D.

⁶Se eligió una representación de matriz triangular inferior por simplicidad. La otra opción sería utilizar la matriz cuadrada de 25×25 para representar el grafo no dirigido, de modo que si una arista conecta a los vértices i y j implicaría que $G_{i,j} = 1$ y $G_{j,i} = 1$.

⁷El valor de σ se asignó como resultado de experimentación constante del sistema. Mediante los experimentos realizados se ha observado que incrementar el valor de σ exige mayor similitud entre los parches de color para llegar a la conclusión de que corresponden a la misma persona, provocando una disminución de los falsos positivos, pero aumentando los falsos negativos. Por el contrario, disminuir el valor de σ exige menor similitud entre los parches y tiene el efecto de disminuir los falsos negativos, pero aumentando los falsos positivos.

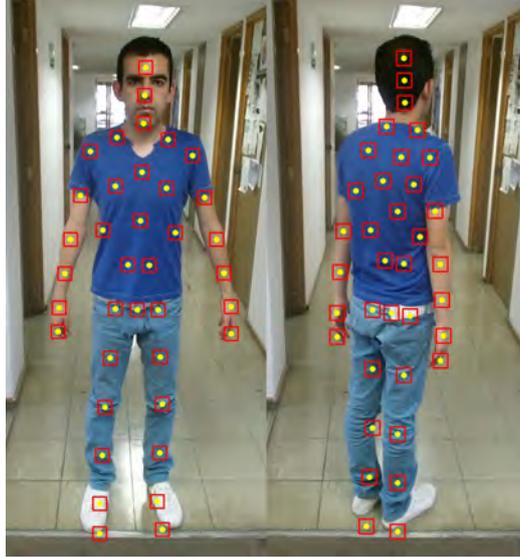


Figura 4.6: Conjunto de parches para comparación fina en dos vistas del mismo usuario

modo fino. En la parte izquierda se muestran dos imágenes de personas; en la parte central se presentan los parches extraídos del centro de la columna vertebral (vértice de interés $G_{1,1}$) y en la parte derecha se muestran los histogramas RGB de la pareja de parches y su correlación.

4.4.3. Identificación a partir de la descripción

En este modo se invoca al SDBS para obtener los vectores de descripción d_1 y d_2 de las personas que se desean comparar. Se aplica una función de categorización $\varphi_i(x)$ a algunos de los elementos de los vectores de descripción para facilitar el proceso de comparación. Por ejemplo, la edad de una persona está expresada en el vector de descripción como un número real por lo que se aplica la función $\varphi_{edad} : \mathbb{R} \rightarrow C_{edad}$, donde $C_{edad} = \{menor, joven, adulto, mayor\}$ es el conjunto de categorías para el atributo edad y $\varphi_{edad}(x)$ se define como:

$$\varphi_{edad}(x) = \begin{cases} menor & si & x \leq 12 \\ joven & si & 12 < x < 18 \\ adulto & si & 18 < x < 60 \\ mayor & si & x \geq 60 \end{cases}$$

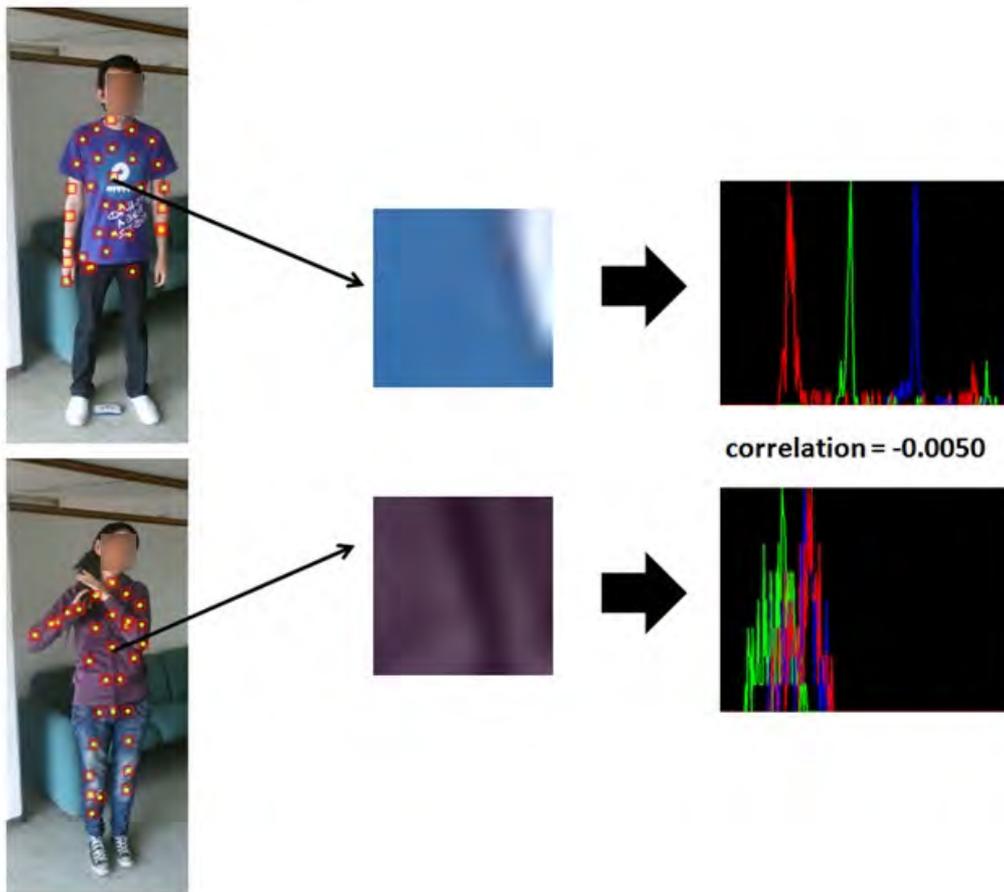


Figura 4.7: Comparación fina de dos personas. En la parte izquierda se muestran dos imágenes de personas diferentes. Utilizando la información del esqueleto que provee el SDLRP se eligen regiones de interés del cuerpo (puntos en amarillo) de las cuales se extraen parches de color (delimitados en la imagen con cuadros rojos). En la parte central se muestra una pareja de parches extraídos de la misma región del cuerpo y en la parte derecha se muestran sus respectivos histogramas RGB. El cálculo de la correlación entre ambos histogramas sirve para determinar si la pareja de parches es suficientemente parecida para emitir un voto que contribuirá a la función de similitud.

Para definir el conjunto de categorías de un atributo se tomó en cuenta la posibilidad de asignar etiquetas en lenguaje natural siguiendo los cuatro criterios propuestos en la sección 2.2 (naturalidad, utilidad, objetividad y entendimiento común). Seguir estos criterios puede generar que las categorías sean desbalanceadas, por ejemplo, el rango de adulto es más grande que el de joven. Una segunda opción es proponer un conjunto de categorías más balanceado en el tamaño de los rangos, con el costo de no contar con etiquetas en lenguaje natural para describirlas de manera simple. Por ejemplo:

$$\varphi_{edad}(x) = \begin{cases} A & \text{si } x \leq 10 \\ B & \text{si } 10 < x \leq 20 \\ C & \text{si } 20 < x \leq 30 \\ D & \text{si } 30 < x \leq 40 \\ E & \text{si } 40 < x \leq 50 \\ F & \text{si } 50 < x \leq 60 \\ G & \text{si } 60 < x \leq 70 \\ H & \text{si } 70 < x \leq 80 \\ I & \text{si } 80 < x \leq 90 \\ J & \text{si } 90 < x \leq 100 \\ K & \text{si } x > 100 \end{cases}$$

En el Cuadro 4.1 se presentan los elementos que forman parte del vector de descripción d' después de aplicar las funciones de categorización. Los elementos 1 a 8 están relacionados con la posición corporal del usuario y su ubicación en el escenario; los elementos 9 a 16 corresponden a atributos físicos de la personas y los elementos 17 a 25 corresponden a atributos de comportamiento. Es importante señalar que los elementos 7 a 25 corresponden a biométricos suaves debido a que el valor de la propiedad es una etiqueta en lenguaje natural.

Para realizar la comparación de las personas se utilizaron los atributos 9 a 16. Si ambas descripciones tienen un atributo con la misma categoría se da un voto. El valor de similitud es el número total de votos obtenidos. Para concluir que ambas descripciones corresponden a la misma persona se exige un umbral alto, de modo que se obtengan casi todos los votos al realizar la comparación.

El producto cartesiano de los conjuntos de categorías de los atributos 9 a 16 $C_{género} \times C_{edad} \times C_{lentes} \times C_{barba} \times C_{bigote} \times C_{patilla} \times C_{estatura} \times C_{colorplayera}$ contiene todos los posibles semblantes o tipos de personas que el sistema

Cuadro 4.1: Vector de descripción después de aplicar las funciones de categorización.

Elemento	Atributo	Categorías
1	posición en coordenada polar (distancia)	no aplica
2	posición en coordenada polar (ángulo)	no aplica
3	orientación respecto al sensor (ángulo)	no aplica
4	inclinación del rostro (ángulo sobre el plano XY)	no aplica
5	inclinación del rostro (ángulo sobre el plano XZ)	no aplica
6	inclinación del rostro (ángulo sobre el plano YZ)	no aplica
7	pose de los brazos	ninguna, un brazo apuntando, un brazo levantado, ambos brazos levantados
8	pose del cuerpo	de pie, sentada, recostada
9	género	masculino, femenino
10	edad	menor, joven, adulto, mayor
11	lentes	sí, no
12	barba	sí, no
13	bigote	sí, no
14	patilla	sí, no
15	estatura	baja, media, alta
16	color de playera	azul, rojo, verde, amarillo, ...
17	sonrisa	sí, no
18	enojo	sí, no
19	desprecio	sí, no
20	asco	sí, no
21	miedo	sí, no
22	felicidad	sí, no
23	tristeza	sí, no
24	sorpresa	sí, no
25	emoción neutra	sí, no

puede identificar. Es posible que dos o más personas compartan el mismo semblante, por ejemplo, que se de la presencia de dos hombres de edad adulta con lentes, barba, bigote, patilla, de estatura media y que usan playera roja. En este caso, el contenido del vector de descripción tendría la misma asignación de valores en los atributos 9 a 16.

El número total de semblantes distintos es igual al producto de las cardinalidades de los conjuntos de categorías involucrados en el producto cartesiano descrito en el párrafo anterior, es decir, $|C_{género}| \cdot |C_{edad}| \cdot |C_{lentes}| \cdot |C_{barba}| \cdot |C_{bigote}| \cdot |C_{patilla}| \cdot |C_{estatura}| \cdot |C_{colorplayera}|$. Un aumento en el número de atributos físicos de la persona o del número de categorías en uno o varios de los atributos tendría como efecto incrementar el número de semblantes distintos que el sistema podría identificar.

No todos los semblantes tienen la misma probabilidad de ocurrir. Existen algunos semblantes más comunes que otros y semblantes que difícilmente podrían ocurrir. Aunque el método propuesto resulta débil en situaciones donde dos o más personas tienen el mismo semblante, resulta útil en situaciones donde se presenta un semblante poco común.

Una de las ventajas de este modo es que no necesita comparar imágenes. Por ejemplo, es posible que una descripción en lenguaje natural proveniente de una conversación se convierta directamente al vector de descripción categorizado y se compare con la descripción de una imagen. De esta forma, se podría realizar identificación de personas a partir de retratos hablados.

4.5. Recapitulación

En este capítulo se explicó la forma en que se diseñó e implementó el Sistema de Reconocimiento de Personas basado en Biométricos Suaves. El sistema está compuesto por distintos módulos en los que se resuelven distintos subproblemas del reconocimiento de personas. Para detección, localización y rastreo se utilizó el SDK de Microsoft para Kinect 2.0, y parte de la descripción se generó a partir de Microsoft Cognitive Services.

A partir de la información generada por los módulos de detección, localización y descripción, se implementó el Sistema de Aprendizaje e Identificación de Personas Basado en Biométricos Suaves, que es la aportación central de este capítulo. Este sistema está vinculado con una Base de Datos de Usuarios Basada en Biométricos Suaves en donde es posible almacenar

la información de las personas que aprende el sistema. Además, es posible almacenar múltiples vistas del mismo usuario, permitiendo que el sistema pueda identificar a una persona sin importar su orientación con respecto al dispositivo sensor.

Se implementaron tres métodos de identificación de personas a partir de biométricos suaves: el modo grueso, el modo fino y el modo basado en la descripción. El modo grueso se basa en extraer la información de color de la ropa en el torso de la persona. Por su parte, el modo fino extrae información de color de múltiples regiones del cuerpo, incluyendo torso, brazos, piernas, pies, etc. El modo basado en la descripción utiliza el género, la edad, la estatura, el color de la ropa y la presencia de vello facial y lentes para comparar a dos personas.

Aunque los tres modos de identificación presentados en esta sección se probaron por separado, la forma en que se integraron dentro del robot de servicio permite que se puedan combinar de manera sencilla invocándolos en cascada. Por ejemplo, en una aplicación particular se podría invocar identificación fina, y en caso de tener éxito proceder con una segunda verificación mediante identificación a partir de la descripción. Esto se explicará a detalle en el siguiente capítulo, en el que se abordará la integración del Sistema de Reconocimiento de Personas basado en Biométricos Suaves propuesto en este capítulo en un robot de servicio.

Capítulo 5

Integración del sistema en un robot de servicio

El sistema descrito en el capítulo anterior se puede usar desde una computadora fija que cuente con un sensor RGB-D. En este modo de uso se requiere que un humano controle el sistema y decida cuándo invocar los procesos de aprendizaje e identificación de personas. Sin embargo, si se desea integrar el sistema a un robot de servicio para ser usado de manera autónoma, resulta necesario situar a los procesos de aprendizaje e identificación en el contexto de una tarea con llamadas al SAIPBBS en momentos específicos. De esta manera el reconocimiento de personas se puede utilizar en comportamientos tan complejos como la combinación de las distintas habilidades del robot lo permita.

Para integrar el SAIPBBS en un robot de servicio se definió un conjunto de conductas siguiendo el concepto y la estructura funcional de un robot de servicio (Pineda *et al*, 2013a). Este modelo conceptual inspirado en la jerarquía de niveles de sistema de Newell (1981) y los niveles computacional, algorítmico e implementacional de Marr (1982) organiza la investigación en robótica en tres capas: i) el robot de servicio, ii) los algoritmos de percepción y acción, y iii) la programación del sistema. La primera de estas capas, que corresponde al nivel del conocimiento de Newell y al nivel computacional de Marr, es una especificación de alto nivel de las funciones de un robot de servicio y es por sí mismo un objeto de investigación (*Íbidem*).

Este modelo propone un conjunto de conductas robóticas básicas definidas desde el nivel del robot de servicio, con una especificación abstracta enfocada en su funcionalidad y no en los algoritmos o en los dispositivos de hardware que utilizan, mismos que corresponden a las otras capas de la jerarquía

propuesta. En la Figura 5.1 se presenta una versión extendida de la biblioteca de conductas del robot Golem (*Íbidem*), que incluye las conductas relacionadas con reconocimiento de personas mediante biométricos suaves. Esta biblioteca de conductas es jerárquica, ya que muchas de las conductas se definen utilizando conductas inferiores de la jerarquía. Como se puede ver en la Figura 5.1, las conductas relacionadas con biométricos suaves forman parte de la conducta de búsqueda *find*, la cual se explicará a detalle en este capítulo.

Las conductas son unidades atómicas que se pueden ensamblar para formar una tarea. El modelo conceptual provee mecanismos para ensamblar las conductas durante la ejecución de tareas, que pueden ser estáticos (cuando la estructura de la tarea se conoce de antemano) o dinámicos (cuando el robot debe atender comandos arbitrarios). Además, el modelo consta de una representación explícita de la estructura de la tarea. Una instanciación del modelo conceptual propuesto es la que se mencionó en el Capítulo 1 y que está integrada por tres elementos: la Arquitectura Cognitiva Orientada a la Interacción, los Modelos de Diálogo y SitLog. Estos elementos se explicarán con más detalle en este capítulo (*Íbidem*).

5.1. Conductas de reconocimiento de personas

El funcionamiento del SAIPBBS se embebe en las conductas *detect_body*, *memorize_body*, *identify_body* y *describe_body*, que internamente se encargan de gestionar las solicitudes de reconocimiento de personas mediante biométricos suaves.

La conducta *detect_body* se encarga de la detección y localización de personas en la escena. Las conductas *memorize_body*, *identify_body* y *describe_body* se encargan del aprendizaje, la identificación y la descripción respectivamente. Estas tres conductas tienen como precondition la ejecución de la conducta *detect_body*, aunque es posible intercalar una conducta entre ellas. Por ejemplo, en una tarea en particular se podría utilizar la conducta *detect_body* y en caso de detectar a una persona enfrente del robot proceder con la conducta *ask* para solicitar su nombre y continuar con la conducta *memorize_body*.

La conducta *see_body* administra automáticamente la secuencia de las

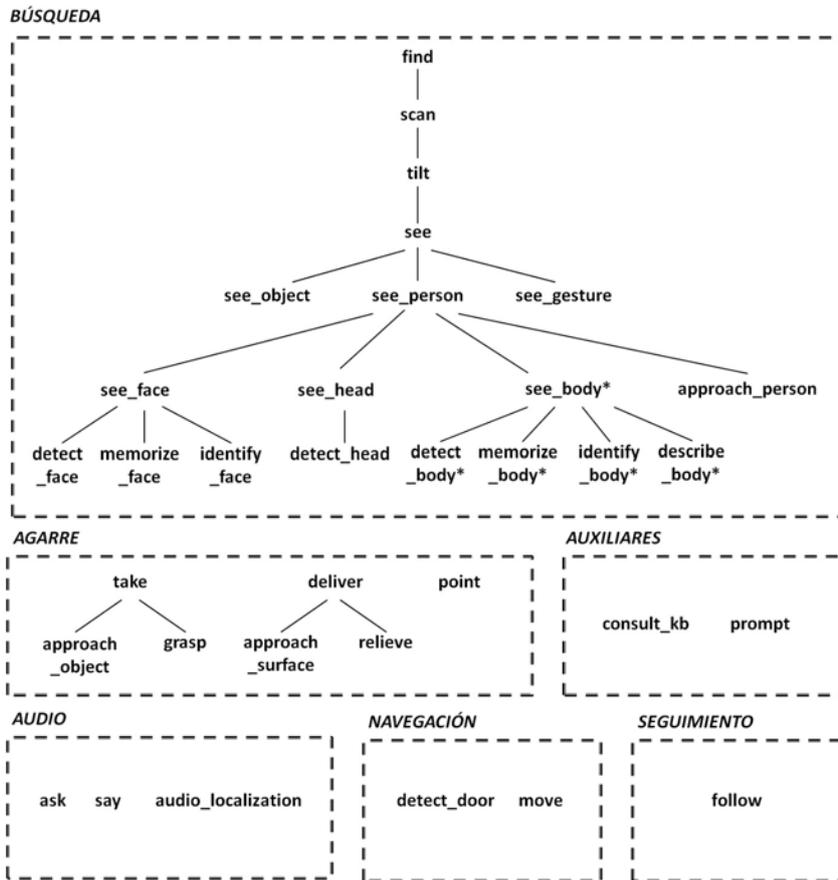


Figura 5.1: Biblioteca de conductas del robot Golem (Pineda *et al*, 2013a) extendida con las conductas de biométricos suaves (marcadas con un asterisco)

conductas anteriores, creando encadenamientos de las conductas básicas dependiendo del modo en el que se invoca. La Figura 5.2 muestra un diagrama de bloques de esta conducta que incluye los distintos modos de uso:

- El modo *detect* ejecuta directamente la conducta *detect_body*.
- El modo *memorize* ensambla serialmente las conductas *detect_body* y *memorize_body*.
- El modo *identify* ensambla serialmente las conductas *detect_body* e *identify_body*.
- El modo *describe* ensambla serialmente las conductas *detect_body* y *describe_body*.

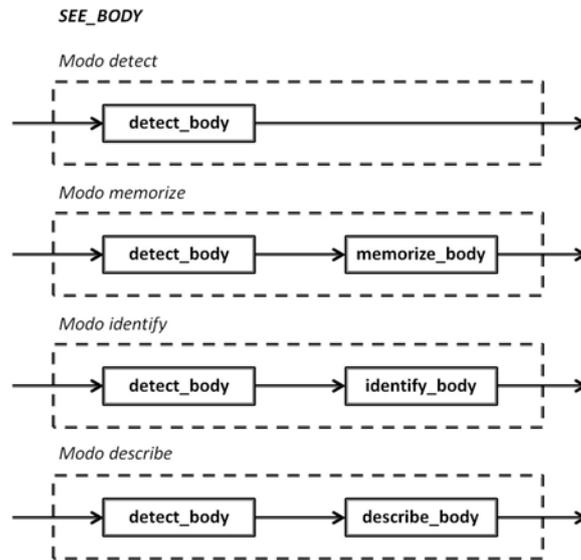


Figura 5.2: Conducta *see_body*. En este diagrama se muestran los distintos modos en que puede operar esta conducta. En el modo *detect* se detecta la presencia de una persona. En el modo *memorize* se detecta a una persona y a continuación se aprenden sus características. En el modo *identify* se detecta a una persona y a continuación se identifica. En el modo *describe* se detecta a una persona y posteriormente se obtiene su descripción. En todos los casos se trabaja con el cuerpo de la persona.

De manera análoga al subárbol cuya raíz es *see_body*, el conjunto de conductas contiene el subárbol cuya raíz es *see_face*, en el cual los procesos de detección, aprendizaje e identificación se enfocan en el rostro de la persona mediante biométricos duros. En la Figura 5.3 se presenta el diagrama de bloques de esta conducta. Dentro de la biblioteca de conductas también se encuentra la rama *see_head*, que incluye una conducta de detección de la región de cabeza y hombros (Romero Cordero, 2015).

5.2. Integración de habilidades motoras en las conductas de reconocimiento

El reconocimiento de personas mediante biométricos suaves se integró en conductas que permiten aprovechar el movimiento del robot, tanto en su capacidad de navegar en el escenario como en el uso de las articulaciones del cuello para ver en la dirección deseada. La conducta *approach_person* está compuesta por una iteración de las conductas *move* y *detect_body* que

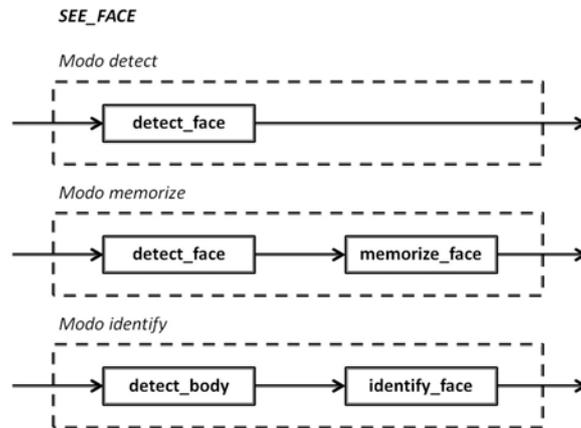


Figura 5.3: Conducta *see_face*. En este diagrama se muestran los distintos modos en que puede operar esta conducta. En el modo *detect* se detecta la presencia de una persona. En el modo *memorize* se detecta a una persona y a continuación se aprenden sus características. En el modo *identify* se detecta a una persona y a continuación se identifica. En todos los casos se trabaja con el rostro de la persona.

termina cuando el robot llega a una posición en la que el usuario queda en el centro del ángulo de visión y se puede ver de cuerpo completo. Para ello se realizan desplazamientos (acercarse o alejarse) y rotaciones para colocarse a una distancia y orientación adecuada del usuario antes de realizar el aprendizaje o la identificación. En la Figura 5.4 se presenta un diagrama que resume el funcionamiento de esta conducta.

La conducta *see_person* es una conducta genérica de observación de personas que a través de sus modos permite elegir el motivo de la observación (detectar, memorizar, identificar o describir), el tipo de rasgo a utilizar (biométricos o biométricos suaves) y si se desea que el robot se aproxime a la persona. La Figura 5.5 presenta un diagrama de bloques de esta conducta que incluye los distintos modos basados en biométricos suaves. La Figura 5.6 presenta los modos basados en biométricos duros.

En la conducta *see_person* también se incluyen modos de funcionamiento que combinan el uso de biométricos suaves y biométricos duros para realizar procesos de detección, memorización o identificación con redundancia. Estos esquemas dan inicio con la detección del cuerpo del usuario y a continuación se procede con la aproximación a la persona. Una vez que el robot se encuentra a la distancia adecuada se realiza el reconocimiento a partir de

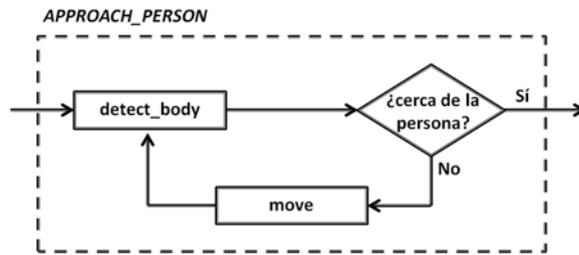


Figura 5.4: Conducta *approach-person*. El diagrama de flujo muestra el uso de las conductas *detect_body* y *move* dentro de un iteración en la que el robot se acerca gradualmente a la persona y que termina cuando se encuentra a una distancia adecuada.

biométricos suaves. A continuación, el robot le informa al usuario que lo mire a la cámara y procede con la fase de reconocimiento mediante rostro. En caso de fallo en la detección del rostro se solicita de nuevo al usuario que vea a la cámara y se intenta otra vez. La Figura 5.7 presenta los modos basados en biométricos suaves y biométricos duros.

La conducta *see-person* en cualquiera de sus modos se puede usar desde conductas superiores en el conjunto jerárquico de conductas (Pineda *et al.*, 2013a). La conducta *see* es un proceso genérico de observación en la posición y orientación del cuello actual del robot y se puede usar para reconocer objetos, personas y gestos. La conducta *tilt* realiza un conjunto de conductas *see* cambiando el ángulo vertical del cuello del robot. La conducta *scan* ejecuta una serie de conductas *tilt* en diferentes ángulos horizontales del cuello del robot. Finalmente, la conducta *find* permite definir un conjunto de puntos en el escenario a los cuales el robot debe ir para ejecutar la conducta *scan*. La estructura de conductas *tilt-scan-find* permite tomar ventaja de los movimientos articulados de la cabeza del robot y de la habilidad de navegación para explorar el ambiente en los procesos de búsqueda de personas, ya sea para detección, aprendizaje o identificación, tanto con biométricos duros como con biométricos suaves.

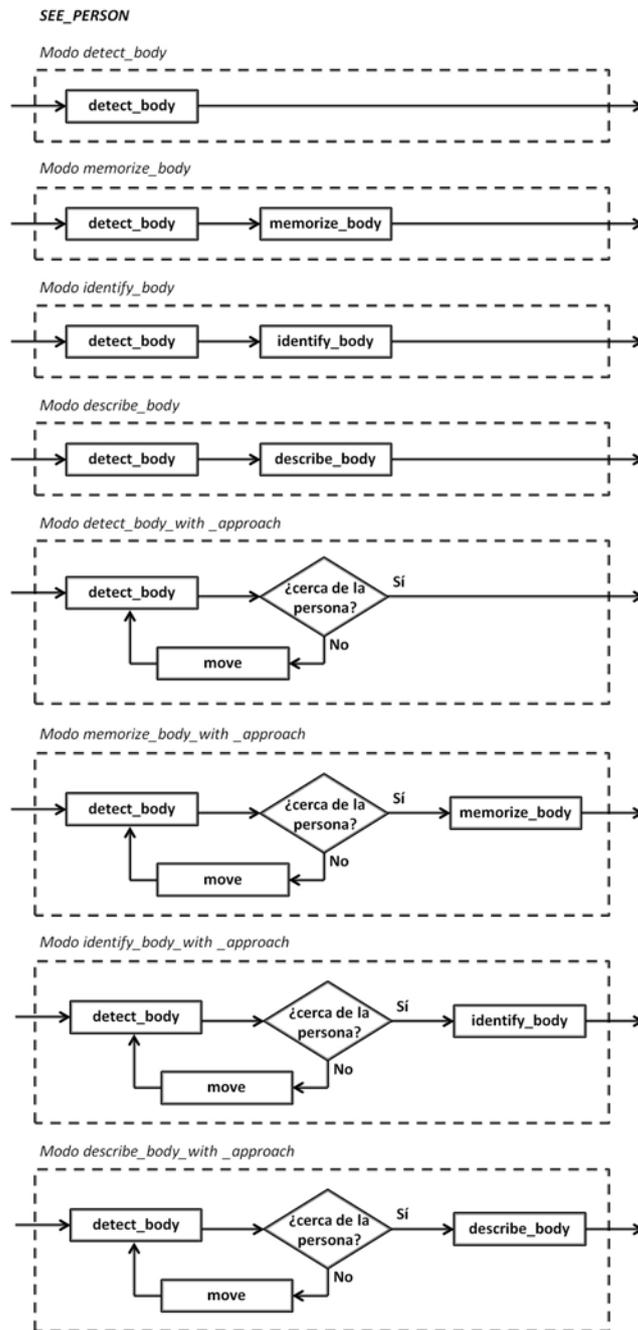


Figura 5.5: Modos basados en biométricos suaves de la conducta *see_person*. Los diagramas de flujo de cada uno de los modos muestran la manera en que se combinan conductas más básicas. En los modos cuyo nombre termina con *with_approach* el robot se aproxima a la persona para tener una mejor visión de ella.

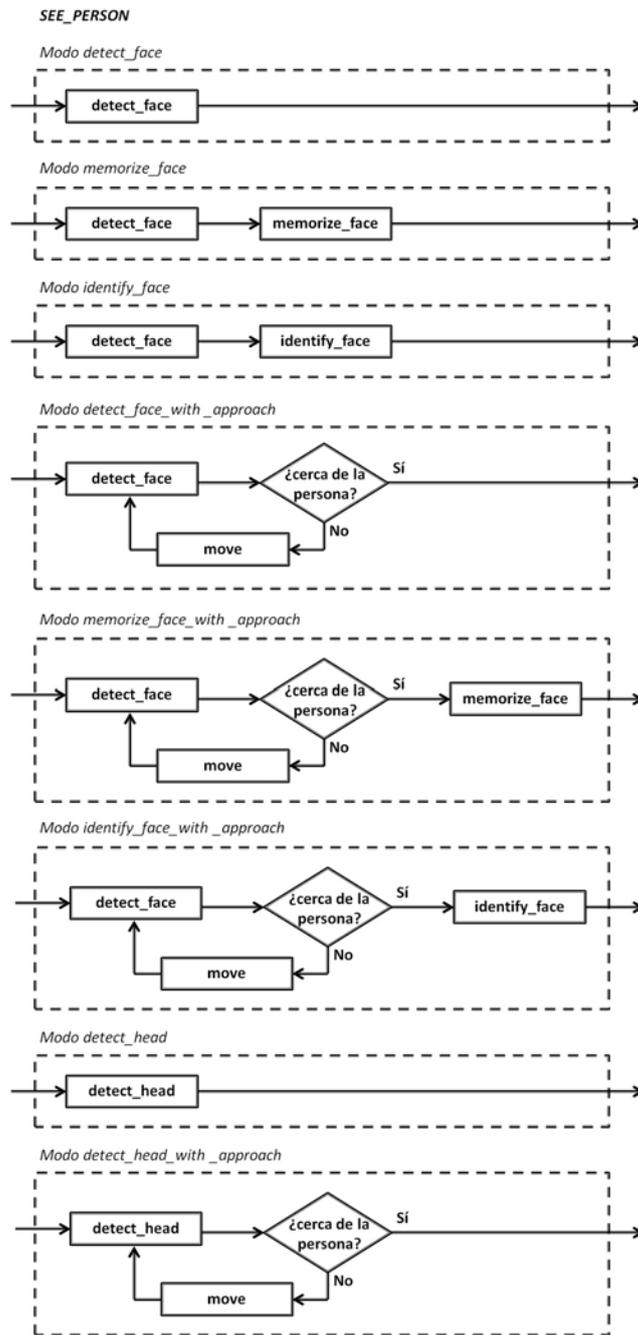


Figura 5.6: Modos basados en biométricos duros de la conducta *see_person*. Los diagramas de flujo de cada uno de los modos muestran la manera en que se combinan conductas más básicas. En los modos cuyo nombre termina con *with_approach* el robot se aproxima a la persona para tener una mejor visión de ella.

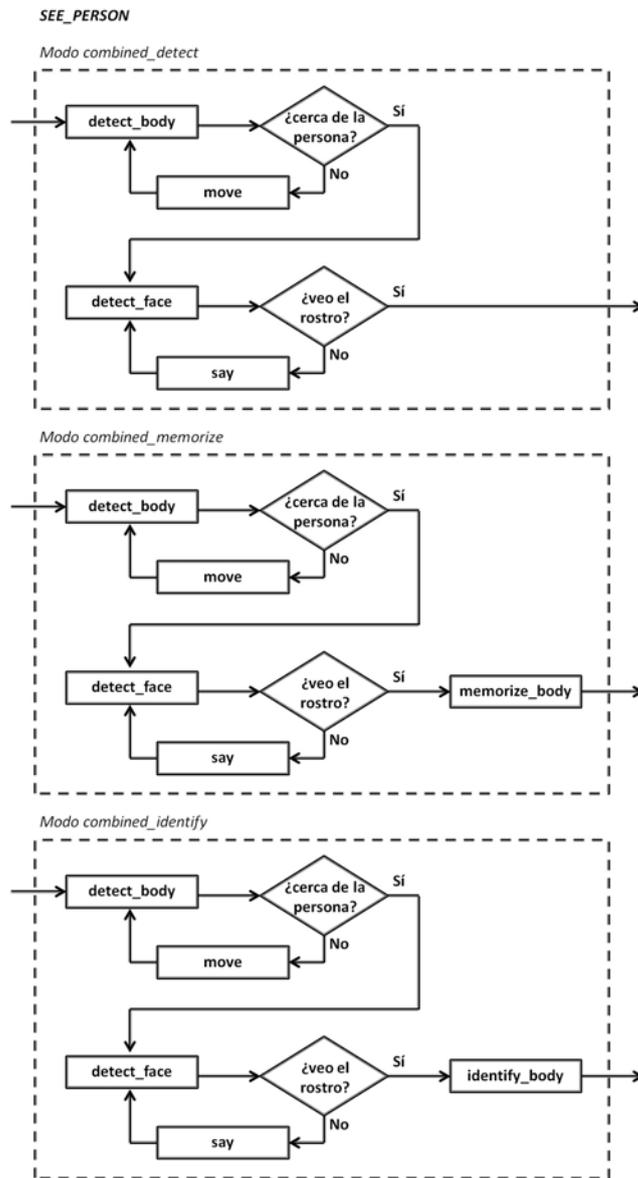


Figura 5.7: Modos basados en la combinación de biométricos suaves y biométricos duros de la conducta *see_person*. Los diagramas de flujo de cada uno de los modos muestran la manera en que se combinan conductas más básicas. En todos los modos el robot se aproxima a la persona para tener una mejor visión de ella y le solicita que lo encare para poder ver su rostro claramente.

5.3. SitLog, Modelos de Diálogo y Arquitectura Cognitiva Orientada a la Interacción

La implementación de las conductas se realizó con el lenguaje de programación SitLog (Pineda *et al.*, 2013b). Este entorno de programación permite “la especificación declarativa de sistemas interactivos complejos con entrada y salida multimodal, incluyendo lenguaje hablado, visión computacional y comportamiento motor” (Pineda *et al.*, 2010:20) y se basa en el uso de Modelos de Diálogo (Pineda, 2008) y de una Arquitectura Cognitiva Orientada a la Interacción (Pineda *et al.*, 2011).

Un Modelo de Diálogo es un protocolo de interacción abstracto de la estructura de una tarea y consiste en un conjunto de situaciones (Pineda *et al.*, 2013a). “Una situación es una abstracción del estado informacional del sistema definido en términos de las expectativas, las acciones que se necesitan realizar cuando una de esas expectativas es satisfecha, y la situación a la que se llega cuando la correspondiente acción se realiza” (Pineda *et al.*, 2010:21-22). Los Modelos de Diálogo se especifican e interpretan mediante SitLog (Pineda *et al.*, 2013a).

El intérprete de SitLog es el componente principal de la Arquitectura Cognitiva Orientada a la Interacción, la cual se ilustra en la Figura 5.8. Esta arquitectura permite relacionar la especificación declarativa de la estructura de la tarea con los algoritmos de acción y percepción y el software del sistema (*Ibidem*).

Como parte de la arquitectura se cuenta con una base de conocimiento en la que se almacena toda la información relevante del escenario en el que trabaja el robot y que es utilizada por SitLog durante la ejecución de una tarea (Pineda *et al.*, 2017b). Dentro de esta base de conocimiento se lleva registro de las personas que intervienen en una tarea particular. Al aprender a una persona se crea un nuevo objeto que comparte el identificador con la BDUBBS del SAIPBBS. De esta manera se mantiene la sincronización entre la información proposicional de la base de conocimiento del robot y las imágenes almacenadas en la BDUBBS.

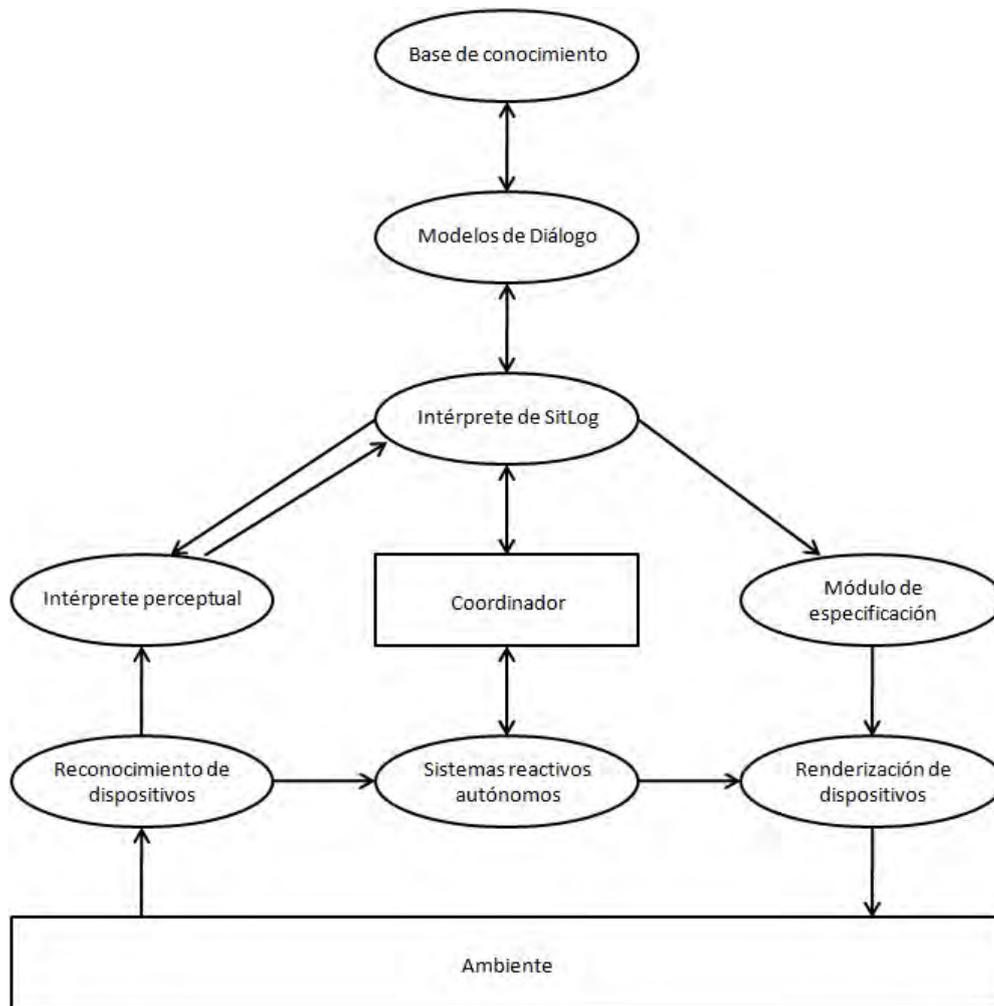


Figura 5.8: Arquitectura Cognitiva Orientada a la Interacción (Pineda *et al.*, 2017b). En el diagrama se muestran los módulos de la arquitectura y la manera en que se relacionan. El ciclo de interacción principal da inicio con el reconocimiento, en el que se transforma la información de entrada del ambiente en una imagen. A continuación se realiza la interpretación perceptual, que asigna significados a las imágenes de acuerdo al contexto, tomando en cuenta las expectativas de la situación actual, mismas que están especificadas en los Modelos de Diálogo. SitLog interpreta la especificación de la estructura de la tarea contenida en los Modelos de Diálogo y relaciona las expectativas y las interpretaciones con las acciones, todas ellas expresadas en un formato proposicional independiente de las modalidades de entrada y salida. Finalmente, se especifican los protocolos de acción y se renderizan en los dispositivos de salida (Pineda *et al.*, 2011).

Cuando se realiza una petición de descripción de una persona al SAIPBBS el resultado se almacena en la base de conocimiento del robot dentro de las propiedades del individuo, las cuales se representan como una lista atributo-valor. Por ejemplo, dentro de la base de conocimiento se puede representar a una instancia de la clase persona de la siguiente manera: [*id* => *arturo*, [*edad* => *adulto*, *genero* => *masculino*], []], cuyo identificador es *arturo* y del que se conocen los valores de las propiedades *edad* y *genero*.

El modelo conceptual presentado en esta sección permite el uso de elementos contextuales en el proceso de identificación de personas. Por ejemplo, si al robot se le da la instrucción “Busca un refresco y entrégaselo a Arturo en la sala”, se puede utilizar el contexto espacial de la tarea para restringir los lugares de búsqueda de la persona. En este ejemplo, la base de conocimiento del robot cuenta con una lista de lugares de búsqueda de personas en cada uno de los cuartos.

A partir de los elementos presentados en este capítulo es posible integrar elementos contextuales, biométricos duros y biométricos suaves en el proceso de identificación de personas. En el siguiente capítulo se presentarán algunos ejemplos de tareas complejas que hacen uso de las conductas de reconocimiento de personas.

5.4. Recapitulación

En este capítulo se explicó la forma en que el Sistema de Aprendizaje e Identificación de Personas Basado en Biométricos Suaves se integró en un robot de servicio. Para ello se utilizó como marco de trabajo el concepto y la estructura funcional de un robot de servicio propuesta por Pineda *et al.* (2013a). Siguiendo este modelo, se definieron las conductas de reconocimiento basadas en biométricos suaves que se integraron a la biblioteca de conductas del robot Golem-III, las cuales pueden ser ensambladas para formar una tarea compleja. El diseño de las conductas atendió a los subproblemas de reconocimiento de personas planteados en el Capítulo 1.

Las conductas de reconocimiento de personas basadas en biométricos suaves se integraron como parte de la jerarquía de conductas de búsqueda, cuya conducta principal es *find*. Esta conducta aprovecha la capacidad del robot para moverse en el ambiente y el uso de las articulaciones de su cuello para realizar una búsqueda de objetos o de personas en el escenario. Además, se propusieron diversas mejoras, incluyendo la aproximación del robot a la

persona de interés para obtener una mejor vista de ella y la combinación de biométricos duros y biométricos suaves como una estrategia de redundancia en el reconocimiento de personas.

Finalmente, se explicaron los tres elementos claves en la instanciación del modelo conceptual de robot de servicio propuesto: SitLog, Modelos de Diálogo y la Arquitectura Cognitiva Orientada a la Interacción. En el siguiente capítulo se presentarán los detalles técnicos del robot Golem-III y se mostrará la forma en que las conductas explicadas en este capítulo se integraron en tareas que puede resolver este robot.

Capítulo 6

Uso de las conductas de aprendizaje e identificación de personas basadas en biométricos suaves en robots de servicio

Las conductas de reconocimiento de personas basadas en biométricos suaves descritas en el capítulo anterior se utilizaron en tareas de robots de servicio. En este capítulo se describe el robot en el que se implementaron las ideas expuestas, las tareas que se propusieron y que requerían el reconocimiento de personas para ser resueltas, los escenarios en las que se llevaron a cabo y los resultados obtenidos.

6.1. Integración en el robot Golem-III

Las conductas de reconocimiento de personas se integraron y probaron en el robot Golem-III (ver Figura 5.9). En esta sección se presenta el hardware y el software que componen al robot (Pineda *et al.*, 2017a).

El robot está compuesto por el siguiente hardware (*Íbidem*):

- Una base robótica *PatrolBotTM*, que cuenta con:
 - Arreglo de 8 sonares.
 - Dos arreglos de 5 bumpers protectivos.

- Altavoces *Infinity*.
- Láser *Sick LMS-500*.
- Computadora interna *Cobra EBX-12*.
- Dos computadoras portátiles *Dell Precision M7510*.
- Telémetro láser *Hokuyo SOKUIKI*.
- Conmutador ethernet *Black Box* de 5 puertos con alimentación USB.
- *Microsoft Kinect 2*.
- Cámara de alta resolución *Point Grey Flea* para puerto USB3.
- Interfaz de audio *8SoundsUSB*.
- Tres micrófonos miniatura.
- Micrófono direccional *RODE VideoMic*.
- Torso, brazos y cuellos construidos por el Grupo Golem.

El software que utilizan los distintos módulos del robot Golem-III se lista a continuación (*Íbidem*):

- Administrador de diálogo.- SitLog, SWI Prolog.
- Base de conocimiento.- SWI Prolog.
- Visión.- MOPED, PCL, Kinect 2 SDK, Microsoft Cognitive Services, OpenCV.
- Audición robótica.- Windows Speech API, JACK.
- Sintetizador de voz.- Windows Speech API.
- Navegación.- ROS Navigation Stack.
- Manipulación de objetos.- Dynamixel RoboPlus.
- Movimiento de cuello y torso.- Dynamixel RoboPlus.
- Comunicación.- ROS.

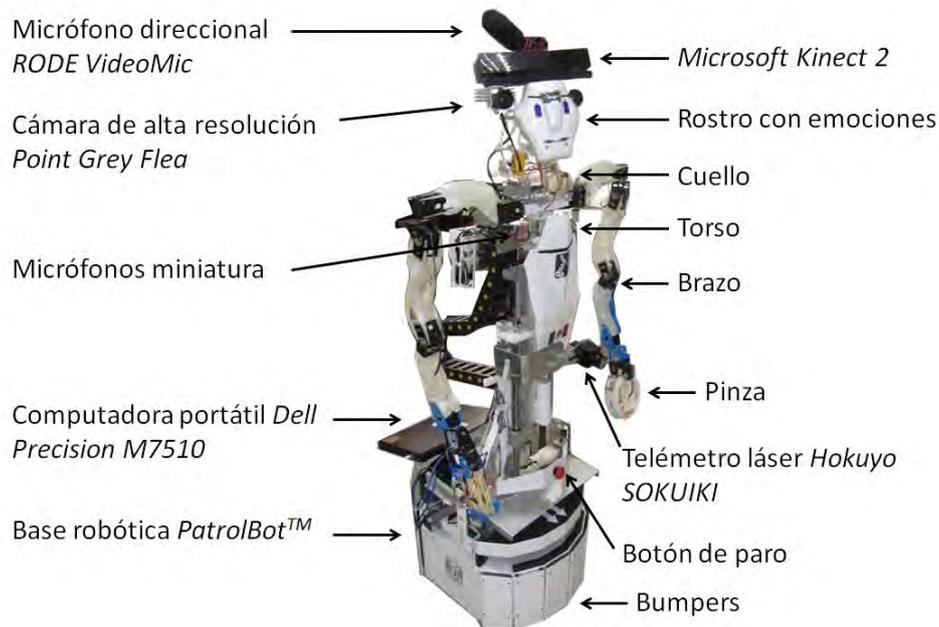


Figura 6.1: El robot de servicio Golem-III (Pineda *et al.*, 2017a)

6.2. Prueba *First & Second*

Esta fue la aplicación piloto que se realizó para evaluar el funcionamiento del sistema de reconocimiento de personas basado en biométricos suaves en el robot Golem-III. Para realizar esta prueba se necesitan dos personas (A y B), y la definición de dos puntos en el escenario (P y Q) separados a una distancia de 1.5 metros. Previo a la ejecución de la prueba se indica a los participantes dónde es el punto en el que se deben colocar cuando el robot les indique y el orden en que deben entrar a la escena. También se les solicita que se posicionen viendo de frente al robot y con los brazos a los costados. Esta prueba consiste de las siguientes fases:

- Inicio.- el robot se encuentra fijo en el punto P viendo hacia el punto Q. El cuello del robot está ajustado de modo que una persona parada en el punto Q quede en el centro del campo visual del robot y pueda ser vista de cuerpo completo.
- Memorización de la primera persona.- el robot solicita que una persona se ponga enfrente de él. La persona A entra y se coloca en el punto Q. El robot memoriza a la persona A y le indica que puede retirarse. La persona A sale del campo visual del robot.

- Memorización de la segunda persona.- se repite el procedimiento del inciso anterior con la persona B.
- Identificación de una persona.- el robot solicita que alguien entre a la escena para identificarlo. En forma aleatoria se decide cuál de las persona regresará a escena. La persona elegida se coloca de nuevo en el punto Q. El robot analiza a la persona y dependiendo del resultado de su proceso de identificación indica “Tú eres la primera persona que memoricé” o “Tú eres la segunda persona que memoricé”.

Un ronda de experimentos consiste en 10 ejecuciones de la prueba en un mismo escenario y con los mismos puntos P y Q. En cada ejecución se utilizan diferentes parejas de personas. Debido a la dificultad de conseguir voluntarios se permitió que una misma persona participara en distintas ejecuciones del experimento siempre y cuando no lo hiciera con la misma pareja.

Las personas que participan son voluntarios que se encuentran en el ambiente donde se realiza el experimento y a quienes se les solicita interrumpir sus actividades para formar parte de los experimentos. En todos los experimentos de este capítulo el perfil de los participantes es el mismo: alumnos de licenciatura, maestría y doctorado, así como investigadores que se encuentran presentes en el lugar del experimento, así como amigos y conocidos de ellos que se encontraban presentes en ese momento y que aceptaron participar en los experimentos. La mayoría de los participantes tienen formación en ciencias o ingeniería de la computación. La mayoría fueron adultos jóvenes y la mayor parte de género masculino. Todos utilizaron vestimenta informal.

Todas las ejecuciones de una ronda se realizaron el mismo día y de manera continua, tratando de lograr la menor diferencia de tiempo entre la primera y la última ejecución, con el objetivo de tener condiciones similares de iluminación y de configuración del ambiente. Se eligieron 10 ejecuciones por ronda debido a que es una cantidad de experimentos que se puede completar con una carga de pilas del robot si se llevan a cabo de manera ágil uno detrás de otro. Cada ronda se hizo en un día diferente, por lo que se permitió que una misma persona participara en más de una ronda, siempre y cuando portara ropa distinta en cada una de ellas.

La conducta usada en esta prueba es *see_person* en los modos *memorize_person* e *identify_person* y el proceso de comparación se realiza en el modo *fino*. En caso de que una ejecución se interrumpa por un fallo que

no esté relacionado directamente con el sistema de detección de personas (por ejemplo, un error generado por un fallo de suministro de energía o un error en el sintetizador de voz), ésta se ignora dentro de la ronda de experimentos y no se contabiliza.

6.2.1. Escenario 1

En el Cuadro 6.1 se muestra la descripción del Escenario 1 (Laboratorio de Golem). En la Figura 6.2 se presenta un ejemplo de una muestra tomada por el robot durante la fase de memorización en este escenario. En la Figura 6.3 se muestran las fases del algoritmo para la localización de parches en esta muestra. En este escenario se llevaron a cabo cinco rondas de experimentos y los números de identificaciones exitosas fueron los siguientes:

- Ronda de experimentos 1.- 9 de 10.
- Ronda de experimentos 2.- 8 de 10.
- Ronda de experimentos 3.- 8 de 10.
- Ronda de experimentos 4.- 10 de 10.
- Ronda de experimentos 5.- 9 de 10.

Un análisis de los fallos en las rondas de experimentos revela que los errores de identificación ocurrieron en parejas cuyo color de vestimenta era muy similar. En la ronda de experimentos más exitosa los colores de la vestimenta de las personas eran claramente distintos, lo que permitió al sistema no tener errores en el proceso de identificación.

Cuadro 6.1: Escenario 1 (Laboratorio de Golem).

Característica	Descripción
Espacio	Cuarto pequeño usado como laboratorio ubicado en el Instituto de Investigaciones en Matemáticas Aplicadas y Sistemas.
Iluminación natural	Los experimentos se realizaron durante el día. La luz del sol entra por ventanas amplias en dos paredes adyacentes del cuarto. Las ventanas cuentan con persianas y cartones blancos recargados para disminuir los efectos de la luz solar.
Iluminación artificial	Se cuenta con lámparas LED en el techo. La iluminación es heterogénea teniendo cambios drásticos de intensidad luminosa dependiendo de la zona del cuarto.
Presencia de otras personas	Poca presencia de otras personas. Se tiene control para solicitar que despejen el área de experimentación.



Figura 6.2: Escenario 1 (Laboratorio de Golem)

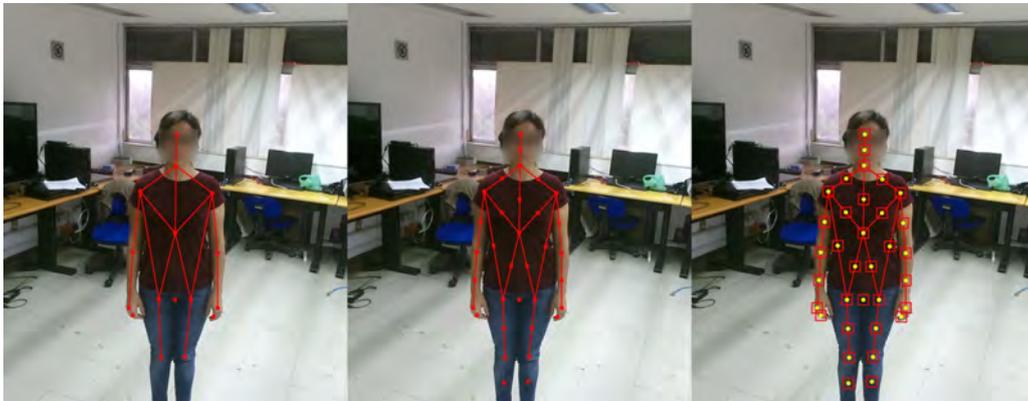


Figura 6.3: Fases de localización de parches en el Escenario 1. A la izquierda se muestra el esqueleto de la persona. En el centro se indica la localización de los puntos de interés para extraer parches. A la derecha se presentan los recuadros que delimitan los parches de color.

6.2.2. Escenario 2

En el Cuadro 6.2 se muestra la descripción del Escenario 2 (Explicada del museo Universum). En la Figura 6.4 se presenta un ejemplo de una muestra tomada por el robot durante la fase de memorización en este escenario. En la Figura 6.5 se muestran las fases del algoritmo para la localización de parches en esta muestra. En este escenario se llevaron a cabo dos rondas de experimentos y los números de identificaciones exitosas fueron los siguientes:

- Ronda de experimentos 1.- 7 de 10.
- Ronda de experimentos 2.- 9 de 10.

Cuadro 6.2: Escenario 2 (Explanada del museo Universum).

Característica	Descripción
Espacio	Patio interior del museo de ciencias Universum. Se realizó una instalación para simular una casa, que incluye unos muros bajos para delimitar los cuartos de la casa y un número pequeño de muebles en cada uno de ellos. Fuera de la casa se puede ver el patio del museo por donde transitan los visitantes y un espacio de trabajo con mesas y personas trabajando.
Iluminación natural	Los experimentos se realizaron durante el día. La luz del sol entra de manera directa por las ventanas en el techo del museo.
Iluminación artificial	Sin iluminación artificial durante los experimentos.
Presencia de otras personas	Presencia moderada de otras personas en la cercanía del robot. Alta presencia de personas fuera del escenario de casa (incluyendo visitantes del museo y personas laborando en las mesas de trabajo). Se tiene control para solicitar que despejen el área dentro de la casa, pero ningún control sobre personas fuera de la casa.



Figura 6.4: Escenario 2 (Explanada del museo Universum)



Figura 6.5: Fases de localización de parches en el Escenario 2. A la izquierda se muestra el esqueleto de la persona. En el centro se indica la localización de los puntos de interés para extraer parches. A la derecha se presentan los recuadros que delimitan los parches de color.

6.3. Prueba *Person Recognition Test 2016*

La segunda prueba que se implementó fue parte de la competencia *RoboCup@Home 2016* (Van Beek *et al.*, 2016). El robot inicia en una posición determinada por un juez y memoriza a un usuario. El robot debe esperar

un momento mientras el usuario rodea al robot y se coloca detrás de él dentro de una multitud (entre 5 y 10 personas dispersas en un área de 5 metros a una distancia de 2 o 3 metros del robot). El robot se da la vuelta e inicia la búsqueda de la persona. Al encontrar a la persona el robot debe acercarse a ella y apuntarla con su brazo. Para esta prueba se utilizó directamente la conducta *find* con los modos *memorize_person* e *identify_person_with_approach* y el proceso de comparación se realiza en el modo fino.

Una ronda de experimentos consiste en 10 ejecuciones de la prueba en un mismo escenario con la misma posición inicial del robot. Se siguieron las mismas consideraciones de selección de participantes que se usaron en la Prueba *First & Second*. Se permitió la participación de una misma persona en diferentes ejecuciones de la ronda ya sea como sujeto de aprendizaje del robot o parte de la multitud, pero se verificó que no hubieran ejecuciones en una misma ronda en las que se repitieran de manera simultánea la persona aprendida y los miembros de la multitud. Las rondas se llevaron a cabo en días diferentes, y se permitió la participación de una misma persona en distintas rondas siempre y cuando portara vestimenta diferente en cada ronda. También se siguió la política de ignorar las ejecuciones fallidas por una causa ajena al sistema de reconocimiento de personas.

6.3.1. Escenario 1

Se realizaron 3 rondas de experimentos en el Escenario 1 (Laboratorio de Golem) con multitudes de 5 personas. Los números de identificaciones exitosas fueron los siguientes:

- Ronda de experimentos 1.- 4 de 10.
- Ronda de experimentos 2.- 2 de 10.
- Ronda de experimentos 3.- 5 de 10.

En este escenario el número de ejecuciones exitosas fue bajo. Un análisis de los fallos de estos experimentos muestra una debilidad de la estrategia ante las condiciones de iluminación. Por ejemplo, si el usuario fue aprendido en una zona bien iluminada y se coloca después en una zona oscura del escenario, el robot no logra identificarlo.

6.3.2. Escenario 2

Se realizó una ronda de experimentos en el Escenario 2 (Explanada del museo Universum) con multitudes de 5 personas. El número de identificaciones exitosas fue 4 de 10, un resultado bajo de manera similar al obtenido en el Escenario 1. De nueva cuenta la presencia de luz natural provocó errores en el proceso de identificación.

6.3.3. Escenario 3

En el Cuadro 6.3 se muestra la descripción del Escenario 3 (Zona de trabajo del Torneo Mexicano de Robótica 2016). En la Figura 6.6 se presenta un ejemplo de una muestra tomada por el robot durante la fase de memorización en este escenario. En la Figura 6.7 se muestran las fases del algoritmo para la localización de parches en esta muestra. Se realizó una ronda de experimentos en este escenario con multitudes de 5 personas. El número de identificaciones exitosas fue 8 de 10. Las condiciones de iluminación artificial del escenario tuvieron un alto impacto en el buen desempeño del sistema.

Cuadro 6.3: Escenario 3 (Zona de trabajo del Torneo Mexicano de Robótica 2016).

Característica	Descripción
Espacio	Zona de trabajo de participantes en un centro de exposiciones.
Iluminación natural	El espacio es cerrado por lo que no hay presencia significativa de luz solar.
Iluminación artificial	Iluminación homogénea en toda la zona de experimentación. La luz es tenue y en combinación con los colores del ambiente genera un tono opaco en las imágenes captadas por la cámara.
Presencia de otras personas	Poca presencia de otras personas en la zona elegida para el experimento. Se tiene control para solicitar que despejen el área de experimentación.



Figura 6.6: Escenario 3 (Zona de trabajo del Torneo Mexicano de Robótica 2016)



Figura 6.7: Fases de localización de parches en el Escenario 3. A la izquierda se muestra el esqueleto de la persona. En el centro se indica la localización de los puntos de interés para extraer parches. A la derecha se presentan los recuadros que delimitan los parches de color.

6.3.4. Escenario 4

En el Cuadro 6.4 se muestra la descripción del Escenario 4 (Zona de trabajo de *RoboCup* 2016). En la Figura 6.8 se presenta un ejemplo de una muestra tomada por el robot durante la fase de memorización en este

escenario. En la Figura 6.9 se muestran las fases del algoritmo para la localización de parches en esta muestra. Se realizó una ronda de experimentos en este escenario con multitudes de 5 personas. El número de identificaciones exitosas fue 9 de 10. Las excelentes condiciones de iluminación artificial del escenario tuvieron un alto impacto en el buen desempeño del sistema.

Cuadro 6.4: Escenario 4 (Zona de trabajo de *RoboCup@Home* 2016).

Característica	Descripción
Espacio	Zona de trabajo de participantes en un centro de exposiciones.
Iluminación natural	El espacio es cerrado por lo que no hay presencia significativa de luz solar.
Iluminación artificial	Iluminación homogénea en toda la zona de experimentación. La luz es fuerte y en combinación con los colores del ambiente genera un tono blanco brillante en las imágenes captadas por la cámara.
Presencia de otras personas	Presencia moderada de otras personas en la zona elegida para el experimento. No se tiene control para solicitar que despejen el área de experimentación.

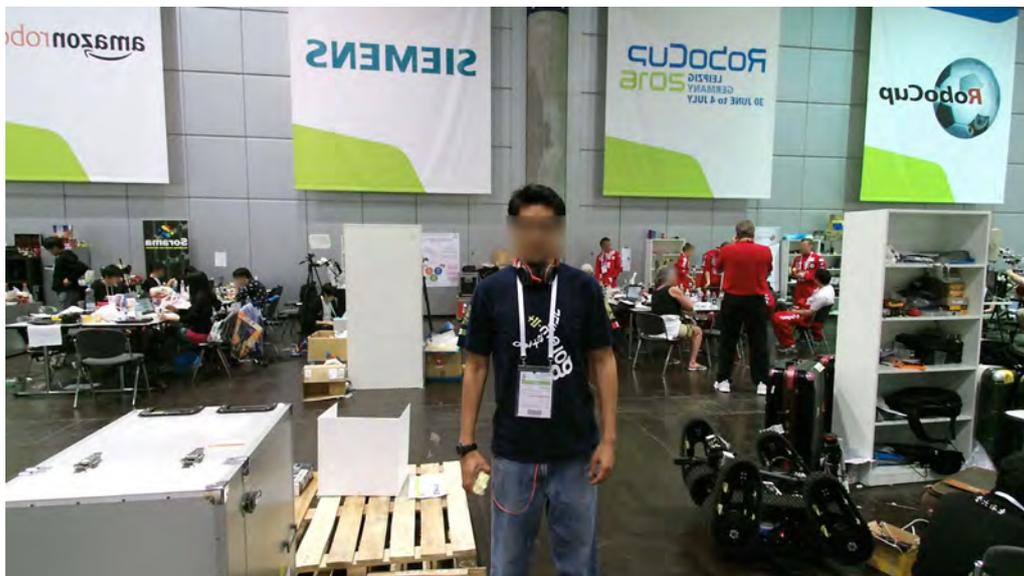


Figura 6.8: Escenario 4 (Zona de trabajo de *RoboCup@Home* 2016)



Figura 6.9: Fases de localización de parches en el Escenario 4. A la izquierda se muestra el esqueleto de la persona. En el centro se indica la localización de los puntos de interés para extraer parches. A la derecha se presentan los recuadros que delimitan los parches de color.

6.3.5. Ejecuciones de la prueba en competencias

En esta sección se explicará de manera anecdótica los resultados de demostrar el funcionamiento de la prueba frente a una audiencia en vivo y en condiciones de evaluación por parte de jueces en el contexto de la competencia *RoboCup@Home*.¹

Esta prueba se ejecutó en el Torneo Mexicano de Robótica 2016. El robot Golem-III logró identificar correctamente a la persona en la multitud. Sin embargo, al tratar de aproximarse a ella el robot detectó que había un obstáculo (otra persona sentada en una silla) e intentó buscar una alternativa para llegar al lugar que quería, lo que le costó tiempo de la prueba. Finalmente decidió ignorar el obstáculo y apuntó desde lejos de manera correcta, dejando claro que había identificado a la persona correcta. Sin embargo, el tiempo de la prueba ya había concluido y el robot no recibió puntos.

La prueba también se ejecutó en *RoboCup 2016*. En el primer intento Golem-III logró identificar, aproximarse y apuntar a la persona dentro del tiempo de la prueba. En esta competencia los puntos no se asignaron al robot debido a que el reporte de resultados estaba en formato .jpg y no en formato .pdf como lo solicitaba el libro de reglas. En la Figura 6.10 se presenta la imagen de la persona memorizada en esta ejecución. En la Figura

¹<http://www.robocupathome.org/>

6.11 se muestran las fases del algoritmo para la localización de parches en esta muestra. En la Figura 6.12 se muestra a la persona identificada en la multitud.

En esta prueba hubo un segundo intento y de nuevo se identificó a la persona correcta, pero ocurrió un error similar al del Torneo Mexicano de Robótica y Golem-III no se pudo aproximar e invirtió tiempo en resolver el problema de navegación. Debido a que se terminó el tiempo, se tuvo que interrumpir la ejecución y el robot no pudo apuntar a nadie en la multitud.



Figura 6.10: Memorización de la persona en la competencia *RoboCup@Home* 2016



Figura 6.11: Fases de localización de parches en la competencia *RoboCup@Home* 2016. A la izquierda se muestra el esqueleto de la persona. En el centro se indica la localización de los puntos de interés para extraer parches. A la derecha se presentan los recuadros que delimitan los parches de color.



Figura 6.12: Identificación de la persona en la competencia *RoboCup@Home* 2016

6.4. Prueba *General Purpose Service Robot 2015*

En esta prueba el robot espera en una posición específica del escenario a que se le dé una instrucción para resolverla. Aunque el repertorio de posibles comandos que se le pueden dar al robot es muy amplio se eligió uno que implica la identificación de personas. El comando que se le da es “*Take this object and bring it to P in the R*”, donde *P* es el nombre de una persona y *R* es el nombre de un cuarto. Como parte de los preparativos de esta prueba, el robot memoriza a un conjunto de usuarios. Además, se define un conjunto de cuartos en la base de conocimiento.

6.4.1. Escenario 1, Versión 1

Se realizaron 2 rondas de experimentos en el Escenario 1 (Laboratorio de Golem). En cada ejecución se definió de manera aleatoria el cuarto en el que el robot debía buscar a la persona. Además, se colocó a una segunda persona para verificar que el robot identificaba a la correcta, aunque la descripción oficial de la prueba no lo señala. Para esta prueba se utilizó directamente la conducta *find* con los modos *memorize_person* e *identify_person_with_approach* y el proceso de comparación se realiza en el modo fino.

Los números de identificaciones exitosas fueron los siguientes:

- Ronda de experimentos 1.- 8 de 10.
- Ronda de experimentos 2.- 6 de 10.

6.4.2. Escenario 1, Versión 2

Es importante mencionar que en la competencia *RoboCup@Home 2016* la forma en que se llevó a cabo esta prueba fue distinta, pues no se incluyó una fase en la que los robots pudieran memorizar a las personas que participarían en la prueba. En vez de ello, se asumió que el robot no conocía de antemano a la persona y simplemente tenía que detectar a alguien en ese cuarto. Se asumía que sólo había una persona en el cuarto. De esta manera, la identificación de la persona se puede realizar únicamente por elementos contextuales. Se elaboró una segunda versión de este experimento y se utilizó la conducta *find* en el modo *detect_person_with_approach*, eliminando la fase de memorización. Se realizó una ronda de 10 experimentos y en

9 de ellos se tuvo éxito en la identificación de la persona por elementos contextuales.

6.5. Recapitulación

En este capítulo se describió el robot de servicio Golem-III, en el cual se integraron las conductas de reconocimiento de personas que utilizan el Sistema de Aprendizaje e Identificación de Personas Basado en Biométricos Suaves. Además, se propusieron un conjunto de aplicaciones para robots de servicio en las que se realiza reconocimiento de personas. La tarea *First & Second* se propuso como experimento piloto, mientras que las tareas *Person Recognition Test 2016* y *General Purpose Service Robot (GPSR) 2015* se tomaron de los libros de reglas de la competencia *RoboCup@Home*.

En el Cuadro 6.5 se presenta un resumen de los experimentos realizados. Las causas de los experimentos que fallaron son diversas. En los escenarios con iluminación natural (Laboratorio de Golem y Explanada del Museo Universum) la tarea *Person Recognition Test 2016* se dificultó debido a que la iluminación en la zona donde se aprendía a la persona y la zona donde se identificaba eran distintas. Este fallo no ocurrió en la tarea *First & Second* a pesar de probarse en los mismos escenarios, y esto se debe a que el aprendizaje y la identificación ocurren en la misma región del escenario. La tarea *Person Recognition Test 2016* mostró mejores resultados en los escenarios con iluminación artificial (Torneo Mexicano de Robótica y *RoboCup@Home 2016*).

Cuadro 6.5: Resumen de los experimentos realizados en el Capítulo 6.

Prueba	Escenario	Ronda	Resultados
<i>First & Second</i>	1) Laboratorio de Golem	1	9 de 10
		2	8 de 10
		3	8 de 10
		4	10 de 10
		5	9 de 10
	2) Explanada del museo Universum	1	7 de 10
	2	9 de 10	
<i>Person Recognition Test 2016</i>	1) Laboratorio de Golem	1	4 de 10
		2	2 de 10
		3	5 de 10
	2) Explanada del museo Universum	1	4 de 10
	3) Torneo Mexicano de Robótica	1	8 de 10
4) RoboCup 2016	1	9 de 10	
<i>GPSR 2015 Versión 1</i>	1) Laboratorio de Golem	1	8 de 10
		2	6 de 10
<i>GPSR 2015 Versión 2</i>	1) Laboratorio de Golem	1	9 de 10

Uno de los fallos frecuentes en los experimentos se relaciona con una detección incorrecta del esqueleto por parte del SDK de Microsoft para Kinect 2.0. Este fallo provoca que la localización de las articulaciones en el esqueleto no correspondan con la postura real de la persona. Esto causa que el algoritmo de identificación falle, pues se extraen parches de color de regiones arbitrarias de la imagen.

Otro de los retos del sistema propuesto ocurrió en los escenarios del Torneo Mexicano de Robótica y *RoboCup@Home 2016*. En estas competencias los participantes suelen utilizar uniformes de acuerdo al equipo al que pertenecen. Aunque en los experimentos realizados se evitó seleccionar personas que trajeran uniforme, en algunos de los casos fue inevitable. Algunos de los fallos ocurrieron cuando el robot confundió a personas que usaban una playera similar. La detección de estas fuentes de error es útil para mejoras a futuro del algoritmo. Esto se incluye como parte de las conclusiones que se presentan en el siguiente capítulo.

En la página <https://www.youtube.com/golemiimas>, en la lista de reproducción *Soft Biometrics Person Recognition*, se pueden ver videos del funcionamiento del robot utilizando algunas de las conductas de este trabajo de investigación.

Capítulo 7

Conclusiones

En este último capítulo se presentan las conclusiones a las que se llegaron como resultado del presente trabajo de investigación. Primero se presentará una recapitulación del trabajo realizado en cada uno de los capítulos, destacando las aportaciones que se hizo en cada uno de ellos. Posteriormente se presentará un resumen del método de reconocimiento de personas propuesto y el resultado de los experimentos. Finalmente se presentarán ideas para trabajo futuro.

7.1. Recapitulación y aportaciones

En el Capítulo 1 se plantearon los objetivos generales y específicos, los cuales se cumplieron de manera exitosa. Además, se hizo un análisis del problema de reconocimiento de personas y se identificaron los subproblemas principales que la componen, que sirvieron como conceptos básicos en el análisis teórico y en la implementación práctica de los capítulos posteriores.

La revisión de la literatura sobre reconocimiento de personas realizada en el Capítulo 2 permitió distinguir de manera clara tres tipos de elementos para abordar los subproblemas de aprendizaje e identificación: los rasgos biométricos duros, los rasgos biométricos suaves y los elementos contextuales. Como resultado del análisis del estado del arte en biométricos suaves se concluyó que debido a que es un área de investigación de reciente creación todavía no existe una estandarización en los conceptos básicos. Se identificaron distintas definiciones de biométricos suaves no compatibles entre ellas y se estableció una manera de organizarlas (Tabla 2.3). Además, se propuso una definición de biométricos suaves como terna objeto-atributo-valor que ayudó a facilitar su entendimiento y que se usó a lo largo del trabajo

de investigación. También se identificaron los tipos de elementos contextuales que se pueden usar para la identificación de personas. Finalmente, se realizó una discusión de cómo los distintos elementos de identificación se pueden usar en robots de servicio y se llegó a la conclusión de que cada uno de ellos presenta ventajas y desventajas en función del tipo de tarea, por lo que sería ideal que un robot contara con la capacidad de utilizarlos todos.

En el Capítulo 3 se hizo una revisión extensa de la literatura de biométricos suaves para identificar los algoritmos que se utilizan en su detección, localización, aprendizaje e identificación. Esta revisión permitió concluir que la mayoría de los experimentos realizados se hacen a partir de muestras fotográficas con un proceso de recolección fuera de línea controlado por humanos. Se encontró una gran cantidad de algoritmos y estrategias y se eligieron algunas de ellas para realizar este trabajo de investigación, en particular la detección y localización de rasgos basada en contexto y el uso de histogramas de color para el aprendizaje y la identificación.

Las estrategias seleccionadas se utilizaron para diseñar un sistema de aprendizaje en línea e identificación de personas basado en biométricos suaves. En el Capítulo 4 se presentó la arquitectura de este sistema y se explicaron los tres algoritmos diseñados para realizar el proceso de comparación de personas: comparación gruesa, comparación fina y comparación a partir de la descripción.

El siguiente paso fue la integración de este sistema en un robot de servicio. En el Capítulo 5 se explicó el marco de trabajo que se basa en tres nociones principales: Modelos de Diálogo, Arquitectura Cognitiva Orientada a la Interacción y Sitlog. Se diseñaron las conductas de reconocimiento de personas y se integraron a la biblioteca de conductas jerárquica y composicional del robot Golem-III. Además, se diseñó una estrategia que combina la identificación por biométricos duros, la identificación por biométricos suaves, el uso de elementos contextuales a partir de la especificación de la tarea y el contenido de la base de conocimiento, así como las conductas motoras y verbales del robot.

Se diseñaron tareas que utilizan las conductas de reconocimiento de personas a partir de biométricos suaves y se evaluó su funcionamiento. La tarea *First & Second* permitió evaluar el sistema cuando el robot se encuentra en reposo. Debido a que el proceso de memorización e identificación se realiza en la misma escena se eliminan muchas de las variables que podrían afectar

su funcionamiento (en particular, el cambio de iluminación). El desempeño de esta prueba fue exitoso.

La segunda tarea fue *Person Recognition Test*, tomado de la competencia *RoboCup@Home* 2016. En esta prueba el robot tiene que moverse en un escenario dinámico, aumentando el número de variables que pueden afectar el desempeño del sistema. Se realizaron pruebas en 4 escenarios diferentes y se concluyó que el tipo de iluminación tiene un impacto alto en los resultados. En particular, la luz natural afecta el desempeño del sistema, mientras que una iluminación artificial homogénea facilita obtener buenos resultados. Esta prueba se ejecutó en vivo en competencias de robots de servicio y se logró realizar de manera exitosa el proceso de identificación, aunque no se obtuvieron los puntajes por cuestiones no relacionadas con el reconocimiento de personas.

La tercera tarea fue *General Purpose Service Robot*, tomado de la competencia *RoboCup@Home* 2015. Se eligió un comando en el que se necesitaba la identificación de una persona para resolver una tarea. En esta prueba, los elementos contextuales jugaron un papel importante en el éxito del robot para resolverla.

7.2. Resumen del algoritmo de aprendizaje e identificación basado en biométricos suaves y experimentos

Al realizar una solicitud de aprendizaje al sistema se especifica cuál de las personas detectadas en la escena debe ser memorizada y se brinda una etiqueta para ser usada como identificador. La observación se almacena en la base de datos de usuarios indexada por la etiqueta y el ángulo de orientación del usuario con respecto al dispositivo sensor. De esta manera es posible almacenar diferentes muestras del mismo usuario.

Cuando se le solicita al sistema realizar la identificación de una persona en la escena se determina su orientación con respecto al dispositivo sensor como primer paso, y se compara con las observaciones almacenadas en la base de datos. Para cada persona en la base de datos se selecciona la observación más cercana en términos de su orientación y se determina la similitud entre la persona observada y la persona en la base de datos. El proceso de identificación tiene los modos grueso, fino y basado en descripción,

que difieren únicamente en la función de similitud que utilizan. Al finalizar el proceso, el sistema escoge al usuario más similar en la base de datos. Si la mejor similaridad es mayor que un umbral predeterminado el sistema regresa el identificador de la persona seleccionada en la base de datos, de lo contrario indica que se trata de una persona desconocida.

En el modo de comparación gruesa se utiliza la información del esqueleto obtenida en la fase de detección para delimitar la región del torso de las personas a comparar. Se extraen los segmentos de imagen de esas regiones y se calculan sus respectivos histogramas RGB. El cálculo de la correlación entre ambos sirve como medida de similitud.

En el modo de comparación fina de dos personas se utiliza la información del esqueleto para elegir regiones de interés del cuerpo de las cuales se extraen parches de color. Los parches obtenidos de la misma región del cuerpo de ambas personas se comparan mediante la correlación de sus histogramas RGB. Este cálculo sirve para determinar si la pareja de parches es suficientemente parecida para emitir un voto que contribuye a la función de similitud.

En el modo de identificación a partir de la descripción se obtienen los vectores de descripción de las personas que se desean comparar. Se aplica una función de categorización a algunos de los elementos que se utilizarán para realizar la comparación. Si ambas descripciones tienen un atributo con la misma categoría se otorga un voto. El valor de similitud es el número total de votos obtenidos. Este modo utiliza el género, la edad, la estatura, el color de la ropa y la presencia de vello facial y lentes para la comparación.

Se procedió con una serie de experimentos para evaluar el funcionamiento del sistema integrado en un robot de servicio, utilizando el modo de identificación fino. Se propusieron aplicaciones para robots de servicio en las que se realiza reconocimiento de personas. Se inventó la tarea *First & Second*, y se utilizaron las tareas *Person Recognition Test 2016* y *General Purpose Service Robot (GPSR) 2015* de la competencia *RoboCup@Home*. Además, se propició la evaluación de la ejecución de las tareas en escenarios diferentes. En particular, se utilizaron dos escenarios con iluminación artificial y dos escenarios con iluminación natural. El sistema mostró un mejor desempeño en escenarios con iluminación artificial. Los fallos fueron causados principalmente por variaciones significativas de iluminación entre las zonas de aprendizaje e identificación de las personas, errores en la generación automática del esqueleto por parte del detector de personas y presencia de

vestimentas similares entre los participantes.

7.3. Trabajo futuro

El problema de aprendizaje en línea e identificación de personas en ambientes dinámicos para robots de servicio es demasiado complejo, y aunque en este trabajo de investigación se lograron buenos resultados en algunos escenarios y tareas particulares, todavía existe un largo camino para lograr un buen desempeño en escenarios y tareas con mayor nivel de dificultad. Este trabajo de investigación se puede continuar y profundizar en muchos frentes, incluyendo la mejora de los algoritmos de reconocimiento, el diseño de conductas de reconocimiento más robustas, el uso adecuado de la redundancia en los procesos de aprendizaje e identificación, el diseño de mecanismos selectivos de rasgos de identificación dependiendo de las circunstancias específicas de la tarea, el diseño de procesos más eficientes para el almacenamiento y la extracción de información en la base de datos de personas, la construcción de modelos tridimensionales de personas más ligeros y muchos otros problemas de investigación.

A corto plazo se pueden abordar los siguientes problemas en el algoritmo de identificación basado en biométricos suaves: sustituir el algoritmo de detección de esqueleto por uno más reciente y robusto, en particular uno basado en aprendizaje profundo; intentar resolver la fragilidad que tiene el sistema ante las variaciones de iluminación probando el funcionamiento en otro espacio de color; aumentar el número de características del vector de descripción en el modo de identificación a partir de la descripción, incluyendo no sólo el color de la playera, sino de otras regiones como pantalón, zapatos, mangas, etc., con el objetivo de tener una descripción más detallada de la persona y tratar de resolver situaciones en que las personas usan prendas similares.

Apéndice A

Clasificación de pose de brazos y cuerpo

En el Capítulo 4 se presentó el Sistema de Estimación de Orientación y Pose (SEOP), que entre sus funciones incluye determinar la pose de brazos y de cuerpo de una persona a partir de su esqueleto. Como se explicó en ese capítulo, el esqueleto utilizado cuenta con 25 vértices que corresponden a puntos de articulación y extremidades importantes del cuerpo humano. Estos vértices se presentaron en la Figura 4.3. A partir de este esqueleto se creó un módulo basado en reglas que permite clasificar la pose de los brazos y del cuerpo de un usuario mediante geometría analítica básica.

La pose de los brazos de una persona se clasifica en: i) Brazos a los costados, ii) Apuntando, iii) Un brazo levantado, iv) Ambos brazos levantados. Estas cuatro poses se ilustran en la Figura A.1. Uno de los criterios básicos para la clasificación fue la altura de la mano a_m con respecto a la altura del hombro a_h que le corresponde (se compara mano izquierda con hombro izquierdo y mano derecha con hombro derecho). El margen de tolerancia η permite delimitar el intervalo de alturas $(a_h - \eta, a_h + \eta)$ cercanas a la altura del hombro y permite definir tres casos:

- Si $a_m \leq a_h - \eta$ la mano se encuentra por debajo del hombro.
- Si $a_h - \eta < a_m < a_h + \eta$ la mano se encuentra a la altura del hombro.
- Si $a_m \geq a_h + \eta$ la mano se encuentra por encima del hombro.

En el caso de la pose de apuntar se estableció como requisito que el brazo se encuentre extendido al máximo de manera rígida y apuntando hacia el horizonte. Se calcula la pendiente m de la recta que pasa por el hombro y su

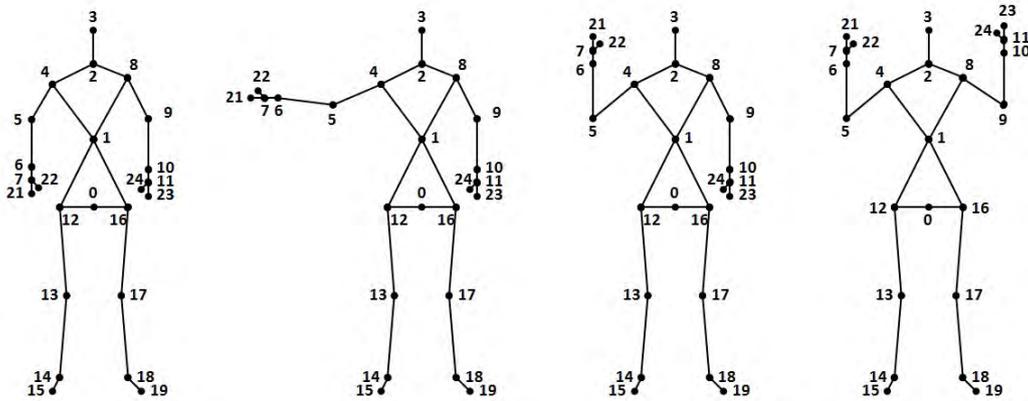


Figura A.1: Posiciones de brazos de una persona. De izquierda a derecha: i) Brazos a los costados, ii) Apuntando, iii) Un brazo levantado, iv) Ambos brazos levantados.

respectivo codo. En un caso ideal se esperaría que al apuntar al horizonte se formara un ángulo de 90° entre el brazo y el tronco, de modo que $m = 0$. También se establece un margen de tolerancia ζ de modo que si $-\zeta < m < \zeta$ se considera que el brazo está apuntando al horizonte. Se elige el codo en vez de la mano debido a que si la persona se encuentra muy cerca del sensor de profundidad su mano puede quedar fuera del rango de visión al apuntar al horizonte.

Para calcular el ángulo al que está apuntando se utilizan las proyecciones sobre el plano XZ de los puntos que corresponden al hombro y al codo del brazo con el que apunta (ver el sistema de referencia en la Figura 4.2). Se aplica una transformación de traslación a ambos puntos de modo que el hombro quede en el origen. Finalmente se calcula el ángulo con respecto al eje X que forma la línea entre ambos.

La pose del cuerpo de una persona se clasifica en: i) De pie, ii) Sentado, iii) Recostado en el piso. Estas tres poses se ilustran en la Figura A.2. El primer criterio para clasificar las poses fue la altura de la cabeza h_c . Se establecieron los límites de altura κ y λ , con $\kappa < \lambda$, que permiten definir tres casos:

- Si $h_c \leq \kappa$ es posible que la persona se encuentre recostada en el piso.
- Si $\kappa < h_c < \lambda$ es posible que la persona se encuentre sentada.
- Si $h_c \geq \lambda$ es posible que la persona se encuentre de pie.

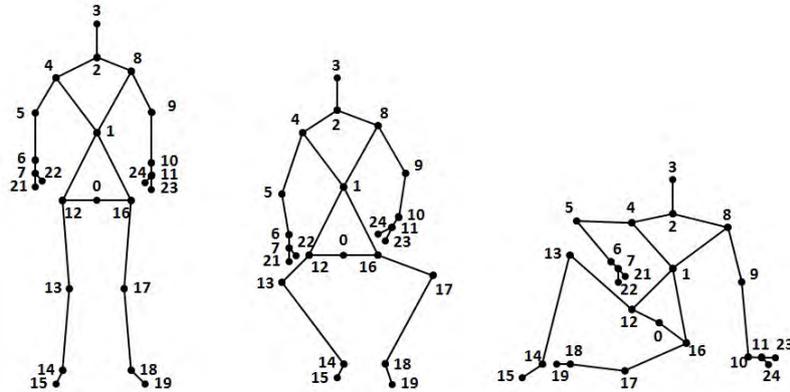


Figura A.2: Posiciones de cuerpo de una persona. De izquierda a derecha: i) De pie, ii) Sentado, iii) Recostado en el piso.

En general, los esqueletos que se obtienen de personas recostadas en el piso son confusos y la mayoría de los puntos que lo conforman no son detectados, por lo que el criterio anterior resulta útil y sencillo para detectar esta pose. Sin embargo, el criterio no es suficiente para separar de manera clara a las personas sentadas de las persona paradas. Por ejemplo, un niño de pie podría ser clasificado como una persona sentada. Por ello se introduce el criterio de comparar la altura de la rodilla h_r con la altura de la cadera h_x . Se establece un margen de tolerancia ξ , de modo que si $h_x - \xi < h_r < h_x + \xi$ se considera que la rodilla está aproximadamente a la altura de la cadera y por lo tanto es posible que la persona se encuentre sentada.

Referencias

Akhter, Arsalan; Mehmood, Tahir; Aslam, Haseeb; Akram, Basit; Syed Osama, Ahmad; Hakeem, Hamza; Ajmal, Ammar. (2016). "Machinilog 2016 Team Description Paper". En *Proceedings of RoboCup 2016*. 8 pp.

Angelova, Anelia; Krizhevsky, Alex; Vanhoucke, Vincent; Ogale, Abhijit y Ferguson, Dave. (2015). "Real-Time Pedestrian Detection with Deep Network Cascades". En *Proceedings of British Machine Vision Conference 2015*, Vol. 2, 12 pp.

Anguelov, Dragomir; Lee, Kuang-chih; Göktürk, Salih Burak y Sumengen, Baris. (2007). "Contextual Identity Recognition in Personal Photo Albums". En *IEEE Conference on Computer Vision and Pattern Recognition*, 7 pp.

Asteriadis, S.; Nikolaidais, A.; Hajdu, A. y Pitas, I. (2006). "An Eye Detection Algorithm Using Pixel to Edge Information". En *International Conference on Computer Vision Theory and Applications*.

Batrinca, Ligia Maria; Lepri, Bruno; Mana, Nadia y Pianesi, Fabio. (2012). "Multimodal Recognition of Personality Traits in Human-Computer Collaborative Tasks". En *Proceedings of the 14th ACM International Conference on Multimodal Interaction*, pp. 39-46.

Belhumeur, P. N.; Hespanha, J. P. y Kriegman, D. J. "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection". En *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, pp. 711-720.

Biederman, Irving; Mezzanotte, Robert J. y Rabinowitz, Jan C. (1982). "Scene Perception: Detecting and Judging Objects Undergoing Relational Violations". En *Cognitive Psychology*, 14, pp. 143-177.

Biorobotics Laboratory. (2016). "Pumas@Home 2016 Team Description Paper". En *Proceedings of RoboCup 2016*. 8 pp.

Bourdev, Lubomir y Malik, Jitendra. (2009). "Poselets: Body Part Detectors Trained Using 3D Human Pose Annotations". En *International Conference on Computer Vision*, pp. 1365-1372.

Camargo, Luis; Ivanovska, Iryna; Moriarty, Alexander; Nguyen, Minh; Thoduka, Santosh; Vazquez, Daniel; Kuestenmacher, Anastassia y Ploeger, Paul G. (2016). "The b-it-bots@Home 2016. Team Description Paper". En *Proceedings of RoboCup 2016*. 7 pp.

Caron, Louis-Martin; Otis, Samuel; Fortin, Jonathan; Doyle, Alexandre; Cousineau, Jeffrey; Haché, Xavier; Lépine, Cassandra; Chartrand, Alexandre; Pouplier, Thierry; Combet-Blanc, Jérôme; Lemay, Raphael y Poirier, Jimmy. (2016). "Walking Machine 2016 Team Description". En *Proceedings of RoboCup 2016*. 9 pp.

Cielniak, Grzegorz y Duckett, Tom. (2004). "People Recognition by Mobile Robots". En *Journal of Intelligent and Fuzzy Systems*, Vol. 15, pp. 21-27.

Chen, Huizhong; Gallagher, Andrew y Girod, Bern. (2012). "Describing Clothing by Semantic Attributes". En *Lecture Notes in Computer Science*, Vol. 7574, pp. 609-623.

Chen, Tzu-Yang; Wu, Min-Yu; Hsieh, Yu-Hsun y Fu, Li-Chen. (2016). "Deep learning for integrated hand detection and pose estimation". En *Proceedings of the 23rd International Conference on Pattern Recognition*, pp. 615-620.

Cho, Taeg Sang; Freeman, William T. y Tsao, Hensin. (2007). "A reliable skin mole localization scheme". En *IEEE 11th International Conference on Computer Vision*, pp. 1-8.

Choi, Jae Young; De Neve, Wesley; Ro, Yong Man y Plataniotis, Konstantinos N. (2010). "Automatic Face Annotation in Personal Photo Collections Using Context-Based Unsupervised Clustering and Face Information Fusion". En *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 20, No. 10, pp. 1292-1309.

Comport, Andrew I.; Marchand, Éric y Chaumette, François. (2005). “Complex articulated object tracking”. En *Electronic Letters on Computer Vision and Image Analysis*, 5(3), pp. 21-31.

Corvee, Etienne y Bremond, Francois. (2010). “Body parts detection for people tracking using trees of Histogram of Oriented Gradient descriptors”. En *IEEE Advanced Video and Signal-Based Surveillance*, pp. 469-475.

Cour, Timothee; Sapp, Benjamin y Taskar, Benn. (2011). “Learning from Partial Labels”. En *Journal of Machine Learning Research*, 12, pp. 1225-1261.

Cushen, George A. y Nixon, Mark S. (2012). “Real-Time Semantic Clothing Segmentation”. En *International Symposium on Visual Computing*, pp. 272-281.

Dantcheva, Antitza; Velardo, Carmelo; D’Angelo, Angela y Dugelay, Jean-Luc. (2010). “Bag of soft biometrics for person identification. New trends and challenges”. En *Multimed Tools Appl*, 51, pp. 739-777.

De Souza, Josemar R.; Mercês, Ramon C.; Matos, Ricardo S.; Pinheiro, Oberdan R. y Simões, Marco A. C. (2016). “BahiaRT@Home 2016 Team Description Paper”. En *Proceedings of RoboCup 2016*. 8 pp.

Demura, Kosei; Demura, Kensei; Nagashima, Kazuki y Enomoto, Koyo. (2016). “The demura.net 2016 Team Description”. En *Proceedings of RoboCup 2016*. 8 pp.

Dhahri, Royaka y Belaid, Samir. (2013). “A Robust Detection and Removal of Beard from Three Dimensional Human Face”. En *5th International Conference on Modeling, Simulation and Applied Optimization*, pp. 1-5.

Everingham, Mark; Sivic, Josef y Zisserman, Andrew. (2009). “Taking the bite out of automated naming on characters in TV video”. En *Image and Vision Computing*, 27, pp. 545-559.

Foresti, Henrique; Finch, Gabriel; Cavalcanti, Lucas; Alves, Felipe; Lacerda, Diogo; Brito, Raphael; Villa Verde, Filipe; Barros, Tiago; Freitas, Felipe; Lima, Leonardo; Ribeiro, Wayne; Mabuse, H. D.; Teichrieb, Veronica

y Teixeira, João Marcelo. (2016). “Emotive Robotics with I-Zak”. En *Proceedings of RoboCup 2016*. 10 pp.

Forsyth, David A.; Malik, Jitendra; Fleck, Margaret M.; Greenspan, Hayit; Leung, Thomas; Belongie, Serge; Carson, Chad y Bregler, Chris. (1996). “Finding Pictures of Objects in Large Collection of Images”. En *ECCV'96: Proceedings of the International Workshop on Object Representation in Computer Vision II*, pp. 335-360.

Fu, Jianlong; Wang, Jinqiao; Li, Zechao; Xu, Min y Lu, Hanqing. (2012). “Efficient Clothing Retrieval with Semantic-Preserving Visual Phrases”. En *Asian Conference on Computer Vision*, pp. 1-14.

Gallagher, Andrew C. y Chen, Tsuhan. (2008). “Clothing Cosegmentation for Recognizing People”. En *IEEE Conference on Computer Vision and Pattern Recognition 2008*, pp. 1-8.

Gallagher, Andrew C. y Chen, Tsuhan. (2009a). “Understanding Images of Groups of People”. En *IEEE Conference on Computer Vision and Pattern Recognition 2009*, pp. 256-263.

Gallagher, Andrew C. y Chen Tsuhan. (2009b). “Using Context to Recognize People in Consumer Images”. En *IPSN Transactions on Computer and Vision Applications*, Vol. 1, pp. 115-126.

Ghofrani, Pouya; Neshat, Zahra y Aghaeinia, Hassan. (2012). “Labeling Algorithm for Face Detection Using Skin and Hair Characteristics”. En *Journal of Electronic, Science and Technology*, Vol. 10, No. 2, pp. 135-141.

Guo, Jiacheng; Yao, Haotian; Ma, Haocheng; Guo, Cong; Dong, Yu; Zhu, Yilin; Peng, Jingsong; Wang, Xukang; He, Shuncheng; Xia, Fei y Zhang, Xunkai. (2016). “Tinker@Home 2016 Team Description Paper”. En *Proceedings of RoboCup 2016*. 8 pp.

Hajiarbabi, Mohammadreza y Agah Arvin. (2015). “Human Skin Detection in Color Images Using Deep Learning”. En *International Journal of Computer Vision and Image Processing*, 5(2), 13 pp.

Hara, Kota; Jagadeesh, Vignesh y Piramuthu, Robinson. (2016). “Fashion Apparel Detection: The Role of Deep Convolutional Neural Network and Pose-dependent Priors”. En *IEEE Winter Conference on Applications of*

Computer Vision, 9 pp.

Hasan, Basela S. y Hogg, David C. (2010). "Segmentation using Deformable Spatial Priors with Application to Clothing". En *Proceedings of British Machine Vision Conference 2010*, pp. 1-11.

Hefin, Brian; Scheirer, Walter y Boulton, T. E. (2012). "Detecting and Classifying Scars, Marks, and Tattoos Found in the Wild". En *IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems*, pp. 31-38.

Hegger, Frederick; Hochgeschwender, Nico; Kraetzschmar, Gerhard K. y Ploeger, Paul G. (2012). "People Detection in 3D Point Clouds Using Local Surface Normals". En *RoboCup 2012: Robot Soccer World Cup XVI*, pp. 154-165.

Heitz, Jeremy y Koller, Daphne. (2008). "Learning Spatial Context: Using Stuff to Find Things". En *European Conference on Computer Vision 2008*, pp. 30-43.

Hu, Zhilan; Yan, Hong y Lin, Xinggang. (2008). "Clothing Segmentation using foreground and background estimation based on the constrained Delaunay triangulation". En *Pattern Recognition*, 41, pp. 1581-1592.

Jain, Anil K.; Dass, Sarat C. y Nandakumar Karthik. (2004). "Soft Biometric Traits for Personal Recognition Systems". En *Proceedings of International Conference on Biometric Authentication*, pp. 731-738.

Jain, Anil K. y Park Unsang. (2009). "Facial Marks: Soft Biometric for Face Recognition". En *IEEE International Conference on Image Processing*, pp. 37-40.

Jain, Anil K.; Ross, Arun y Prabhakar, Salil. (2004). "An Introduction to Biometric Recognition". En *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 14, No. 1, pp. 4-20.

Johnson, Matthew Alastair. (2008). *Semantic Segmentation and Image Search*. Phd Thesis, University of Cambridge, 161 pp.

Julian, P.; Dehais, C.; Lauze, F.; Charvillat, V.; Bartoli, A. y Choukroun, A. (2010). "Automatic Hair Detection in the Wild". En *IEEE International*

Conference on Pattern Recognition, pp. 4617-4620.

Kalal, Zdenek; Mikolajczyk, Krystian y Matas, Jiri. (2012). "Tracking-learning-detection". En *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 34, No. 7, pp. 1409–1422.

Kakumanu, P.; Makrogiannis, S. y Bourbakis, N. (2007). "A survey of skin-color modeling and detection methods". En *Pattern Recognition*, 40, pp. 1106-1122.

Karahan, Şamil y Akgül, Yusuf Sinan. (2016). "Eye detection by using deep learning". En *24th Signal Processing and Communication Application Conference*, pp. 2145-2148.

Kim, Heung-Nam; El Saddik, Abdulmotaleb y Jung, Jin-Guk. (2012). "Leveraging personal photos to inferring friendships in social networks services". En *Expert Systems with Applications*, 39, pp. 6955-6966.

Kumar, Ajay y Wang Bichai. (2015). "Recovering and matching minutiae patterns from finger knuckle images". En *Pattern Recognition Letters*, 68, pp. 361-367.

Lao, Brian y Jagadeesh, Karthik. (2015). "Convolutional Neural Networks for Fashion Classification and Object Detection". En

Le, T. Hoang Ngan; Luu, Khoa; Seshadri, Keshav y Savvides, Marios. (2012). "Beard and mustache segmentation using sparse classifiers on self-quotient images". En *19th IEEE International Conference on Image Processing*, pp. 165-168.

Leber, Raphael; Saraydaryan, Jacques; Jumel, Fabrice; Evers, Kathrin y Vouillon, Thibault. (2016). "CPE Lyon Robot Forum, 2016 Team Description Paper". En *Proceedings of RoboCup 2016*. 8 pp.

Lee, Jung-Eun; Jin, Rong; Jain, Anil K. y Tong, Wei. (2012). "Image Retrieval in Forensics: Tatto Image Database Application". En *IEEE Multimedia*, Vol. 19(1), pp. 40-49.

Lee, Kuang-chih; Anguelov, Dragomir; Sumengen, Baris y Goktur, Salih Burak. (2008). "Markov Random Field Models for Hair and Face Segmentation". En *IEEE Conference on Automatic Face and Gesture*

Recognition, pp. 1-6.

Lee, Yong Jae y Grauman, Kristen. (2011). "Face Discovery with Social Context". En *British Machine Vision Conference*, 11 pp.

Li, Xuelong; Maybank, Stephen J.; Yan, Shuicheng; Tao, Dacheng and Xu, Dong. (2008). "Gait Components and Their Application to Gender Recognition". In *IEEE Transactions on Systems, Man and Cybernetics, Part C: Applications and Reviews*, Vol. 38, No. 2, pp. 145-155.

Lin, Dahua; Kapoor, Ashish; Hua, Gang y Baker, Simon. (2010). "Joint People, Event, and Location Recognition in Personal Photo Collections using Cross-Domain Context". En *ECCV Lecture Notes in Computer Science*, vol. 6311, pp. 243-256.

Lin, Hui; Lu, Huchuan y Zhang Lihe. (2006). "A New Automatic Recognition System of Gender, Age and Ethnicity". En *The Sixth World Congress on Intelligent Control and Automation*, Vol. 2, 4 pp.

Lipowesky, Uri; Mamo, Omri y Cohen, Avihai. (2008). "Using integrated color and texture features for automatic hair detection". En *IEEE 25th Convention of Electrical and Electronics Engineers in Israel*, pp. 51-55.

Liu, Ziwei; Luo, Ping; Qiu, Shi; Wang, Xiaogang y Tang, Xiaoou. (2016). "DeepFashion: Powering Robust Clothes Recognition and Retrieval with Rich Annotations". En *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1096-1104.

Lunenburg, J.J.M.; van den Dries, S.; Appeldoorn, R.P.W.; Wijnands, R.W.J. y van de Molengraft, M.J.G. (2016). "Tech United Eindhoven @Home 2016 Team Description Paper". En *Proceedings of RoboCup 2016*. 8 pp.

Masiero, Andrey A.; De Rizzo Meneghetti, Douglas; Contador, Leonardo; Vasconcelos, Lucas; Tonidandel, Flavio y Aquino Junior, Plinio T. (2016). "RoboFEI@Home 2016 Team Description Paper". En *Proceedings of RoboCup 2016*. 9 pp.

Marr, David. (1982). *Vision*. W.H. Freeman and Company, Nueva York.

Mehrabi, Esmaeil; Babaians, Edwin; Ahmadi, Alireza; Sheikhjafari, Ameneh; Gharghabi, Shaghayegh; Khazaei Korghond, Navid y Shiry

Ghidary, Saeed. (2016). "AUT @Home 2016 Team Description Paper". En *Proceedings of RoboCup 2016*. 9 pp.

Memmesheimer, Raphael; Seib, Viktor; Heuer, Gregor; Schmidt, Patrik; Thies, Darius; Mykhalchyshyna, Ivanna; Klöckner, Johannes; Schmitz, Martin; Yann Wettengel, Niklas; Geilen, Nils; Schötz, Richard; Polster, Florian y Paulus, Dietrich. (2016). "RoboCup 2016 - homer@UniKoblenz (Germany)". En *Proceedings of RoboCup 2016*. 9 pp.

Meyer zu Borgsen, Sebastian; Korthals, Timo y Wachsmuth, Sven. (2016). "ToBI - Team of Bielefeld. The Human-Robot Interaction System for RoboCup@Home 2016". En *Proceedings of RoboCup 2016*. 9 pp.

Naaman, Mor; Yeh, Ron B; Garcia-Molina, Hector y Paepcke, Andreas. (2005). "Leveraging Context to Resolve Identity in Photo Albums". En *ACM/IEEE Joint Conference on Digital Libraries*, pp. 178-187.

Newell, Allen. (1981). "The Knowledge Level: Presidential Address". En *AI Magazine*, 2(2), pp. 1-20.

Nowak, Eric; Jurie, Frédéric y Triggs, Bill. (2006). "Sampling Strategies for Bag-of-Features Image Classification". En *Proceedings of the European Conference on Computer Vision*, pp. 490-503.

Nguyen, Minh Hoai; Lalonde, Jean-Francois; Efros, Alexei A. y De la Torre, Fernando. (2008). "Image-based Shaving". En *Computer Graphics Forum Journal*, 27(2), pp. 627-635.

Nixon, Mark S.; Correia, Paulo L.; Nasrollahi, Kamal; Moeslund, Thomas B.; Hadid, Abdenour y Tistarelli, Massimo. (2015). "On soft biometrics". En *Pattern Recognition Letters*, 68, pp. 218-230.

Okada, Hiroyuki; Nagai, Takayuki; Nakamura, Tomoaki; Sugiura, Komei; Iwahashi, Naoto e Inamura, Tetsunari. (2016). "Team eR@sers 2016 in the @Home League Team. Description Paper". En *Proceedings of RoboCup 2016*. 8 pp.

Peng, Kun y Chen, Liming. (2005). "A Robust Algorithm for Eye Detection on Gray Intensity Face without Spectacles". En *Journal of Computer Science & Technology*, 5(3), pp. 127-132.

Perez-Sala, Xavier; Escalera, Sergio y Angulo, Cecilio. (2012). "Survey on 2D and 3D Human Pose Recovery". En *CCIA Frontiers in Artificial Intelligence and Applications*, vol. 248, pp. 101-110.

Pineda, Luis A. (2008). "Specification and interpretation of multimodal dialogue models for human-robot interaction". En *Artificial Intelligence for Humans: Service Robots and Social Modeling*, SMIA, pp. 33-50.

Pineda, Luis A.; Meza, Ivan V. y Salinas, Lisset. (2010). "Dialogue Model Specification and Interpretation for Intelligent Multimodal HCI". En *Advances in Artificial Intelligence - Iberamia 2010*, pp. 20-29.

Pineda, Luis A.; Meza, Ivan; Aviles, Hector; Gershenson, Carlos; Rascon, Caleb; Alvarado, Montserrat y Salinas, Lisset. (2011). "IOCA: Interaction-Oriented Cognitive Architecture". En *Research in Computing Science*, 54, pp. 273-284.

Pineda, Luis A.; Rodriguez, Arturo; Fuentes, Gibran; Rascon, Caleb y Meza, Ivan V. (2013a). "Concept and Functional Structure of a Service Robot". En *International Journal of Advanced Robotic Systems*, pp. 1-15.

Pineda, Luis A.; Salinas, Lisset; Meza, Ivan; Rascon, Caleb y Fuentes, Gibran. (2013b). "SitLog: A Programming Language for Service Robot Tasks". En *International Journal of Advanced Robotic Systems*, pp. 1-12.

Pineda, Luis A.; Rascon, Caleb; Fuentes, Gibran; Rodriguez, Arturo; Ortega, Hernando; Reyes, Mauricio; Hernández, Noé y Cruz, Ricardo. (2016). "The Golem Team, RoboCup@Home 2016". En *Proceedings of RoboCup 2016*. 7 pp.

Pineda, Luis A.; Rascon, Caleb; Fuentes, Gibran; Rodríguez, Arturo; Ortega, Hernando; Reyes, Mauricio; Hernández, Noé; Cruz, Ricardo; Vélez, Ivette y Ramírez, Marco. (2017a). "The Golem Team. Robocup@Home 2017". En *Proceedings of RoboCup, 2017*.

Pineda, Luis A.; Rodriguez, Arturo; Fuentes, Gibran; Rascon, Caleb y Meza, Ivan V. (2017b) "A Light Non-Monotonic Knowledge-Base for Service Robots". En *Intelligent Service Robotics*, pp. 1-13.

Poppe, Ronald. (2012). "Facing scalability: Naming faces in an online social network". En *Pattern Recognition*, 45, pp. 2335-2347.

Prihasto, Bima; Choirunnisa, Shabrina; Nurdiansyah, Muhammad Ishak; Mathulaprangsan, Seksan; Chu, Vivian Ching-Mei; Chen, Shi-Huang y Wang, Jia-Ching. (2016). "A survey of deep face recognition in the wild". En *Proceedings of International Conference on Orange Technologies*, pp. 76-79.

Qin, Siyang; Kim, Seongdo y Manduchi, Roberto. (2017). "Automatic skin and hair masking using fully convolutional networks". En *Proceedings of IEEE International Conference on Multimedia and Expo*, pp. 103-108.

Rabinovich, Andrew; Vedaldi, Andrea; Galleguillos, Carolina; Wiewiora, Eric y Belongie, Serge. (2007). "Objects in Context". En *IEEE International Conference in Computer Vision*, pp. 1-8.

Read, Jesse; Pfahringer, Bernhard; Holmes, Geoff y Frank, Eibe. (2009). "Classifier chains for multi-label classification.". En *Proceedings of 13th European Conference on Principles and Practice of Knowledge Discovery in Databases and 20th European Conference on Machine Learning*, pp. 254-269.

Reid, D. A.; Samangoei, S.; Chen, C.; Nixon, M. S. y Ross A. (2013). "Soft Biometrics for Surveillance: An Overview". En *Handbook of Statistics*, Vol. 31.

Rigas, Ioannis y Komogortsev, Oleg V. (2015). "Eye movement-driven defense against iris print-attacks". En *Pattern Recognition Letters*, 68, pp. 316-326.

Rodríguez Lera, Francisco J.; Matellán, Vicente y Martín Rico, Francisco. (2016). "LeonRobot Team Description Paper. RoboCup@home 2016". En *Proceedings of RoboCup 2016*. 8 pp.

Romero Cordero, Rogelio Adrian. (2015). "Sistema de detección y rastreo de personas en tiempo real para el robot Golem-II+". Tesis de licenciatura, Universidad Nacional Autónoma de México, 132 pp.

Shafi, Muhammad y Chung, Paul W. H. (2009). "A Hybrid Method for Eyes Detection in Facial Images". En *International Journal of Electrical and Electronics Engineering*, pp. 231-236.

Shuai, Wei; Liu, Jiangchuan; Wang, Xiping; Zhou, Feng y Chen, Xiaoping.

(2016). “WrightEagle@Home 2016 Team Description Paper”. En *Proceedings of RoboCup 2016*. 9 pp.

Sidibe, D.; Montesinos, P. y Janaqi, S. (2006). “A simple and efficient eye detection method in color images”. En *International Conference Image and Vision Computing New Zealand*, 5 pp.

Sidorov, Maxim y Minker, Wolfgang. (2014). “Emotion Recognition and Depression Diagnosis by Acoustic and Visual Features: A Multimodal Approach”. En *Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge*, pp. 81-86.

Sivic, Josef; Zitnick, C. Lawrence y Szeliski Richard. (2006). “Finding people in repeated shots of the same scene”. En *Proceedings of the British Machine Vision Conference*, pp. 909-918.

Song, Yang y Leung Thomas. (2006). “Context-aided Human Recognition – Clustering”. En *ECCV 3, Lecture Notes in Computer Science*, 3953, pp. 382-395.

Sucar, L. Enrique; Morales, Eduardo; Martínez-Carranza, José; Coronado, Luis V.; Carrillo, David; Marquez, Francisco y González, Juan Carlos. (2016). “Markovito’s Team Description RoboCup@Home 2016”. En *Proceedings of RoboCup 2016*. 8 pp.

Tan, Jeffrey Too Chuan; Duan, Feng; Zamzuri, Hairi bin; Abdul Rahman, Mohd Azizi bin; Ismail, Zool Hilmi bin; Lim, Cheng Siong; Uchimura, Yutaka y Arai, Tamio. (2016). “KameRider @Home 2016 Team Description”. En *Proceedings of RoboCup 2016*. 8 pp.

Toshev, Alexander y Szegedy, Christian. (2014). “DeepPose: Human Pose Estimation via Deep Neural Networks”. En *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition 2014*, pp. 1653-1660.

Turk, Matthew A. y Pentland, Alex P. (1991). “Face recognition using eigenfaces”. En *Computer Vision and Pattern Recognition*, pp. 586-591.

Turkan, Mehmet; Pardas, Montse y Cetin, A. Enis. (2008). “Edge projections for eye localization”. En *SPIE Optical Engineering*, Vol. 47, 6 pp.

Ueda, Ryuichi; Hayashibara, Yasuo; Fujie, Shinya; Aoki, Yuya; Inoue, Hirofumi; Matsuzaki, Hiroto; Natusako, Kazuya; Ohno, Akie; Terato, Shotaro e Irie, Kiyoshi. (2016). "CIT Brains@Home 2016 Team Description Paper". En *Proceedings of RoboCup 2016*. 7 pp.

Van Beek, Loy; Chen, Kai; Holz, Dirk; Matamoros, Mauricio; Rascon, Caleb; Rudinac, Maja; Ruiz des Solar, Javier y Wachsmuth, Sven. (2016). "Robocup@Home 2016: Rule and regulations".

Ventura, Rodrigo; Mateus, André; Garcia, João; Basiri, Meysam; Miraldo, Pedro; Santos, Pedro y Lima, Pedro U. (2016). "SocRob@Home: Team Description Paper for RoboCup@Home 2016". En *Proceedings of RoboCup 2016*. 8 pp.

Viola, Paul y Jones, Michael. (2001). "Robust Real-time Object Detection". En *International Journal of Computer Vision*, Vol. 57, pp. 137-154.

Wang, Jian-Gang y Yau, Wei-Yun. (2013). "Real-time Beard Detection by Combining Image Decolorization and Texture Detection with Applications to Facial Gender Recognition". En *IEEE Workshop on Computational Intelligence in Biometrics and Identity Management*, pp. 58-65.

Wang, Nan y Ai, Haizhou. (2011). "Who Blocks Who: Simultaneous Clothing Segmentation for Grouping Images". En *IEEE International Conference on Computer Vision*, pp. 1535-1542.

Wang, Nan; Ai, Haizhou y Lao Shihong. (2010). "A Compositional Exemplar-Based Model for Hair Segmentation". En *Proceedings of the 10th Asian Conference on Computer Vision*, Part III, pp. 171-184.

Wang, Zixuan y Yan, Jinyun. (2013). "Who and Where: People and Location Co-Clustering". En *IEEE International Conference on Image Processing*, 5 pp.

Weber, Michael; Bäuml, Martin y Stiefelhagen, Rainer. (2011). "Part-based Clothing Segmentation for Person Retrieval". En *8th IEEE International Conference on Advanced Video and Signal-Based Surveillance*, pp. 361-366.

Wongsuwan, Kandith; Chaiso, Krit; Kaewnak, Nut; Songkrasin,

Konlayut; Yingsukamol, Nathas; Visetsuthimont, Thanachote y Sukvichai, Kanjanapan. (2016). "SKUBA 2016 Team Description". En *Proceedings of RoboCup 2016*. 9 pp.

Wright, John; Yang, Allen Y.; Ganesh, Arvind; Sastry, S. Shankar y Ma, Yi. (2009). "Robust Face Recognition via Sparse Representation". En *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 31, No. 2, pp. 210–227.

Yacoob, Yaser y Davis, Larry. (2006). "Detection and analysis of hair". En *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 28, No. 7, pp. 1164-1169.

Yamaguchi, Kota; Kiapour, M. Hadi; Ortiz, Luis E. y Berg, Tamara L. (2012). "Parsing Clothings in Fashion Photographs". En *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition 2012*, pp. 3570-3577.

Yang, Ming y Yu, Kai. (2011a). "Real-Time Clothing Recognition in Surveillance Videos". En *18th IEEE International Conference on Image Processing*, pp. 2937-2940.

Yang, Ming; Zhu, Shenghuo; Lv, Fengjun and Yu, Kai. (2011b). "Correspondence Driven Adaptation for Human Profile Recognition". In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 505-512.

Yang, Ming-Hsuan; Kriegman, David J. y Ahuja, Narendra. (2002). "Detecting Faces in Images: A Survey". En *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 1, pp. 34-58.

Yang, Xiaodong; Yuan, Shuai y Tian, YingLi. (2011c). "Recognizing Clothes Patterns for Blind People by Confidence Margin based Feature Combination". En *Proceedings of the 19th ACM International Conference on Multimedia*, pp. 1097-1110.

Yao, Susu; Khan, Ishtiaq Rasool y Farbiz, Farzam. (2011). "Clothing Segmentation and Recoloring using Background Subtraction and Back Projection Method". En *18th IEEE International Conference on Image Processing*, pp. 3137-3140.

Zafeiriou, Stefanos; Zhang, Cha y Zhang, Zhengyou. (2015). “A survey on face detection in the wild: Past, present and future”. En *Computer Vision and Image Understanding*, Vol. 138, pp. 1-24.

Zaman, Khan Rafiqul y Adnan, Ibraheem Noor. (2012). “Hand Gesture Recognition: A Literature Review”. En *International Journal of Artificial Intelligence & Applications*, Vol. 3, No. 4, pp. 161-174.

Zeng, Zhihong; Pantic, Maja; Roisman, Glenn I. y Huang, Thomas S. (2007). “A Survey of Affect Recognition Methods: Audio, Visual and Spontaneous Expressions”. En *Proceedings of the 9th International Conference on Multimodal Interfaces*, pp. 126-133.

Zhang, Cha y Zhang, Zhengyou. (2010). *A Survey of Recent Advances in Face Detection*. Technical Report, Microsoft Research, 17 pp.

Zhang, Zhaoxiang; Hu, Maodi y Wang, Yunhong. (2011). “A Survey of Advances in Biometric Gait Recognition”. En *Proceedings of the 6th Chinese Conference on Biometric Recognition*, pp. 150-158.