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

RECONOCIMIENTO ROBUSTO DE ROSTROS EN AMBIENTES DINÁMICOS

TESIS PARA OPTAR AL GRADO DE DOCTOR

EN INGENIERÍA ELÉCTRICA

MAURICIO ALFREDO CORREA PÉREZ

PROFESOR GUÍA:
JAVIER RUIZ DEL SOLAR

MIEMBROS DE LA COMISIÓN:
DOMINGO MERY QUIROZ
JORGE SILVA SÁNCHEZ
MIGUEL TORRES TORRITI

SANTIAGO DE CHILE

JULIO 2012

RESUMEN DE LA TESIS
PARA OPTAR AL GRADO DE
DOCTOR EN INGENIERIA ELECTRICA
POR: MAURICIO ALFREDO CORREA PEREZ
FECHA: 31 DE JULIO DE 2012
PROF. GUÍA: JAVIER RUIZ DEL SOLAR

En la actualidad un problema fundamental para los sistemas robóticos que basan su sistema sensorial en la utilización de cámaras de video y sistemas de visión computacional es detectar y reconocer objetos de interés en ambientes no controlados. Por otro lado, el análisis del rostro juega un papel muy importante en la construcción de un sistema de Interacción Humano-Robot (HRI) que permita a los humanos interactuar con sistemas robóticos de un modo natural. En este trabajo de tesis se diseña e implementa un sistema de visión que opera en ambientes no controlados, y que es capaz de detectar y reconocer rostros humanos en forma robusta, utilizando métodos de visión activa e integrando diferentes tipos de contexto.

Se plantea una metodología para la construcción del sistema de visión propuesto en forma general y se define cuales son los módulos principales que lo componen. Entre los cuales están los módulos de detección y reconocimiento de rostros, en particular el uso de contexto y un módulo de visión activa. Estos módulos permiten descartar falsas detecciones y realizar modificaciones a las observaciones para así mejorar el rendimiento del sistema de reconocimiento de rostros.

Se desarrolla un simulador que se utiliza para validar el sistema general y en particular evaluar el funcionamiento de los diferentes módulos planteados. Este simulador es una poderosa herramienta que permite realiza evaluaciones de métodos de detección y reconocimiento de rostros ya que genera las observaciones de un agente dentro de un mapa virtual con personas. De los experimentos en el simulador y en otros ambientes se puede concluir que los módulos de contexto realizan un aporte significativo en el rendimiento del sistema de visión, mejorando las tasas de reconocimiento y reduciendo las tasas de falsos positivos en las detecciones de rostros. La tasa de reconocimiento aumenta de 78.41% a 86.77% con el uso de filtros de contexto. El uso de visión activa permite que la tasa de reconocimiento mejore de 86.77% a 92.92%, ya que permite que se construya una mejor galería (en caso que la galería se construye *online*), y mejorar la pose del robot con respecto a la persona en la etapa de reconocimiento.

Se desarrolla un sistema robusto para la detección y la identificación de seres humanos en entornos domésticos el cual es evaluado en un robot de servicio. La principal función es evaluar el funcionamiento del sistema de visión propuesto en una aplicación real. Se agrega un nuevo sensor (cámara térmica) y se agregan nuevos módulos al sistema (Detección de Piel Visible y Térmica, Detección y reconocimiento de Rostros Térmico, Detección de Personas). Los resultados de la evaluación del sistema en una aplicación real (prueba enmarcada en la competencia de robótica RoboCup, que se llama "*Who is Who*") confirman que el uso de contexto mejora el rendimiento del sistema, permitiendo aumentar la tasa de reconocimiento de 54% a 74% y reduciendo el numero de falsos positivos a 0. Nuevamente la visión activa fue un factor importante para mejorar el desempeño del sistema en general, en todos los experimentos influyó de forma positiva en el funcionamiento del sistema.

A mi amada esposa Marcela

A mis Padres

Agradecimientos

Me gustaría que estas líneas sirvieran para expresar mi más profundo y sincero agradecimiento a todas aquellas personas que con su ayuda directa o indirecta han colaborado en la realización del presente trabajo.

A mi esposa Marcela, sin su apoyo, colaboración e inspiración habría sido imposible llevar a cabo esta tesis. A mis padres, Renato y Sonia quienes me han ayudado en todo aspecto posible.

Debo agradecer de manera especial y sincera al Profesor Javier Ruiz del Solar por sugerirme realizar esta tesis doctoral bajo su dirección. Su apoyo y confianza en mi trabajo y su capacidad para guiar mis ideas ha sido un aporte invaluable, no solamente en el desarrollo de esta tesis, sino también en mi formación como investigador. Le agradezco también el haberme facilitado siempre los medios suficientes para llevar a cabo todas las actividades propuestas durante el desarrollo de esta tesis.

Agradezco a los profesores Domingo Mery, Jorge Silva y Miguel Torres por revisar y corregir esta tesis.

Gracias también a mis queridos compañeros, que me apoyaron y ayudaron durante este tiempo en el doctorado. A Rodrigo Verschae por su apoyo, amistad y ayuda. A Daniel Herrmann, Fernando Bernuy, Felipe Smith, Isao Parra, Paul Vallejos, Pablo Guerrero, Rodrigo Palma, Rodrigo Asenjo y Jose Delpiano por su amistad, ideas y consejos. Y a todos los integrantes del laboratorio de Robótica y del laboratorio de Visión Computacional con quienes he compartido durante tanto tiempo. Y como no agradecer a la Señora Eliana por toda su ayuda en estos años.

Finalmente, agradezco a CONICYT por el financiamiento, apoyo y ayuda para la realización de tesis doctoral y a MECESUP a través del Proyecto FSM 0601. También me gustaría agradecer a FONDECYT ya que esta tesis fue financiada parcialmente por el proyecto 1090250. Y un agradecimiento especial al Centro Avanzado de Tecnología para la Minería (Proyecto CONICYT FBO09).

Índice General

Índice de Figuras

Índice de Tablas

Capítulo 1
Aportes de la Tesis

En este capítulo se presentan las contribuciones de este trabajo de tesis y las publicaciones relevantes generadas por esta tesis y otros trabajos realizados en la Universidad mientras se realizó el doctorado.

Desarrollo de sistema de reconocimiento robusto de rostros en ambientes dinámicos

Un aporte importante en este tema es que se propone una metodología general para la construcción de sistemas de reconocimiento de rostros que usen contexto y visión activa.

Otro aporte es el uso de visión térmica para complementar y mejorar el sistema de visión propuesto. Se implementa un sistema que usa imágenes térmicas y visibles para buscar y reconocer personas en un ambiente dinámico. Otros aportes realizados son:

- Se integran diferentes instancias de contexto para mejorar el sistema de visión que posee un robot de servicio sin que esto implique un aumento significativo en el tiempo de procesamiento requerido.
- Se propone el uso de un módulo de visión activa, que retroalimenta el módulo de actuación, utilizando la información obtenida desde el sistema de visión. Este módulo permite modificar las observaciones realizadas y con ello mejorar el rendimiento del sistema de reconocimiento de rostros.
- Se presentan estudios comparativos de métodos de reconocimiento de rostros, estudios que son utilizados para elegir el mejor método de reconocimiento de rostros que cumple con las características buscadas que son: (i) *Operación Online*: Que no existan etapas de entrenamiento o aprendizaje offline; (ii) *Operación en Tiempo Real:* el sistema debe ser capaz de responder sin grandes retrasos; (iii) *Un rostro por persona*: Una sola imagen por persona en la base de datos debe ser suficiente para que el sistema de visión tenga un buen rendimiento en la identificación de las personas; y (iv) *Ambiente dinámico*: El sistema debe ser capaz de tener un buen rendimiento en ambientes no controlados (iluminación, fondo, etc.).

- Se crean nuevas bases de datos, las cuales quedarán disponibles para su uso en estudios futuros (*Illumination Database*, *Rotation Database*, *Distance Database*, *Arena Database*, *HRI Database Distance*, *HRI Database Rotation*, *HRI Database Expression*).

Ambiente virtual para evaluación de sistemas de detección y reconocimiento de rostros

Otro aporte de esta tesis es el diseño y construcción de un entorno virtual que permite evaluar sistemas de detección de rostros en condiciones no controladas (pose, iluminación, expresión, etc.). Esta poderosa herramienta permite a un agente navegar y observar imágenes con rostros reales, a diferentes distancias, ángulos y con la iluminación interior o exterior. Hasta donde sabemos no existen otros simuladores de este tipo disponibles. Durante el desarrollo de este sistema se realizaron los siguientes aportes:

- Se diseña y construye un dispositivo de adquisición de imágenes portátil que mediante una cámara CCD montada en una estructura giratoria permite la captura de imágenes alrededor de los usuarios.
- Se crea una base de datos de rostros que posee imágenes de rostros con rotaciones en *yaw* (121 ángulos diferentes) y *pitch* (3 ángulos diferentes), y fondos e iluminación no controlados, dada la variabilidad de las imágenes del mismo usuario permite generar vistas del rostro desde casi cualquier ángulo. No existen bases de datos parecidas disponibles para realizar evaluaciones de métodos de detección o reconocimiento de rostros.
- Se realizan estudios comparativos en métodos de detección y reconocimiento de rostros.

Publicaciones

Las publicaciones generadas que están relacionadas directamente con el trabajo en esta tesis son las siguientes:

1. Mauricio Correa, Gabriel Hermosilla, Rodrigo Verschae, Javier Ruiz-del-Solar. Human Detection and Identification by Robots using Thermal and Visual Information in Domestic Environments, Journal of Intelligent and Robotic Systems, Vol. 66, No.1-2, pp. 223-243, 2012. **(ISI)**
2. Javier Ruiz-del-Solar, Rodrigo Verschae, Mauricio Correa. Recognition of Faces in Unconstrained Environments: A Comparative Study. EURASIP Journal on Advances in Signal Processing (Recent Advances in Biometric Systems: A Signal Processing Perspective), Vol. 2009, Article ID 184617, 19 pages. **(ISI)**

3. Mauricio Correa, Javier Ruiz-del-Solar, Parra-Tsunekawa, I., Rodrigo Verschae (2011). A Realistic Simulation Tool for Testing Face Recognition Systems under Real-World Conditions. Lecture Notes in Computer Science 6556 (RoboCup Symposium 2010), pp. 13–24.
4. Mauricio Correa, Javier Ruiz-del-Solar, Parra-Tsunekawa, I. (2010). A Virtual Environment for realistic Testing and Training of Face Detection and Recognition

Systems, 19th IEEE Int. Symposium in Robot and Human Interactive Communication – Ro-Man 2010, Sept. 12-15, 2010, Viareggio, Italy.

5. Gabriel Hermosilla, Javier Ruiz-del-Solar, Rodrigo Verschae, Mauricio Correa, "Face Recognition using Thermal Infrared Images for Human-Robot Interaction Applications: A Comparative Study", 6th Latin American Robotics Symposium, LARS 2009 (CD Proceedings), Valparaiso, Chile, 2009.
6. Mauricio Correa, Javier Ruiz-del-Solar, Bernuy, F. (2009). Face Recognition for Human-Robot Interaction Applications: A Comparative Study. Lecture Notes in Computer Science 5399 (RoboCup Symposium 2008) pp. 473-484.
7. Rodrigo Verschae, Javier Ruiz-del-Solar, Mauricio Correa, "Face Recognition in Unconstrained Environments: A Comparative Study", In Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition, Marseille France, 2008.

Otras publicaciones generadas durante el trabajo en esta tesis son las siguientes:

1. Rodrigo Verschae, Javier Ruiz-del-Solar, Mauricio Correa, "A unified learning framework for object detection and classification using nested cascades of boosted classifiers", In Machine Vision Applications, Springer-Verlag New York, Inc., vol. 19, no. 2, Secaucus, NJ, USA, pp. 85-103, 2008. **(ISI)**
2. Gabriel Hermosilla, Javier Ruiz-del-Solar, Rodrigo Verschae, Mauricio Correa. A Comparative Study of Thermal Face Recognition Methods in Unconstrained Environments, Pattern Recognition, Vol. 45, No. 7, pp. 2445-2459. **(ISI)**

3. Javier Ruiz-del-Solar, Mauricio Correa, Pablo Hevia-Koch, Rodrigo Verschae. (2011). UChile HomeBreakers 2010 Team Description Paper, RoboCup Symposium 2011, Istanbul Turkey (CD Proceedings).
4. Javier Ruiz-del-Solar, Mauricio Correa, Lee-Ferng, J., Hevia-Koch, P., Parra, I., Mascaró, M. (2010). UChile HomeBreakers 2010 Team Description Paper, RoboCup Symposium 2010, 19-25 June 2010, Singapore (CD Proceedings).
5. Javier Ruiz-del-Solar, Mauricio Mascaró, Mauricio Correa, Fernando Bernuy, Romina Riquelme, Rodrigo Verschae, "Analyzing the Human-Robot Interaction Abilities of a General-Purpose Social Robot in Different Naturalistic Environments", In Lecture Notes in Computer Science (RoboCup Symposium 2009), vol. 5949, pp. 308-319, 2010.
6. Mauricio Correa, Rodrigo Verschae, Javier Ruiz-del-Solar, Jong Lee-Ferng, Nelson Castillo, "Real-Time Hand Gesture Recognition for Human Robot Interaction. Lecture Notes in Computer Science (RoboCup Symposium 2009)", In , vol. 5949, pp. 46-57, 2010.
7. Jong Lee-Ferng, Javier Ruiz-del-Solar, Mauricio Correa, Rodrigo Verschae, "Hand Gesture Recognition for Human Robot Interaction in Uncontrolled Environments", In Workshop on Multimodal Human - Robot Interfaces, 2010 IEEE International Conference on Robotics and Automation, Anchorage, Alaska, 2010.
8. Jong Lee-Ferng, Javier Ruiz-del-Solar, Rodrigo Verschae, Mauricio Correa, "Dynamic Gesture Recognition for Human Robot Interaction", 6th Latin American Robotics Symposium, LARS 2009 (CD Proceedings), Valparaiso, Chile, 2009.
9. Javier Ruiz-del-Solar, Mauricio Correa, Bernuy, F., Cubillos, S., Mascaró, M., Vargas, J., Norambuena, S., Marinkovic, A., and Galaz, J. (2008). UChile HomeBreakers 2008

Team Description Paper, RoboCup Symposium 2008, July 15 – 18, Suzhou, China (CD Proceedings).
10. Javier Ruiz-del-Solar, Rodrigo Verschae, Paul Vallejos, Mauricio Correa, "Face analysis for human computer interaction applications", In VISAPP (Special Sessions), pp. 23-30, 2007.

Capítulo 2
Introducción

2.1 Antecedentes

2.1.1 Fundamentación general

En la actualidad un problema fundamental para los sistemas robóticos que basan su sistema sensorial en la utilización de cámaras de video y sistemas de visión computacional es el de detectar y reconocer objetos de interés en ambientes no controlados. Incluso conociendo con anterioridad alguna información sobre el objeto: color, tamaño u otra; el problema no ha sido completamente resuelto por algún grupo de investigación hasta el momento. De hecho, el problema se acentúa cuando se trata de condiciones de iluminación muy variables y poco homogéneas, con la existencia de un fondo no uniforme y de objetos en movimientos. En otras palabras, cuando nos referimos a ambientes no controlados estamos hablando de ambientes altamente dinámicos, como aquellos en los cuales los seres humanos usualmente nos desenvolvemos. Un ejemplo es un ambiente de hogar, en que cada lugar posee una iluminación distinta, el fondo es no homogéneo, hay obstáculos que no siempre se encuentran en el mismo lugar, así como hay objetos en movimiento dentro del lugar.

Por otro lado, el análisis del rostro juega un papel muy importante en la construcción de sistemas de Interacción Humano-Robot (HRI) que permitan a los humanos interactuar con sistemas robóticos de un modo natural. Al igual que la voz y la firma manuscrita, el rostro de una persona es uno de los indicadores biométricos más usado por las personas. La capacidad para reconocer e identificar a otros seres humanos a partir de la información obtenida de la imagen de sus rostros es una de las características distintivas de nuestro sistema visual, donde éste muestra una alta especialización en comparación con los sistemas visuales de otros animales. El análisis del rostro permite la localización y la identificación de humanos, así como la interacción y la comunicación visual con ellos. Por lo tanto, la interacción humano-robot basada en el uso de información facial debería alcanzar la misma eficacia, diversidad y complejidad que tiene la interacción entre humanos.

En la literatura existen múltiples enfoques para abordar el problema de detección y reconocimiento de objetos, en particular el problema de detección y reconocimiento de rostros. La detección de rostros busca encontrar la ubicación y tamaño de los rostros presentes en una imagen arbitraria, por otro lado el reconocimiento de rostros busca determinar la identidad de

las personas presentes en una imagen arbitraria. Entre los enfoques que se han desarrollado en los últimos años en reconocimiento de rostros están [117][125][89][1], estos enfoques van desde los clásicos métodos basados en *Eigenspace* (por ejemplo, *eigenfaces* [68]), hasta sistemas sofisticados basados en la información térmica, o en modelos 3D (ejemplos [1][8][92]). Sin embargo, el problema del reconocimiento de rostros en ambientes complejos no ha sido completamente resuelto [36].

En el ámbito de los sistemas de visión robótica existen varias formas de enfrentar el problema; la primera consiste en ocupar criterios holísticos[1], esto es, considerar el rostro en su conjunto. Esta idea parte de la base que la representación digital de un rostro ya representa una cuantificación del indicador biométrico. Además, se sabe con seguridad que la imagen digital contiene la información necesaria para realizar el reconocimiento. En efecto, las personas son capaces de reconocer un rostro a partir de una imagen digital, desplegada como una matriz de píxeles que cuantifican la intensidad de gris en las zonas respectivas del rostro. Basado en lo anterior, el criterio holístico forma un vector de características usando todo el conjunto de píxeles que forman el rostro. Luego, el aspecto más importante de la metodología holística consiste en ignorar las posiciones relativas entre píxeles, se toma el valor de cada píxel sin importar en que parte del rostro se ubica. De esta forma, un sistema que analiza el conjunto de valores de intensidades de gris, sin saber cuáles son vecinos a cuáles, es incapaz de ver las características locales en la imagen y se ve obligado a analizar los datos en forma conjunta. En otra forma de abordar el problema, las imágenes son representadas como un conjunto de dos dimensiones, usando la intensidad de los valores de los píxeles, estos se comparan con una sola o varias características que representan a todo el rostro, algunos ejemplos de este tipo de métodos son [109][51].

Los modelos holísticos pueden tener un rendimiento muy bueno y la ventaja de un tiempo de procesamiento reducido, pero por otro lado, la dificultad de la construcción de buenos modelos representativos e invariantes a diferentes condiciones en que los rostros están presentes en la escena, hace demasiado difícil encontrar soluciones robustas. En los métodos no holísticos que comparan diferentes características geométricas de las caras usando la intensidad de los valores de los píxeles, tienen mayor robustez a las variaciones en la escena pero la complejidad y los tiempos de procesamiento son mayores.

Por otra parte, hay un creciente interés en los robots de servicio en la comunidad robótica. Según la Federación Internacional de Robótica (IFR), un robot de servicio es un robot que opera de forma parcial o totalmente autónoma, para realizar servicios útiles para el bienestar de los humanos y del equipamiento, excluyendo operaciones de manufactura. Un robot doméstico es un robot diseñado para interactuar con los seres humanos en una casa u otro ambiente similar y para proporcionar diferentes tipos de servicios (limpieza, cocina, entretenimiento, compañía, vigilancia, por nombrar sólo algunos). El ambiente del hogar se define como "cualquier lugar donde la gente vive su vida cotidiana", que puede incluir, por ejemplo, una cocina, un dormitorio, o un jardín. Aunque algunos robots domésticos para usos específicos ya son populares (por ejemplo, los robots aspiradora [46]), todavía estamos lejos de tener robots domésticos para fines generales.

---

[1] La holística se refiere a la manera de ver las cosas enteras, en su totalidad, en su conjunto, en su complejidad, pues de esta forma se pueden apreciar interacciones, particularidades y procesos que por lo regular no se perciben si se estudian los aspectos que conforman el todo, por separado.

Entre las habilidades básicas de robots de servicio doméstico están la capacidad de moverse de forma autónoma en el ambiente doméstico, la capacidad de reconocer y manipular objetos (vasos, libros, anteojos, medicamentos, sillas, manijas de puertas, etc.) y la capacidad de identificar seres humanos e interactuar con ellos utilizando interfaces intuitivas como el habla, los gestos, y la información facial. Esta tesis se centró en la detección e identificación de humanos mediante información visual. Esta tarea es vital en robots domésticos de uso general que son utilizados por usuarios no expertos.

Los sistemas robóticos de servicio actuales suelen tener una limitada capacidad de procesamiento. La idea es concentrarse en la construcción de un método de reconocimiento de rostros que cumpla con las siguientes características: (i) *Operación Online*: Que no existan etapas de entrenamiento o aprendizaje offline. El sistema debe ser capaz de construir la galería[2] en forma incremental completamente online. Este requerimiento es porque se necesita agregar rostros mientras el sistema está funcionando y no se puede detener cada vez para hacer un entrenamiento; (ii) *Operación en Tiempo Real:* el sistema debe ser capaz de responder sin grandes retrasos: El proceso de análisis completo, incluyendo detección, alineamiento y reconocimiento de rostros, debe funcionar al menos a 5 *fps*; (iii) *Un rostro por persona*: Una sola imagen por persona en la base de datos debe ser suficiente para que el sistema de visión tenga un buen rendimiento en la identificación de rostros; y (iv) *Ambiente dinámico:* El sistema debe ser capaz de tener un buen rendimiento en ambientes no controlados (variabilidad en iluminación, fondo, etc.).

El reconocimiento de rostros en ambientes controlados es un problema casi resuelto (según estudios comparativos recientes, ver [117] [126] [89] [1] [82]). Sin embargo, el reconocimiento facial en entornos no controlados es aún un problema abierto [101] [36]. Números especiales de revistas recientes [16], *workshops* [32], y las bases de datos [61] se dedican específicamente a este tema. Los principales factores que aún perturban en gran medida el proceso de reconocimiento de rostros en entornos no controlados son [101] [65]: (i) las condiciones de iluminación variables, especialmente iluminación *outdoor*, (ii) las variaciones de los rostros fuera del plano, y (iii) las variaciones de las expresiones faciales. El uso de sensores más complejos (térmicos, alta resolución, y las cámaras 3D), modelos 3D del rostro, los modelos de iluminación, y conjuntos de imágenes de cada persona (que cubren diversas variaciones de los rostros) son algunos de los enfoques que se utilizan para hacer frente a los problemas antes mencionados [101][65].

A pesar de la gran variedad de enfoques que abordan el problema de detección y reconocimiento de rostros, la mayoría de estos no involucran el uso de información de contexto, algo que es esencial para la mejora y generalización del proceso de detección y reconocimiento de rostros. Los sistemas existentes de detección y reconocimiento de rostros realizan, en su mayoría, el proceso en base a la información previamente conocida sobre él mismo (color, forma, características, etc.), estando prácticamente ausente de la literatura la idea de usar fuentes de información de contexto, de diferentes niveles, como son la coherencia espacial de los objetos, la coherencia temporal de los mismos, el uso de múltiples sensores que complementen la información entre ellos, etc.

---

[2] *Galería* es el conjunto de imágenes de rostros que se les extraen características y se encuentran almacenadas. Se utilizan para reconocer al rostro actual. Estas imágenes están relacionadas a un *ID*.

2.1.2 Definición del problema a abordar

Primero se definirá claramente el término contexto aplicado a sistemas de visión robótica que se manejará en este trabajo. El término "contexto" carece de una definición clara y en la literatura se pueden encontrar diversas definiciones: en [104] es entendido como "cualquier y toda información que pueda influir en la forma se percibe una escena y los objetos dentro de ella", por otro lado, en [17] "la información contextual puede ser cualquier información que no es producida directamente por la presencia de un objeto. Se puede obtener a partir de los datos de imagen, etiquetas o anotaciones de imagen y la presencia o localización de otros objetos". En esta tesis el contexto debe ser siempre visto desde el punto de vista de un rostro o cara, siendo el rostro el elemento en la imagen que se desea reconocer o identificar de alguna manera. Se definirá contexto como: *cualquier información observada o previamente conocida por el robot, que ha sido determinada y/o almacenada en cualquier instante de funcionamiento del sistema y que tenga una relación espacial, temporal, semántica, física u otra, con el objeto de referencia para el cual estamos evaluando su contexto*. Los principales sensores que se considerarán en este trabajo serán una cámara de video y una cámara térmica. La información obtenida desde estas fuentes se complementará con la de los sistemas de posición que normalmente poseen los sistemas robóticos, por lo que el contexto considerado estará restringido fundamentalmente a la información visual y de posición, junto con la información previamente conocida que se le pueda entregar al robot en función del problema que debe resolver.

Otro término que se definirá será el de visión activa, éste se refiere a tener un sistema donde exista una realimentación entre los módulos de estrategia-actuación y visión. De esta forma se pueden tomar decisiones con el objetivo de modificar las percepciones realizadas por el agente y manejadas por el sistema de visión. Por ejemplo, en el caso de un robot de servicio, un sistema de visión activa podría otorgarle al sistema robótico que desempeña la tarea de reconocimiento, la capacidad de tomar la decisión de mirar en la dirección más probable hacia donde se encuentra una persona, realizar un seguimiento de rostros, modificar una observación acercándose al usuario, etc. Otro término importante es la multiresolución, que se refiere a la capacidad de analizar sólo partes de la imagen original donde se establece que hay una alta probabilidad de encontrar un objeto de interés, esta zona es analizada con mayor interés y utilizando diferentes tamaños de imágenes. Mediante la aproximación de las posiciones y velocidades de los objetos y del sistema robótico, se pueden determinar para cada imagen, los sectores donde es más probable encontrar algún objeto. Este análisis permite acelerar el proceso de visión.

Esta tesis se enfocará en resolver el problema de la detección y el reconocimiento de rostros. La detección de rostros busca encontrar la ubicación y tamaño de los rostros presentes en una imagen arbitraria, por otro lado el reconocimiento de rostros busca determinar la identidad de las personas presentes en una imagen arbitraria. Para esto se construirá un sistema de visión para robots humanoides que opere en ambientes no controlados, y sea capaz de detectar y reconocer rostros humanos en forma robusta. De los diferentes niveles de contexto que se considerarán en esta tesis se analizará cuales mejoran el rendimiento del sistema de detección y reconocimiento de rostros en ambientes no controlados. Se diseñará un modelo para la integración de la información de los diferentes niveles de contexto considerados. Se mejorarán los diferentes módulos del proceso de visión, utilizando la información contextual y se integraran etapas de visión activa y multiresolución.

Se considerarán los siguientes niveles de contexto [38]:

- **Contexto de bajo nivel (nivel de píxeles)**: Se refiere a la información de contexto aportada por los píxeles del entorno a un grupo de pixeles de referencia que actúa como objeto de interés. Por ejemplo, este tipo de contexto es utilizado en el Capítulo 6 en el análisis térmico cuando se realiza detección de piel, ya que se determinan las regiones de las imágenes que contienen piel utilizando el algoritmo de segmentación de piel *Skindiff* que utiliza la información de la vecindad para lograr robustez.

- **Contexto físico espacial**: Se refiere a la información determinada a través de las leyes físicas que rigen el ambiente donde opera el robot. El modelo físico del ambiente define un modelo visual, por ejemplo la existencia de un piso sobre el cual deben estar los objetos que no tienen la capacidad de volar, la existencia de un vector de gravedad, entre otras.

- **Contexto de configuración de objetos (coherencia espacial)**: Un objeto puede tener asociada una relación espacial específica con otros objetos. Por ejemplo un rostro está generalmente sobre un cuello y sobre los hombros.

- **Contexto de la situación**: Saber qué actividad se está realizando, en conjunto con la escena o ambiente en que se está inmerso, determinan un contexto a nivel de la situación en que se encuentra el robot. Por ejemplo, la situación puede ser: "navegando, en ambiente *casa*, el 24 de enero", en este caso se puede utilizar un mapa de la habitación. Este nivel de contexto puede ser ingresada por el usuario previamente para aportar información extra. Por ejemplo, si el robot se encuentra buscando personas dentro de un ambiente doméstico, la información respecto a si las personas están sentadas o paradas puede ayudar a la eliminación de detecciones falsas fijando parámetros del contexto físico espacial.

Considerar la información de contexto tendrá como objetivo mejorar diferentes etapas del sistema de detección y reconocimiento de rostros. Por una parte, cada nivel de contexto realimentará a los diferentes módulos del sistema de visión, mejorando una tarea en particular. Por otro lado, el sistema de integración de la información contextual realimentará la decisión final de la detección de cada objeto. Finalmente, los niveles de contexto podrán realimentar a los sistemas de reconocimiento y caracterización de los objetos. En particular, se trabajará extensamente en el módulo de reconocimiento de rostros, ya sea en el espectro visible como en el espectro térmico. Se integrará toda esta información en un sistema de decisión contextual y se diseñarán los sistemas de visión activa y multiresolución.

2.2 Objetivos

2.2.1 Objetivo General

Diseñar e implementar un sistema de visión para robots de servicio que opere en ambientes no controlados, y sea capaz de detectar y reconocer rostros humanos en forma robusta, utilizando métodos de visión activa e integrando diferentes tipos de contexto.

2.2.2 Objetivos específicos

- Definir que tipos de contexto son relevantes para el desarrollo del sistema de visión propuesto y definir la metodología necesaria para el tratamiento de la información y la integración de los diferentes niveles de contexto.
- Desarrollar los distintos módulos de análisis de la información visual, integrando los tipos de contexto que estén relacionados con el problema.
- Desarrollar e integrar un módulo de reconocimiento de rostros que integre los módulos desarrollados anteriormente en un sistema de visión robótica robusto.
- Probar y evaluar el sistema de visión propuesto en un robot de servicio en alguna prueba estándar.
- Construir una base de datos donde las imágenes representen mejor el ambiente en donde se desenvolvería un robot de servicio.
- Desarrollar un simulador en el que se puedan generar las observaciones de un agente dentro de un mapa con personas, que posteriormente se pueda utilizar para la evaluación del sistema de visión propuesto.

2.3 Hipótesis

- El uso e integración de información proveniente de diferentes tipos de contexto puede mejorar el rendimiento del sistema de detección y reconocimiento de rostros y objetos, sin que esto signifique un aumento significativo en el tiempo de procesamiento requerido:
    - La información proporcionada por el contexto físico, como la información inherente al objeto (forma, tamaño, color, etc.), puede ser utilizada para definir cualidades de los rostros que se quieren detectar o reconocer. Por otro lado, las restricciones a las que está sujeto el objeto (leyes físicas naturales), pueden ser utilizadas para descartar objetos en configuraciones que no cumplen con estas leyes (rostros volando, bajo la línea del horizonte, etc.). Esto se puede aplicar a las imágenes en espectro visible y térmico.
    - La información sobre la situación en la que está inmerso el robot puede ser utilizada para mejorar la detección y reconocimiento de rostros, ya que se conoce que tarea se está realizando y cuales son las restricciones del ambiente en que se encuentra inmerso el robot.
    - La información proporcionada por el contexto de configuración de objetos, o coherencia espacial, puede ser utilizada para descartar detecciones falsas ya que se conocen las relaciones físicas que deben cumplir los objetos.
    - La información aportada en las imágenes en espectro térmico puede complementar a la información obtenida con las imágenes en espectro visible. Permitiendo descartar detecciones falsas y aportando información cuando las condiciones de luminosidad no sean las óptimas.
- Retroalimentar el módulo de actuación utilizando la información obtenida desde el nuevo sistema de visión permite modificar las observaciones realizadas y con ello mejorar el rendimiento del sistema de visión robótica (Visión Activa).

2.4 Estructura de la tesis

La tesis está estructurada de la siguiente forma: El Capítulo 3 presenta una revisión bibliográfica de los métodos de reconocimiento de rostros, del uso de contexto en diferentes aplicaciones y el uso de cámaras térmicas. En el Capítulo 4 se presentará la metodología con que se construye el sistema de visión propuesto y los módulos principales que lo componen. En el Capítulo 5 se presenta un entorno virtual para entrenar y probar sistemas de detección de rostros en condiciones no controladas que se utiliza para validar el funcionamiento del sistema de visión propuesto y evaluar las diferentes hipótesis de esta tesis, además de realizar un estudio comparativo de los módulos detección y reconocimiento de rostros. Luego en el Capítulo 6 se analiza el sistema de visión propuesto evaluando el funcionamiento de los diferentes módulos de contexto implementados. Además se prueba el sistema en una aplicación real, en este caso se realiza una prueba enmarcada en la competencia de robótica RoboCup [96], que se llama "*Who is Who*". Se propone un sistema robusto para la detección y la identificación de los seres humanos en entornos domésticos para ser usado en un robot de servicio. La detección robusta de personas se logra mediante el uso de fuentes de información térmica y visual que se integran para detectar objetos que son *candidatos a humanos*. Estos candidatos son procesados con el fin de verificar la presencia de los seres humanos y su identidad con la información frente a los espectros térmico (*ET*) y visible (*EV*). Finalmente, el Capítulo 7 presenta conclusiones, discusiones y el trabajo futuro.

Capítulo 3
Trabajo Relacionado

Aquí se presenta una amplia revisión bibliográfica. Esta sección está dividida en diferentes áreas que se abordarán en esta tesis. Muchos de los textos citados servirán de referencia y se utilizarán en el desarrollo de este trabajo.

3.1 Sistema visual humano

En general, el sistema humano de reconocimiento de rostros utiliza un amplio espectro de estímulos, obtenidos a partir de muchos, si no todos, los sentidos que posee el ser humano (visual, auditivo, olfativo, táctil y gusto). Estos estímulos se utilizan de forma individual o colectiva para el almacenamiento y recuperación de imágenes del rostro. En muchos casos se utiliza también el conocimiento contextual, es decir, el entorno juega un papel importante en el reconocimiento de rostros en relación a donde se supone que se encuentren. Es inútil (utilizando la tecnología existente) tratar de desarrollar un sistema que puede imitar *todas* estas notables capacidades de los seres humanos. Sin embargo, el cerebro humano tiene sus limitaciones en el número total de personas que puede "recordar" con precisión. Una de las principales ventajas de un potencial sistema informático es su capacidad para manejar grandes conjuntos de datos con imágenes de rostros. En la mayoría de las aplicaciones, las imágenes son en uno o múltiples puntos de vista 2-D, lo que obliga a los algoritmos a sólo utilizar información visual.

Existen muchos estudios de psicología y neurociencia [94][103][23][10][102] que tratan el tema del reconocimiento y las conclusiones de estos estudios tienen mucha relevancia para las personas interesadas en el diseño de algoritmos o sistemas de reconocimiento ya que son utilizadas para mejorar los sistemas implementados. Por otro lado, un mejor sistema de reconocimiento puede proporcionar mejores herramientas para la realización de estudios en la psicología y la neurociencia [85].

3.2 Sistemas de reconocimiento de rostros

Durante los últimos años el problema del reconocimiento de rostros se ha convertido en uno de los temas de investigación más activos entre las aplicaciones de reconocimiento de patrones. El interés en el tema ha sido potenciado por diversas aplicaciones: Vigilancia, interacción humano-robot, seguridad, etc. Aun cuando los sistemas de reconocimiento de

rostros han mostrado una evolución importante en algunos *benchmarks* [77][79], el problema del reconocimiento de rostros está lejos de ser un problema resuelto.

Aun cuando existen muchas publicaciones que hacen comparaciones entre distintos métodos de reconocimiento de rostros [77][117] [50][19], elegir cual es el mejor es difícil, ya que no hay que evaluar sólo el método en las bases de datos estándar [77], sino que hay que evaluar los métodos bajo diferentes condiciones (iluminación, fondo, ruido, etc.). Para seleccionar el método que se utilizará para reconocer los rostros se realizará una evaluación exhaustiva de los mejores métodos, y ver cual tiene un mejor rendimiento bajo diferentes condiciones.

Actualmente los métodos de reconocimiento pueden dividirse en varias categorías. La primera consiste en ocupar criterios holísticos, esto es, considerar el rostro en su conjunto. Esta idea parte de la base que la representación digital ya representa una cuantificación del rostro, el sistema analiza el conjunto de intensidades de gris sin saber la relación que hay entre ellos, por lo que es incapaz de ver características locales y analiza los datos en forma conjunta. En la segunda categoría las imágenes son representadas como un conjunto de dos dimensiones, usando el valor de los píxeles directamente, el rostro es representado por una o más características locales (*calces locales*) y luego se comparan estas características utilizando diversas medidas de distancia. Otra categoría es utilizar un modelo 3D del rostro, trabajos como [57][129][88] en que se construyen modelos 3D de los rostros, el problema de estos métodos es que son costosos computacionalmente, por lo que no son viables para una aplicación online como la que se quiere implementar.

Después de la aparición de PCA [66][68], los métodos holísticos, que usan la región completa del rostro como entrada al sistema, fueron extensamente estudiados. Estos métodos básicamente usan *análisis de componentes principales* (PCA – *Principal Component Analysis*) [66][68][110][74][118], *análisis de discriminante lineal* (LDA – *Linear Discriminant Analysis*) [80][54][21][15] o *análisis de componentes independientes* (ICA – *Independent Component Analysis*) [67]. Los rostros son proyectados a un espacio de menor dimensión para manejar el problema de la dimensionalidad.

Otro método es EBGM (*Elastic Bunch Graph Matching*). Este método se basa en que todos los rostros humanos comparten una estructura topológica similar. Los rostros están representados en forma de gráficos, con los nodos situados en los puntos fiduciales (ojos, nariz, etc.) y los bordes están marcados con vectores de distancia 2-D. Cada nodo contiene un conjunto de 40 coeficientes *Gabor wavelets* complejos en diferentes escalas y orientaciones (fase y amplitud). Se les llama "*jets*". El reconocimiento se basa en gráficos de etiqueta. Un gráfico de etiqueta es un conjunto de nodos conectados por los bordes, los nodos están etiquetados con *jets*, los bordes están etiquetados con distancias [59][60].

Otros métodos de reconocimiento de rostros son los de aprendizaje estadístico, entre los cuales encontramos a SVM [35], dado un conjunto de puntos que pertenecen a dos clases, una máquina de soporte vectorial (SVM) encuentra el hiperplano que separa la fracción más grande posible de puntos de la misma clase en el mismo lado, al mismo tiempo que maximiza la distancia entre las dos clases. Se utiliza PCA para extraer las características de imágenes de rostros y luego las funciones de discriminación entre cada par de imágenes son aprendidas por un SVM. Otro método conocido es *boost* [127], la idea detrás de *Boosting* es emplear una secuencia de clasificadores débiles como un clasificador fuerte. Viola&Jones construyó el primer sistema de detección de caras en tiempo real usando *Adaboost* [83][84].

Recientemente, se han mostrado resultados muy prometedores en el reconocimiento de rostros en métodos de tipo *calces locales*, [109][124][90][111][3][4][56]. La idea general de los métodos de *calces locales* es primero encontrar varias características locales y luego clasificar los rostros comparándolos y combinándolos con las correspondientes características locales.

Luego de una intensa búsqueda en la bibliografía se eligieron varios métodos que cumplen con las características necesarias mencionadas en el capítulo anterior: Operación online, operación en tiempo real, un rostro por persona, y que funcione en ambientes dinámicos. Los métodos elegidos para ser analizados durante el desarrollo de esta tesis son:

- **Histogramas LBP**: El reconocimiento de rostros mediante histogramas de características LBP (*local binary patterns* [9]) fue propuesto inicialmente en [109]. En el planteamiento original se definen tres niveles diferentes de localidad: a nivel de píxel, a nivel regional y a nivel global. Los dos primeros niveles de la localidad se realizan dividiendo la imagen LBP (imagen original con trasformada LBP) del rostro en pequeñas regiones; se extraen características utilizando histogramas, estos histogramas se utilizan para una eficiente representación de la información de textura (ver Figura 1). El nivel global de localidad, es decir, descripción del rostro, se obtiene concatenando las características LBP locales.

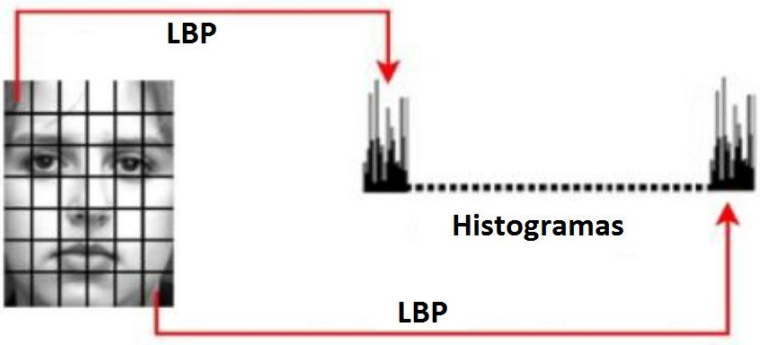


Figura 1: Histogramas LBP.

El operador LBP (ver Figura 2) es uno de los mejores descriptores de textura y ha sido utilizado en varias aplicaciones, inicialmente fue diseñado para la descripción de texturas y luego se adaptó al reconocimiento de rostros. Este operador posee ventajas como su invarianza a cambios monótonos en niveles de grises y además su eficiencia computacional, lo cual es importante en procesamiento de imágenes. Principalmente la idea de usar el operador LBP para el reconocimiento de rostros es motivada en que los rostros pueden ser vistos como una descomposición de micro-patrones los cuales pueden ser descritos por este operador.

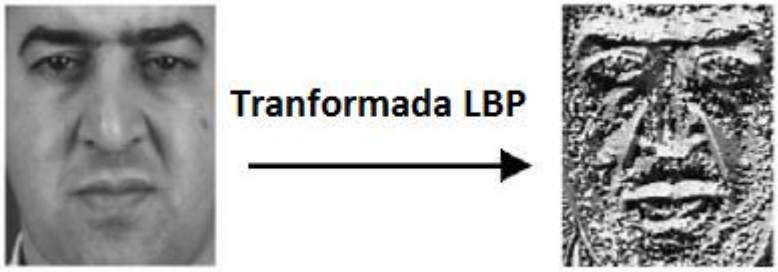


Figura 2: Transformada LBP.

La transformación usando el operador LBP asigna una etiqueta a cada pixel de la imagen comparando el pixel central con sus vecinos de una ventana de 3x3 generando como resultado un código binario. Así el histograma obtenido de las etiquetas puede ser usado como un descriptor de textura. El esquema clásico de la descripción del operador LBP es mostrado en la Figura 3.

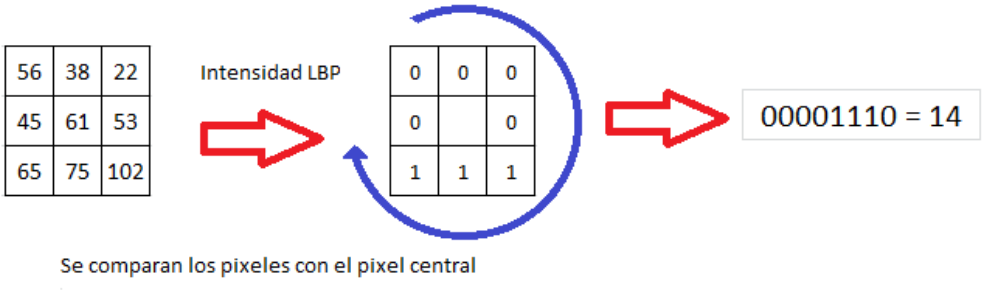


Figura 3: Características LBP.

En este método se utiliza una extensión del método LBP llamada "patrones uniformes". A la imagen resultante del operador LBP se le aplica la detección de patrones uniformes, la cual consiste en encontrar patrones binarios que contengan como máximo 2 transiciones de 0 y 1 o viceversa. Ejemplos de patrones uniformes son 00000000 (0 transiciones) y 11001111 (2 transiciones). Sin embargo, el patrón 11001001 (4 transiciones) no lo es. Luego estos patrones uniformes son usados de tal manera de construir un histograma, donde cada bin del histograma es llenado con un patrón uniforme y todos los patrones no uniformes son asignados en un sólo bin.

El histograma LBP contiene información acerca de la distribución de los micro-patrones locales del rostro, tales como líneas, puntos y diferentes áreas del rostro. Para una representación eficiente del rostro, se debe mantener también la información espacial del rostro. Por este motivo es que la imagen del rostro es dividida en diferentes regiones rectangulares donde es calculado el histograma sobre los códigos LBP obtenidos en cada región. Finalmente, los histogramas de cada región son concatenados en uno sólo que representa la imagen del rostro.

El reconocimiento se realiza utilizando un clasificador "vecino más cercano" usando una de las siguientes medidas de similitud: histograma intersección, de log-verosimilitud y estadística de χ^2 (Chi cuadrado).

Intersección de Histograma:

$$D(S,M) = \sum_i \min(S_i, M_i) \qquad (3.1)$$

Logaritmo de la verosimilitud estadística:

$$L(S,M) = -\sum_i S_i \log M_i \qquad (3.2)$$

Estadística χ^2 (Chi cuadrado):

$$\chi^2(S,M) = \sum_i \frac{(S_i - M_i)^2}{S_i + M_i} \qquad (3.3)$$

- **Descriptores Gabor Jets**: En [51] se comparan diferentes métodos de reconocimiento de rostros que utilizan *calces locales*. El estudio analiza varias formas de extracción de características, métodos de clasificación, y combinaciones de diferentes clasificadores. Teniendo en cuenta los resultados del estudio, los autores construyen un sistema que integra la mejor elección posible en cada paso. Este sistema utiliza Gabor Jets [9] como descriptores locales (se utilizan 5 escalas y 8 orientaciones de los filtros de Gabor).

  Los filtros Gabor son sinusoides espaciales localizadas por una ventana gaussiana que permiten la extracción de características en las imágenes por medio de la selección de su frecuencia, orientación y escala.

  Un filtro Gabor se define matemáticamente como:

$$\psi_{\theta,\lambda}(z) = \frac{\left\| k_{\theta,\lambda} \right\|^2}{\sigma^2} e^{-\frac{\left\| k_{\theta,\lambda} \right\|^2 \|z\|^2}{2\sigma^2}} \left[e^{ik_{\theta,\lambda} z} - e^{-\frac{\sigma^2}{2}} \right] \qquad (3.4)$$

  donde θ y λ se definen como la orientación y escala del filtro Gabor respectivamente, $z = (x, y)$, $\| \|$ representa el operador norma, y el vector de onda $k_{\theta,\lambda} = \frac{k_{\max}}{\lambda} e^{i\theta}$, $k_{\max}$ representa la frecuencia máxima. La escala es dada por $\lambda = 4\sqrt{2^n}$ con $n = 0,..,4$ y la orientación es dada por $\theta = n\pi / 8$ con $n = 0,..,7$.

  En la representación de características Gabor [9], sólo las magnitudes Gabor son usadas porque la fase del filtro cambia linealmente con pequeños desplazamientos. Distintos filtros Gabor son construidos en base a distintos tamaños de σ. Con $\sigma = \lambda$, los tamaños de las ventanas para el filtro Gabor son $6\sigma + 1$ desde el centro de la ventana.

  Se define un Gabor jet como un conjunto de 8 filtros Gabor con la misma escala λ y posición (x, y) pero con distinta orientación θ. Cada Gabor jet es localizado uniformemente sobre la imagen, separado por una longitud de onda λ.

  La idea principal en la extracción de características usando Gabor jet es construir una grilla en la imagen espaciada en λ. De esta forma se pueden calcular los filtros Gabor centrados en cada nodo de la grilla y así obtener el Gabor jet de manera eficiente y rápida debido a que no se usa la imagen completa sino sólo los puntos de la grilla. Cada jet es representado por 8 componentes en cada nodo. La cantidad de jet obtenidos en la imagen depende del tamaño de la región a usar, por ejemplo si se usa un área de 37x37, se encuentran 110 jets: 64 jets para $\lambda = 4$, 25 jets para $\lambda = 4\sqrt{2}$, 16 jets para $\lambda = 8$, 4 jets para $\lambda = 8\sqrt{2}$ y 1 jet para $\lambda = 16$.

  Cada uno de los jets obtenidos anteriormente es comparado con todos los candidatos usando el producto interno normalizado y los resultados son combinados usando el método *Borda Count* [9]. La idea de usar *Borda Count* es obtener un puntaje final para usarlo como una medida de distancia. Si se quiere obtener cual es la imagen de la galería más parecida a una imagen de test, se compara el jet de cada nodo de la grilla de la imagen de test con el mismo nodo en cada una de las imágenes de la galería usando el producto interno normalizado. Usando *Borda Count* se genera un ranking para cada uno de los nodos de la grilla. Finalmente se suman los resultados obtenidos en cada

imagen de la galería y se obtiene un puntaje el cual es tomado como una medida de distancia.

En resumen se tiene una grilla con puntos que están uniformemente distribuidos a través de las imágenes, separados por una *longitud de onda*. En cada posición de la grilla, en cada escala y para cada imagen se calculan los Gabor Jets, luego se comparan utilizando producto punto normalizado, y finalmente estos resultados se combinan mediante el método *Borda Count*.

- **Descriptores SIFT (scale-invariant feature transform)** [22]**:** Los métodos que utilizan SIFT son cada vez más populares y han experimentado un desarrollo impresionante en los últimos años. Este método se basa en el uso de puntos de interés locales como descriptores. Normalmente, los puntos de interés locales son extraídos de forma independiente, en la imagen de prueba y una imagen de referencia, a continuación, los descriptores se comparan y se obtiene una transformación entre las dos imágenes. El sistema desarrollado por Lowe [22] utiliza los descriptores SIFT y una hipótesis probabilística en la etapa de rechazo. Este sistema es una opción popular en las aplicaciones de reconocimiento de objetos, por su capacidad de reconocimiento, y funcionamiento en tiempo real. Sin embargo, tiene un inconveniente: el gran número de detecciones falsas. Este inconveniente puede ser superado por la utilización de varias etapas de rechazo, como por ejemplo, en el sistema L&R [49]. Este sistema ya se ha utilizado en la construcción de sólidos sistemas de verificación de huellas dactilares [49] y verificación off-line de firmas [47].
- **Histogramas WLD (Weber Local Descriptor)**[3]: El método WLD, es un descriptor simple, poderoso y robusto, el cual está basado en la ley de Weber [52]. La ley de Weber se basa en el hecho que la percepción humana de los patrones no depende solamente del cambio en el estimulo (tal como sonido o iluminación) sino también de la intensidad original del estimulo. Específicamente, WLD consiste en 2 componentes: excitación diferencial y orientación. La componente de excitación diferencial es una función del radio entre 2 términos: el primero es la diferencia relativa de intensidad del actual pixel contra sus vecinos y el otro es la intensidad del propio pixel actual. El componente de orientación es la orientación del gradiente del pixel actual. Para una imagen dada, se usan las 2 componentes del método WLD para construir y concatenar el histograma WLD.

  El algoritmo WLD utiliza el cálculo de 2 componentes: excitación diferencial y orientación. Para luego construir histogramas. La Figura 4 muestra el cálculo del descriptor WLD.

  En la excitación diferencial se utilizan las diferencias entre el pixel actual y los vecinos como los cambios del pixel actual. Se intenta encontrar variaciones salientes dentro de la imagen con el fin de simular la forma en que percibe el ser humano.

  Así la excitación diferencial $\xi(x)$ se calcula como:

$$\xi(x_c) = \arctan\left[\sum_{i=0}^{p-1}\left(\frac{x_i - x_c}{x_c}\right)\right] \qquad (3.5)$$

[3] Método de reconocimiento de rostros presentado en [20].

donde $x_i (i = 0,1,.., p-1)$ representa el vecino i del pixel central x_c y p representa el número total de vecinos.

El componente orientación de WLD es la orientación del gradiente, la cual se calcula como:

$$\theta(x_c) = \arctan\left(\frac{v_s^{11}}{v_s^{10}}\right) \qquad (3.6)$$

donde $v_s^{10} = x_5 - x_1$ y $v_s^{11} = x_7 - x_3$.

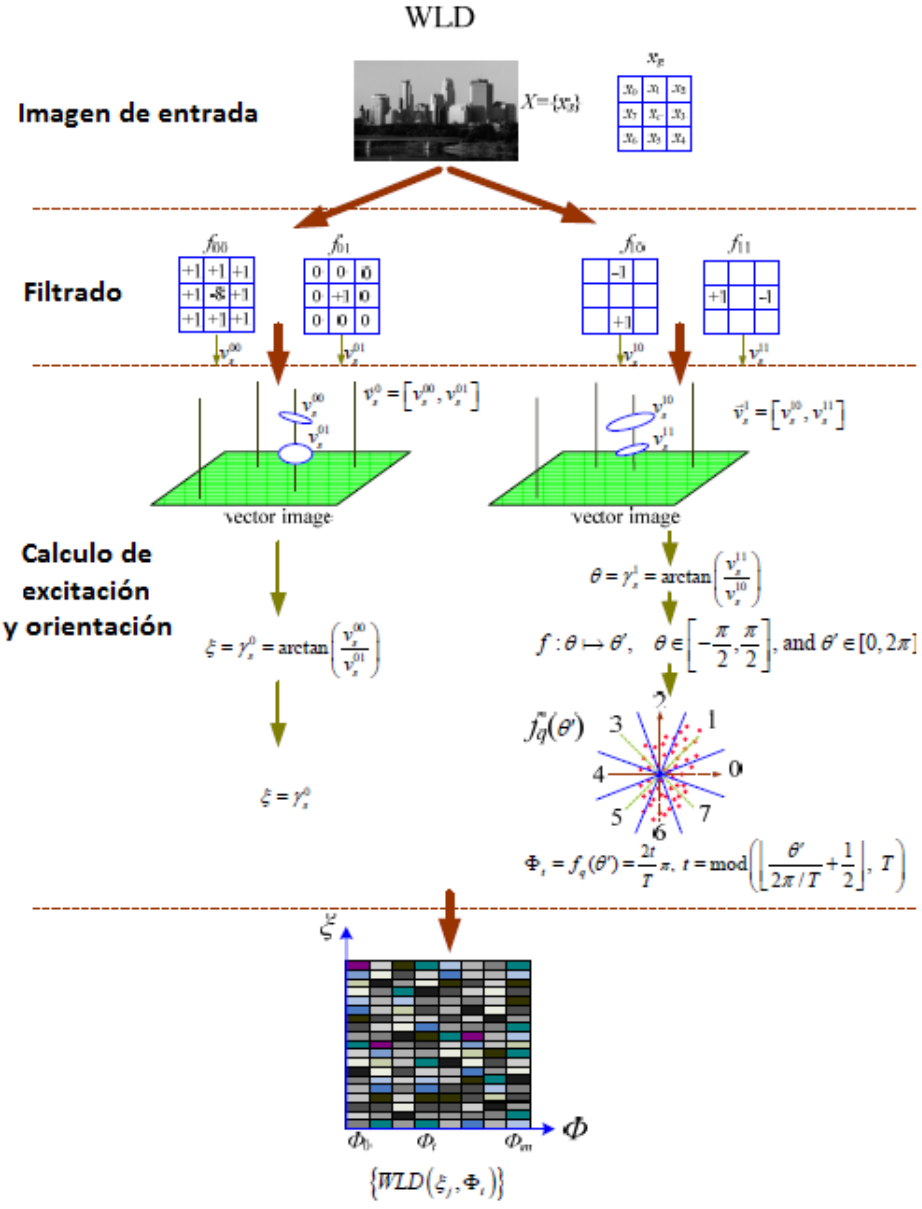


Figura 4: Calculo de descriptor WLD (Tomado de [52]).

Por simplicidad, la orientación θ se cuantiza en T orientaciones dominantes. Antes de la cuantización, se realiza un mapeo de $\theta \in [-\pi/2, \pi/2]$ a $\theta' \in [0, 2\pi]$, de la siguiente forma:

$$\theta' = \arctan 2\left(v_s^{11}, v_s^{10}\right) + \pi \qquad (3.7)$$

donde

$$\arctan 2\left(v_s^{11}, v_s^{10}\right) = \begin{cases} \theta, & v_s^{11} > 0 \; y \; v_s^{10} > 0 \\ \pi - \theta, & v_s^{11} > 0 \; y \; v_s^{10} < 0 \\ \theta - \pi, & v_s^{11} < 0 \; y \; v_s^{10} > 0 \\ -\theta, & v_s^{11} < 0 \; y \; v_s^{10} > 0 \end{cases} \qquad (3.8)$$

De esta forma la función de cuantización es representada como:

$$\Phi_t = \frac{2t}{T}\pi, \quad t = \mathrm{mod}\left(\left[\frac{\theta'}{2\pi/T} + \frac{1}{2}\right], T\right) \qquad (3.9)$$

El cálculo del histograma es realizado como lo muestra la Figura 5. El histograma 2D $\{WLD(\xi_j, \Phi_t)\}$, $j = 0,1,...,N-1$, $t = 0,1,...,T-1$, donde N es la dimensión de la imagen y T es el número de orientaciones dominantes, es construido colocando en cada columna la orientación dominante y en cada fila al histograma de la excitación diferencial con C bines.

Para obtener un descriptor más discriminativo, el histograma 2D $\{WLD(\xi_j, \Phi_t)\}$ es convertido a un histograma 1D llamado H. Tal como muestra la Figura 5, cada columna del histograma 2D es proyectada a un histograma $H(t)$, $t = 0,1,...,T-1$. Se reagrupan las excitaciones diferenciales ξ_j en T sub-histogramas $H(t)$, correspondiendo a cada una orientación dominante. Cada sub-histograma $H(t)$ es dividido en M segmentos $H_{m,t}$ para formar una matriz de histogramas. Cada columna de la matriz de histogramas corresponde a la orientación dominante y cada fila al segmento de excitación diferencial. Luego la matriz de histogramas es reorganizada como un histograma 1D H. Cada fila de la matriz es concatenada como un sub-histograma donde luego son concatenados para originar el histograma 1D $H = \{H_m\}, m = 0,1,...,M-1$.

Cada sub-histograma $H(t)$ es eventualmente dividido en M segmentos, el rango de la excitación diferencial ξ_j también es dividido en M intervalos. Además cada segmento de sub-histograma $H_{m,t}$ es compuesto por S bines, es decir $H_{m,t} = \{h_{m,t,s}\}, s = 0,1,...,S-1$. El número de celdas C en cada columna del histograma 2D $\{WLD(\xi_j, \Phi_t)\}$ es dado por $C = MxS$.

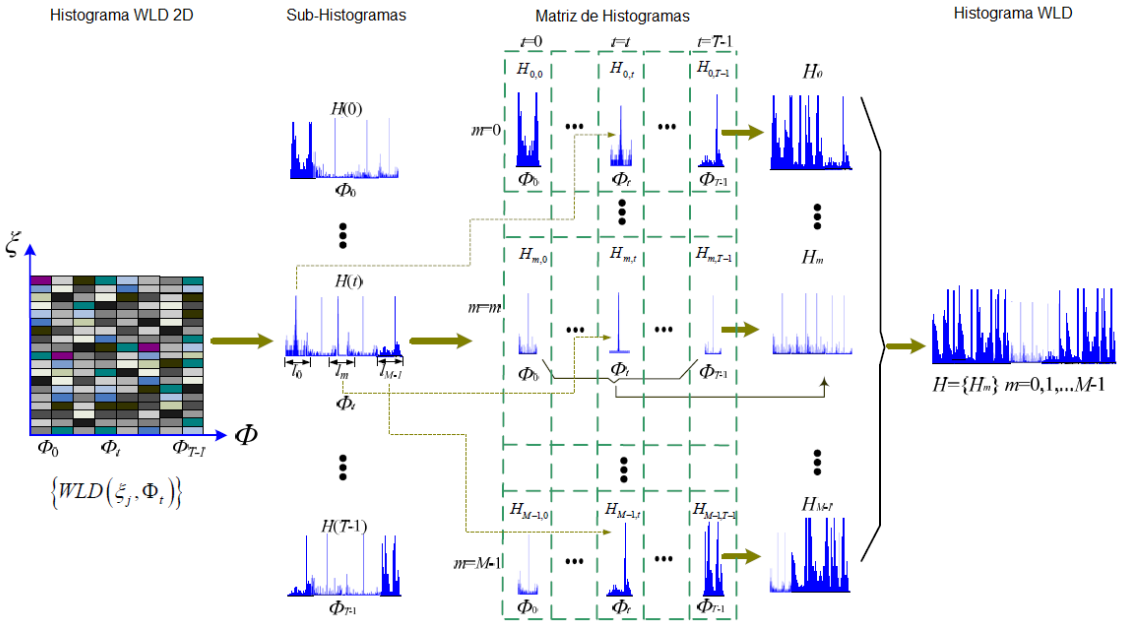


Figura 5: Calculo histograma WLD (Tomado de [52]).

El reconocimiento es realizado de la misma forma que en el caso del método LBP ya que se usa un clasificador de vecino más cercano en el espacio de características. Se usan las mismas 3 medidas de disimilitud: Intersección de Histograma, Logaritmo de la verosimilitud estadística y Estadística Chi cuadrado

- **ERCF (*Extremely Randomized Clustering Forest*)**: En [27] se propone un método para aprender una medida de similitud, lo que permite discriminar si un par de imágenes de objetos corresponden o no al mismo objeto (los objetos en particular pueden ser rostros). El método está diseñado para ser utilizado en problemas de reconocimiento de objetos, y hace uso de ERCF [27] y descriptores SIFT [22]. El aprendizaje se realiza con clases de objetos específicos, tales como rostros frontales o autos. El método consiste básicamente en tres etapas. En la primera etapa, se seleccionan los pares de parches (recortes de partes de los objetos) similares, medido en términos de una correlación cruzada normalizada. En la segunda etapa, cada par de parches es codificado (cuantificado) por medio de un ERCF de descriptores SIFT. ERCF es una representación de la imagen que se construye utilizando árboles de clasificación. Cada árbol de clasificación se genera utilizando los descriptores SIFT y se utiliza para codificar el par de parches. En la tercera fase, los pares de parches codificados se utilizan para construir un vector de características, que es finalmente utilizado para evaluar la similitud de la imagen utilizando un clasificador lineal.

3.3 Contexto

En general, el sistema de reconocimiento de rostros humanos utiliza un amplio espectro de los estímulos (visuales, auditivos, olfativos, táctiles) [2][39][45]. Estos estímulos se utilizan de forma individual o colectiva para el almacenamiento y recuperación de imágenes del rostro. El sistema visual humano (SVH) utiliza distintos niveles de contexto para las actividades que realiza cada día. Un ejemplo de esto es que los humanos podemos decir casi sin equivocarnos al mirar una imagen si una zona es piel o no. Primero analizamos presencia de personas, luego buscamos zonas donde debería haber piel (cabeza, brazos, manos, etc.), pero si nos

preguntaran si un píxel dado es o no piel sin ver el **contexto** en que está inmerso, posiblemente nos equivocaríamos.

En la literatura se plantea distintas formas de utilizar el contexto dependiendo de las aplicaciones, por ejemplo [5][95]. En los sistemas de visión artificiales actuales, fundamentalmente en el dominio de robótica y más específicamente en el contexto de la RoboCup, están recién comenzando a considerar el contexto presente en la información visual, capturada por la cámara de video, como metodología para mejorar el rendimiento de las detecciones y la velocidad de los procesos. En este sentido, los sistemas propuestos en la literatura se han enfocado generalmente a utilizar algún nivel de contexto particular y no en usar los múltiples niveles de contexto presentes en una secuencia de video y en cierto conocimiento a priori y/o aprendido de los problemas abordados.

En [38] se propone una primera aproximación para abordar el problema de considerar el contexto en forma general, tratando de unificar los diferentes niveles que se consideraron relevantes para el problema de jugar fútbol en la RoboCup. La solución propuesta en [108], es una primera aproximación hecha en 1993 y es relativamente limitada respecto de los niveles de contexto utilizados. La solución propuesta en [38], por otra parte, intenta generalizar aún más y se basa en establecer un modelo probabilístico en dos etapas:

- La primera etapa consiste en modelar cada nivel de contexto en forma independiente, donde se consideran fundamentalmente: (i) contexto de coherencia entre objetos detectados en un mismo cuadro; (ii) contexto temporal de coherencia entre objetos detectados en diferentes cuadros; (iii) contexto con una representación global formada en el largo plazo. Se establece, por otra parte, la importancia de considerar a futuro otros niveles de contexto como son: (a) contexto holístico o conocimiento sobre la escena observada (situación actual); (b) contexto de alto nivel en las acciones (actividad que está realizando), considerando el hecho de que se está realizando un determinado tipo de acción (un robot podría jugar fútbol, hacer las tareas del hogar o navegar en un ambiente desconocido mediante un mismo programa basado en rutinas comunes); (c) contexto de alto nivel sobre la configuración funcional de los objetos, es decir, utilizar la información conocida o aprendida sobre las configuraciones normales de los objetos (los objetos no flotan, los objetos se mueven o no se mueven).
- La segunda etapa del sistema consiste en integrar esta información. La solución al respecto dada en [38] es un sistema heurístico, que promedia ponderadamente los niveles de contexto considerados. Sin embargo, queda abierta la posibilidad de estudiar mejores soluciones para esta etapa, además de ver de qué manera, cada nivel de contexto, puede ayudar a diferentes etapas del proceso de reconocimiento y no sólo ser considerado como un conjunto que debe ser integrado completamente al final del proceso.

Respecto de los trabajos que utilizan algún nivel de contexto en forma independiente, para apoyar un sistema de detección, la mayoría de éstos utilizan información holística de la imagen. En [6][7], se propone una metodología que utiliza un análisis de contexto en base a información holística calculada con una imagen de entrada. En [6], por ejemplo, se utiliza como información holística una imagen derivada de una transformada de Fourier en ventanas sobre la imagen original. Esta información define vectores de características. Mediante esta información y utilizando un modelo probabilístico sobre la existencia, posición y escala de los objetos en función del vector de características determinado holísticamente con la imagen de entrada, se determinan los sectores más probables donde encontrar los objetos de interés y su escala más probable en la escena, con el fin de circunscribir la búsqueda en dichos sectores de

la imagen. Esto puede ser una primera etapa de un sistema de visión multiresolución. En [5] se presenta un sistema que intenta contextualizar la escena donde se está operando, para definir que objetos son más probables de encontrar dado el lugar donde se está.

En [55] se propone también el uso de contexto aportado por la escena completa donde se está buscando un objeto como en [6], sin embargo, se introduce un nivel de contexto que podríamos denominar como contexto de configuración de objetos. Las detecciones de objetos particulares, que en inicio están apoyadas por la caracterización holística de la escena, apoyan, a su vez, la detección de objetos relacionados con los ya encontrados (por ejemplo, la distancia entre dos objetos que debe ser constante). De hecho, en este trabajo queda abierto el problema de cómo determinar el orden de detección para hacer más eficiente esta búsqueda.

En general, en la literatura se muestra que algunos niveles de contexto en particular son importantes para mejorar el rendimiento de los sistemas de visión [38]. Sin embargo, es escasa la literatura sobre la integración de diversos niveles de contexto, así como sobre una metodología para un sistema de visión basado en contexto. En esta tesis se desarrollará un sistema para integrar distintos tipos de contextos a la tarea del reconocimiento de rostros.

3.4 Visión Térmica

La detección de humanos en entornos reales de hogar es una tarea difícil, principalmente debido a las condiciones de iluminación variable, fondos complejos, y todas las diferentes poses de un cuerpo humano con respecto a la cámara del robot. De hecho, el cuerpo humano es un objeto complejo y deformable, con muchos grados de libertad, cuya apariencia puede cambiar en gran medida cuando se asigna a una imagen 2D. Por lo tanto, el problema de la detección de un cuerpo humano o de una parte del cuerpo con una cámara CCD o CMOS estándar, que trabajan en el espectro visible, está lejos de ser resuelto. Dependiendo de las circunstancias específicas, los seres humanos pueden ser detectados mediante el uso de la información sobre sus rostros, siluetas, piel, o el movimiento, así como mediante el uso de información de profundidad. Ninguno de estos métodos es de uso múltiple y cualquiera de ellos puede fallar dependiendo de las circunstancias específicas. Por ejemplo, detección de rostro y silueta depende de la relación específica de los seres humanos presentan (por ejemplo, un rostro no se puede detectar cuando el ser humano se observa desde la parte posterior), detección de piel depende en gran medida de las condiciones de iluminación y en el fondo (por ejemplo, la piel puede fácilmente confundirse con otros materiales como la madera), la detección de movimiento depende en gran medida de las condiciones de iluminación y el movimiento relativo de los seres humanos (por ejemplo, un ser humano en una posición estática, no puede ser detectado), y detección de personas con información en profundidad requiere un mayor análisis para distinguir entre las partes del cuerpo y otros objetos.

Los sensores térmicos, sin embargo, permiten la detección robusta de cuerpos humanos independientemente de las condiciones de iluminación (la luz no es necesaria) y de la pose (la radiación térmica de un cuerpo humano se puede detectar en cualquier pose), y su rango de detección es de hasta varios metros, lo cual es suficiente para entornos domésticos. Además, los seres humanos también pueden ser identificados mediante el análisis de sus rostros en el espectro térmico [41][43]. Tomando todas estas propiedades en cuenta, parece natural que se incluyan cámaras térmicas en los robots de servicio actuales y futuros. El precio de las cámaras térmicas ya no es un factor para no usarlos en los robots domésticos, ya que el precio se ha reducido considerablemente en los últimos años, siendo comparable con el precio de los sensores láser de rango medio y las cámaras de tiempo de vuelo (TOF), ambos usados con frecuencia en robots domésticos. En la presente tesis, en el robot posee una cámara FLIR TAU

320 térmica [34]. Esta cámara tiene una resolución de 324x256 píxeles, y es sensible en el 7.5 a 13.5 μ m de longitud de onda infrarrojo.

Ha habido un gran número de artículos en la literatura reciente que se ocupan de la detección e identificación humana. En cuanto a la tecnología de sensores, varias obras se basan en el uso de la visión estéreo [105][64], la visión monocular [70][14], el sonar y visión [120], el láser y la visión [11], y visión térmica [71][72][75][12]. Por ejemplo, en [105] hay un sistema de visión estéreo que utiliza una imagen de profundidad para la detección y seguimiento de personas, [120] utiliza un sonar en combinación con un detector de color de piel para detectar los rostros, y [11] utiliza un láser para detectar las piernas de los seres humanos y un sistema de visión para encontrar los rostros. Uno de los principales beneficios de utilizar la visión térmica es la simplificación de la segmentación del cuerpo humano o partes del cuerpo humano desde el fondo. En particular existen bastantes trabajos en que realizan la fusión entre imágenes térmicas e imágenes normales para el reconocimiento de personas [106][40][107]. Estos trabajos son interesantes ya que la información que proporciona las imágenes es complementaria. En general en este trabajo se desarrolla esta idea de fusionar diferentes fuentes de información para lograr un mejor desempeño, pero las imágenes térmicas no solo se utilizan para reconocer a los rostros presentes sino que también para la detección de personas.

En el caso de la detección de cuerpos humanos, donde se ha realizado más investigación es en el problema de la detección de peatones (ver [29] para una encuesta). Algunos enfoques se basan en el uso de imágenes de infrarrojo lejano [71][13][72][75][12]. Estos enfoques utilizan plantillas probabilísticas [75], objetos calientes simétricos de tamaño específico y relaciones de aspecto [12], o el filtrado temporal [13] (por ejemplo, el filtro de Kalman). Un detector de cabezas funciona mejor que un detector de cuerpos al utilizar clasificadores estadísticos como se muestra en [71].

En el caso de detección visual de peatones, trabajos recientes se han centrado en la detección de peatones en caso de frenado de un automóvil, incluyendo: (i) una comparación de la utilización de diferentes características (características globales y locales (PCA coeficientes) [73], wavelets de Haar, y los campos receptivos locales), y clasificadores (*support vector machines*, redes neuronales *feed-forward* y los clasificadores *k-nearest neighbor*), (ii) el uso de un sistema basado en la generación de ROI estéreo, detección basada en la forma, clasificación basada en texturas y verificación basada en visión estéreo [37], (iii) el uso de algoritmos de detección en cascada para una clase general de modelos definidos por una gramática, en la que los modelos pueden representar a cada parte de forma recursiva como una mezcla de otras partes [33]. En cuanto a los tipos de características, otros trabajos relevantes incluyen el uso de nuevas características tales como histogramas de Gradiente Orientados (HOG) [24], las características HOG de tamaño variable[130], el uso regiones de matrices de covarianza [113][112], el uso conjunto de características de movimiento y la apariencia [116][25], y la comparación de tipos de diferentes características [86][122].

En cuanto a los clasificadores, en la literatura se incluye el uso de un clasificador Viola y Jones [84] como clasificadores en cascada [116][86][130][112], el uso de matrices de covarianza, junto con una variedad de Riemann [113][112], y la comparación de los métodos existentes [122]. Además, nuevas bases de datos se han propuesto recientemente (por ejemplo, base de datos Caltech peatonal [26][87] y la base de datos DaimlerChrysler peatonal [29][113] [36]) para este problema en particular. Sin embargo, cabe destacar que la detección de peatones es una aplicación completamente diferente a la detección y la identificación de

personas en un entorno doméstico, y que ambas aplicaciones tienen diferentes retos que cumplir.

Para la detección e identificación de personas, la información facial es una de las más utilizadas. Los trabajos existentes sobre la detección de rostros utilizando algoritmos de aprendizaje automático se ha aplicado casi exclusivamente a las imágenes en espectro visible, con poco trabajo dedicado a la utilización de imágenes térmicas [58]. Los mejores métodos de detección de rostros se basan en el uso de algoritmos de aprendizaje automático, como Máquinas de Vectores Soporte (SVM), *Convolutional Neural Networks* (CNN), y los *Boosting Classifiers* [115][44]. El paradigma de detección de rostros más popular se basa en el uso de las cascadas de *Boosting Classifiers*, permitiendo la detección robusta y eficiente de rostros [84]. Los detectores de rostros térmico y visible implementados en este trabajo se basan en este paradigma. No se ha encontrado en la literatura el uso de este tipo de detectores para imágenes térmicas, por lo que el uso en este trabajo sería su primera aplicación.

Capítulo 4
Metodología Propuesta

El sistema de visión que se quiere desarrollar tiene como objetivo a detección, segmentación, localización y reconocimiento de ciertos objetos de interés en imágenes (por ejemplo, rostros humanos). En particular en este trabajo se quiere diseñar e implementar un sistema de visión para robots humanoides que opere en ambientes no controlados, y sea capaz de detectar y reconocer rostros humanos en forma robusta, utilizando métodos de visión activa e integrando diferentes tipos de contexto.

En este capítulo se presenta la metodología con que se construye el sistema de visión propuesto y cuales son los módulos principales que lo componen.

En el Capítulo 5 se presenta un entorno virtual para entrenar y probar sistemas de detección de rostros en condiciones no controladas (pose, iluminación, expresión, etc.). Los elementos claves de esta herramienta son la base de datos de rostros reales y las imágenes de fondo que son capturadas bajo condiciones reales, y en ambientes no controlados. Se utilizará esa herramienta para validar el funcionamiento del sistema de visión propuesto y evaluar las diferentes hipótesis de esta tesis, además de realizar un estudio comparativo de los módulos de detección y reconocimiento de rostros en ambientes domésticos.

Luego en el Capítulo 6 se valida el sistema de visión propuesto en una aplicación real, en este caso se realiza una prueba enmarcada en la competencia de robótica RoboCup [96], que se llama "*Who is Who*". Se propone un sistema robusto para la detección y la identificación de los seres humanos en entornos domésticos para ser usado en un robot de servicio. La detección robusta de personas se logra mediante el uso de fuentes de información térmica y visual que se integran para detectar objetos que son *candidatos a humanos*. Estos candidatos son procesados con el fin de verificar la presencia de los seres humanos y su identidad con la información frente a los espectros térmico (*ET*) y visible (*EV*).

4.1 Arquitectura general

El diagrama de bloques de la arquitectura del sistema de visión robusto de rostros se muestra en la Figura 6. Las entradas del sistema corresponden a: (i) la imagen de las cámaras de video, ya sea la cámara visible o la cámara térmica; (ii) odometría del robot; (iii) conjunto de datos que pueden provenir de etapas de conocimiento a priori de cualquier especie, como por ejemplo una base de datos de rostros o información de contexto, y (iv) datos de otros sensores (diferentes a cámaras) como laser, kinect, etc. La salida del sistema es un vector de información para cada objeto de interés dentro de la imagen, que en este caso son rostros. Este vector contendrá toda la información adquirida por el sistema, como por ejemplo la ubicación del rostro (la posición relativa al agente determinada por el sistema), la identidad de la persona (en caso que sea reconocida) y otros datos en caso de que sean necesarios dependiendo de la aplicación (por ejemplo en el Capítulo 6, el vector de información la ubicación de los blobs de cuerpo, la posición de los rostros detectados, la identidad de los rostros, etc.). Las flechas indican la dirección del flujo de la información entre los diferentes módulos. En Figura 6 el sistema de visión posee 4 sensores: Odometría del robot, Laser, Cámara Visible y Cámara térmica. Esta información es entregada a los perceptores que en este caso son: un detector de rostros, un detector de ojos y un reconocedor de rostros, estos bloques extraen la información de las imágenes: las posición de los rostros presentes en las imágenes y el ID de los rostros, luego toda la información es utilizada por los diferentes módulos del sistema: el módulo de coherencia física verifica si la información extraída por los perceptores cumple con las leyes físicas que rigen el ambiente donde opera el robot; el módulo de coherencia espacial verifica que los rostros detectados sean coherentes en altura con las de una persona; el módulo de mapa de personas utiliza la información anterior y la información de la localización para estimar la posición de las personas dentro del mapa y las agrega al mapa si no hayan sido detectados anteriormente; el módulo de visión activa modifica la posición del robot si es necesario para modificar la observación y mejorar el rendimiento de alguno de los perceptores. Toda la información es recibida y verificada por el módulo de filtro de contexto que entrega un vector con la información adquirida. El sistema en general utiliza la información de las últimas detecciones para mantener actualizado los diferentes módulos en caso que sea necesario. Para mas detalles de cada módulo ver descripción general en sección 4.2, y para mas detalles de los perceptores ver sección □.

4.2 Módulos - Descripción general

En términos generales, este sistema de visión contará con los siguientes módulos:

- **Módulo de preprocesamiento de imagen**. En este módulo se realizan procesamientos de bajo nivel con el objetivo de mejorar la calidad de la imagen de entrada, como por ejemplo ecualización de histogramas.
- **Módulo de perceptores**. En estos módulos se encuentran distintos métodos para detectar o clasificar objetos de interés, por ejemplo detector de rostros en espectro visible, detector de ojos, detector de rostros en espectro térmico, etc. Estos perceptores se detallan en sección □.
- **Módulo de localización**. Este módulo entrega al sistema la información referente a la localización del agente en el mundo, su posición y su orientación.

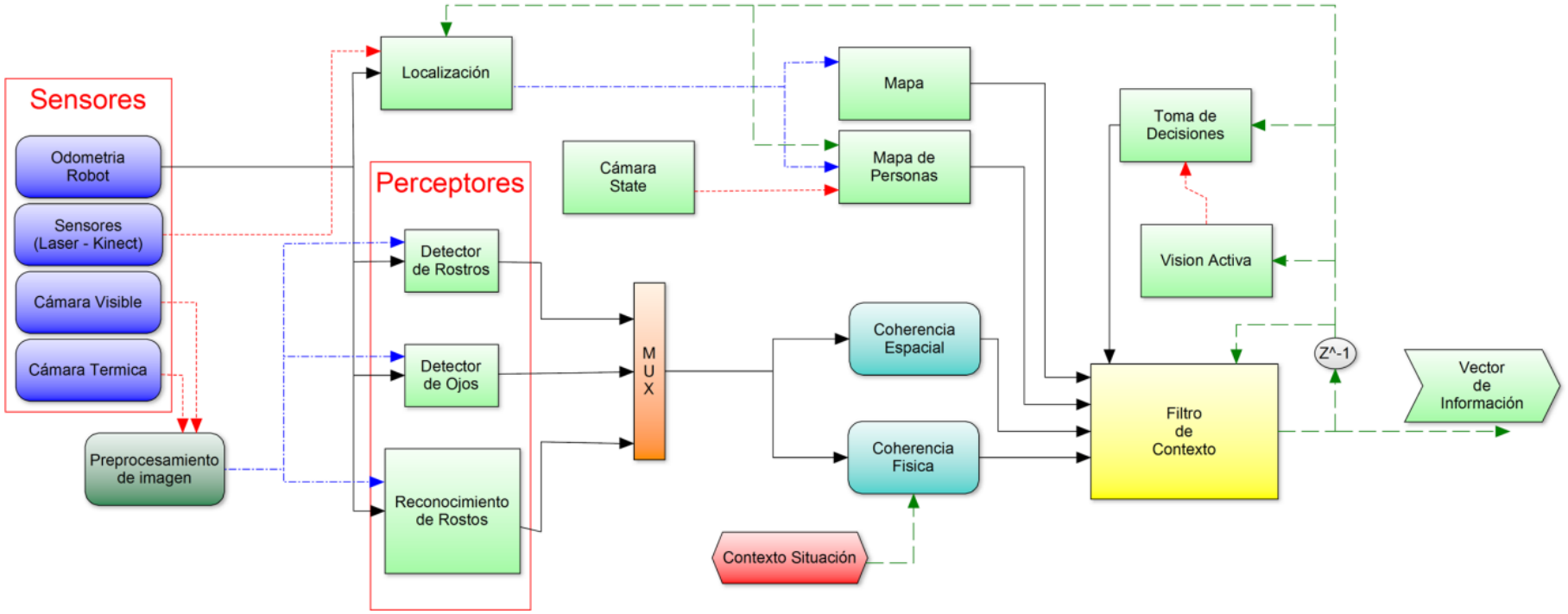


Figura 6: Diagrama general de la arquitectura del sistema de visión. Las flechas indican dirección del flujo de información.

- **Módulos de manejo de contexto**. En este módulo existen diversos sub-módulos que se encargan de evaluar diferentes instancias de contexto, utilizando la información adquirida por los diferentes perceptores y o la información ingresada por el usuario. Dentro de los sub-módulos de contexto están:
    1. **Contexto físico**: Se refiere a la información determinada a través de las leyes físicas que rigen el ambiente donde opera el robot. El modelo físico del ambiente define un modelo visual, por ejemplo la existencia de un piso sobre el cual deben estar los objetos que no tienen la capacidad de volar, la existencia de un vector de gravedad, entre otras.
    2. **Contexto de coherencia espacial**: Un objeto puede tener asociada una relación espacial específica con otros objetos. Por ejemplo un rostro está generalmente sobre un cuello y sobre los hombros. Este tipo de contexto es utilizado en el Capítulo 6 en el análisis térmico cuando se detectan personas.
    3. **Filtro de Contexto**: Este módulo se encarga de revisar y validar la información obtenida por el resto de los módulos, viendo que sean coherentes entre sí.

- **Contexto de la situación**: Este módulo es una entrada al sistema, la idea es saber qué actividad se está realizando, en conjunto con la escena o ambiente en que se está inmerso, determinan un contexto a nivel de la situación en que se encuentra el robot. Este nivel de contexto puede ser ingresada por el usuario previamente para aportar información extra.
- **Módulo de visión activa**. Este módulo se encarga de conectar el sistema de toma de decisiones con el sistema de visión (este módulo no esta incluido en el diagrama general). El objetivo de este módulo es mejorar la percepción del robot. Además se puede incluir una búsqueda local del objeto de interés para mejorar los tiempos de procesamiento cuando sea necesario. De esta manera se podrán detectar rostros en determinados instantes de tiempo, de forma más robusta y eficiente.
- **Mapa**: Este módulo está encargado de manejar la información del ambiente para permitir la navegación, parte de la información de este módulo es ingresada por el usuario previamente como el tamaño de mapa.
- **Módulo de Mapa de personas**: Este módulo está encargado de guardar la información de las diferentes detecciones de personas realizadas por el sistema en el tiempo para tener una estimación de la localización de estas. Además este módulo se encargará de verificar si la persona detectada ya se encuentra dentro del mapa o no.
- **Módulo de *Camera State*:** Este módulo está encargado de tomar las detecciones realizadas con las cámaras y proyectar esa información en el mundo real utilizando las características físicas del agente. De esta forma las observaciones se guardan en el mapa de personas.

4.3 Perceptores

En esta sección se describirán los principales perceptores utilizados en este sistema propuesto, además se detallará el proceso de selección del método de reconocimiento elegido.

4.3.1 Detector de rostros y ojos

Uno de los perceptores más importantes es el detector de rostros. La detección de rostros se basa en el uso de un método multi-escala de detección de objetos (ver el diagrama de bloques en la Figura 7) previamente desarrollado por grupo de visión computacional de la Universidad de Chile [115], que utiliza clasificadores de tipo *boosted* en cascada. El mismo trabajo se utiliza para construir detectores de rostros capaces de detectar los rostros en los espectros visible y térmico. Sin embargo, aunque ambos tipos de detectores comparten la misma estructura, el proceso necesario para la construcción de cada detector es diferente debido principalmente a la utilización de diferentes imágenes de entrenamiento en cada caso. Según lo investigado, los clasificadores estadísticos, y en particular los clasificadores de tipo *boosted* en cascada, no se han utilizado para la detección de rostros en imágenes térmicas hasta la fecha. Por lo anterior este es el primer detector de rostros que utiliza clasificadores de tipo *boosted* en cascada en imágenes térmicas.

La detección funciona de la siguiente manera: En primer lugar, para detectar rostros en diferentes escalas, se realiza un análisis multi-resolución de las imágenes usando una reducción de escala de la imagen de entrada por un factor fijo - por ejemplo, 1.2 - (módulo de análisis Multi-resolución). Posteriormente, se extraen las ventanas de 24x24 píxeles en el módulo de extracción de ventana para cada una de las imágenes de entrada escaladas. Entonces, las ventanas son analizadas por el clasificador (Módulo de Clasificación). Finalmente, en el módulo de Final de detecciones, las ventanas que se clasifican como positivos (es decir, que contiene un rostro) se fusionan (normalmente un rostro se detectan en diferentes escalas y posiciones) para obtener el tamaño y la posición de las detecciones finales.

Los conceptos clave utilizados en este detector son: cascadas anidadas, *boosting* y clasificadores de partición del dominio. Clasificadores en cascada constan de varias capas (etapas) de clasificadores con complejidad creciente para obtener una mayor velocidad de procesamiento, junto con una gran precisión. La idea principal de los clasificadores en cascada es que el procesamiento de la mayoría de las ventanas no-objeto es lo más rápido posible, y se procesan las ventanas de objetos y objetos similares, como más cuidado. *AdaBoost* es utilizado para encontrar y combinar varias hipótesis débiles, y para la selección de características. Las cascadas anidadas permiten una mayor precisión en la clasificación y velocidad de procesamiento mediante la reutilización en cada capa de la confianza (resultado de cada etapa) dada por su antecesor, y la cascada se compone de varias capas integradas (anidadas), cada una con un clasificador de tipo *boosted*. Una cascada anidada, compuesta de *M* capas, se define como la unión de los *M* clasificadores *boosted* H_C^k, y cada uno se define por:

$$H_c^k(x) = H_c^{k-1}(x) + \sum_{t=1}^{T_k} h_t^k(x) - b_k \qquad (4.1)$$

con $H_C^0(x) = 0$, h_t^k un clasificador débil, T_k es el número de clasificadores débiles en la etapa *k*, y b_k un umbral (bias) que define el punto de operación del clasificador fuerte. La salida de H_C^k es un valor real que corresponde a la confidencia del clasificador, y para su cálculo se utiliza el valor ya calculado de la capa previa de la cascada, y la clase es asignada usando el signo del valor. Se utiliza una partición del dominio de las hipótesis débiles, cada uno se asigna un valor de confidencia que estima la confiabilidad de cada predicción. La predicción de los clasificadores débiles depende solo de en qué bloque cayó un ejemplo dado para una característica dada:

$$h(f(x)) = c_j \quad \ni \quad f(x) \in F_j \qquad (4.2)$$

Para cada clasificador, el valor es asociado con cada bloque de la partición (c_j). Las salidas, c_j desde cada clasificador débil, obtenida durante el entrenamiento, son almacenadas en una LUT (*LookUp-Table o tabla de consulta*) para acelerar la evaluación.

Para el entrenamiento y la validación del detector de rostros las siguientes bases de datos fueron utilizadas:

- Espectro Visible. Entrenamiento: 5,000 imágenes de rostros frontales y 3,500 imágenes de no-rostros. Validación: 5,000 imágenes de rostros frontales y 1,500 imágenes de no-rostros. Las imágenes fueron obtenidas desde diferentes fuentes, y todas ellas fueron tomadas bajo condiciones reales, incluyendo variaciones de iluminación, fondo, raza, etc.

El proceso de entrenamiento de detector de rostros se describe en [115].

Otro perceptor importante es el detector de ojos que se utiliza principalmente para la alineación de los rostros detectados. El detector de ojo sigue las mismas ideas que tiene el detector de rostros, es decir, tiene sus mismos módulos de procesamiento. La única diferencia es que la búsqueda de los ojos se realiza en la parte superior del área del rostro, es decir, el módulo de extracción ventana extrae ventanas de las regiones donde ya han sido detectados rostros. Un detector de ojos izquierdo se utiliza para procesar la parte superior izquierda del rostro detectado, y un detector de ojos derecho se utiliza de la misma manera en la parte superior derecha del rostro. En realidad sólo un detector de los ojos tiene que ser entrenado (en este caso se entrenó el detector de ojo izquierdo), el otro es un espejo (*flop*), de la versión que fue entrenada. Debido a que no hay más de dos ojos por rostro, el módulo de análisis de superposición devuelve a lo sumo un ojo izquierdo y un ojo derecho. El detector de ojo izquierdo es un clasificador de tipo *boost* que consta de una cascada de 1-capa, sus clasificadores débiles se basan en las características rectangulares, que funciona sobre las ventanas de 24 x 24 píxeles, y por lo anterior solo se puede procesar los rostros de 50 x 50 píxeles o más si se quiere realizar un detección de ojos. Se utiliza sólo una capa debido a dos razones: (1) un conjunto negativo representativo para el entrenamiento puede obtenerse mediante el muestreo de ventanas no-centradas de los ojos, y (2) el tiempo de procesamiento no es importante porque sólo una pequeña región de la imagen, el rostro, necesita ser analizada y la escala del rostro es conocida. Más detalles del detector de ojos se describen en [115].

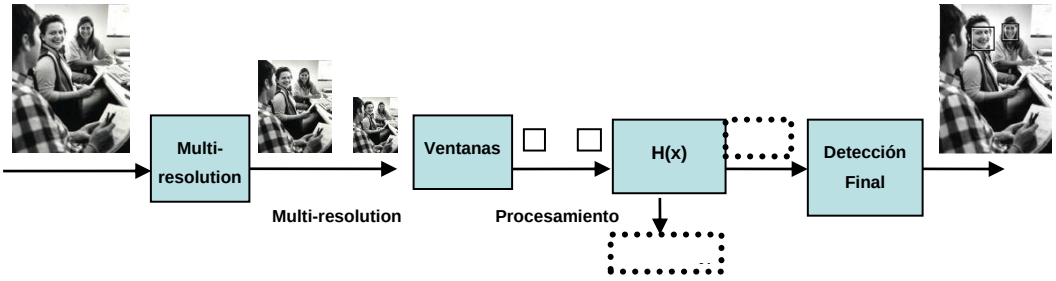


Figura 7. Diagrama de bloques de la detección de rostros.

4.3.2 Reconocedor de rostros

Aun cuando existen muchos métodos de reconocimiento de rostros, elegir cuál es el mejor es difícil, ya que no hay que sólo evaluar el método en las bases de datos estándar [77], sino que hay que evaluar los métodos bajo diferentes condiciones (iluminación, fondo, ruido, etc.). Para seleccionar el método que se utilizó para reconocer los rostros en esta tesis se realizó una evaluación exhaustiva de los mejores métodos, y ver cuál tiene un mejor rendimiento bajo diferentes condiciones. En [101] se presentó un estudio comparativo de métodos de reconocimiento de rostros considerando los requisitos planteados en la sección 2.1.1: (i) *Operación Online*; (ii) *Operación en Tiempo Real*; (iii) *Un rostro por persona*; y (iv) *Ambiente dinámico*.

Así, en este estudio se seleccionaron dos métodos de *local-matching*, un método holístico, y dos nuevos métodos de *image-matching* teniendo en cuenta el cumplimiento de los requisitos antes mencionados, y el desempeño en otros estudios comparativos [125] [50][51][48]. Los dos métodos de *local-matching* son histogramas de características LBP [9] y Gabor-Jet descriptor con *Borda count* [51], se seleccionaron teniendo en cuenta su rendimiento en [51][125]. Entre los métodos holísticos se incluye un PCA generalizada (Análisis de Componentes Principales) con la distancia euclidiana y características LBP para lograr invarianza de iluminación [48] (la restricción de una sola imagen por persona no permite incluir otros métodos holísticos). Además, dos métodos de reconocimiento facial basado en métodos avanzados de comparación de imágenes: SIFT [19] y ERCF (*Extremely Randomized Clustering Forest*) de descriptores SIFT que utiliza clasificadores lineales. Este último método, a pesar de no ser en tiempo real, se incluye a efectos de comparación, debido a los excelentes resultados que ha obtenido en la base de datos LFW [61].

El estudio comparativo se realizó con las base de datos FERET [51] y LFW [36]. Los métodos utilizados están descritos en la sección 3.2.

Tabla 1. Test FERET *fa-fb* y *fa-fc*. Tasas de Reconocimiento.

Test de ruido en la posición de los ojos y oclusión en parte del rostro en *fa-fb*. OR: Original. OC: Original con Oclusión. En negrita los mejores resultados bajo cada condición. Métodos que tienen una diferencia de menos del 1% son considerados con el mismo rendimiento.

Método	**100x185**						**203x251**					
	fa-fb					***fa-fc***	***fa-fb***					***fa-fc***
A-B-C	**OR**	**Ruido en la posición de los Ojos**			**OC**		**OR**	**Ruido en la posición de los Ojos**			**OC**	
		2.5%	**5%**	**10%**				**2.5%**	**5%**	**10%**		
H-HI-10	95.6	**95.0**	**91.3**	**81.8**	93.6	12.9	95.1	23.7	22.4	16.4	93.4	50.0
H-MSE-10	95.6	**95.0**	**91.3**	**81.8**	93.6	12.9	95.1	23.7	22.4	16.4	93.4	50.0
H-XS-10	95.7	94.7	**92.3**	**82.2**	78.4	14.9	95.1	41.3	39.4	31.0	86.1	60.8
H-HI-40	**96.5**	**96.0**	89.7	70.9	95.1	57.2	96.5	41.0	39.7	27.5	95.2	85.1
H-MSE-40	**96.5**	**96.0**	89.7	70.9	95.1	57.2	96.5	41.0	39.7	27.5	95.2	85.1
H-XS-40	95.5	93.6	87.0	67.4	92.1	47.4	97.4	76.6	71.4	53.8	95.0	88.1
H-HI-80	**97.2**	**95.6**	90.1	71.5	**96.7**	71.1	96.9	61.1	55.7	40.6	96.6	91.8
H-MSE-80	**97.2**	**95.6**	90.1	71.5	**96.7**	71.1	96.9	61.1	55.7	40.6	96.6	91.8
H-XS-80	96.3	94.1	88.3	68.0	94.4	62.9	97.4	87.8	83.9	64.9	**96.7**	92.8
PCA-MSE-200	73.1	55.9	40.7	16.2	63.6	52.1	---	---	---	---	---	----
PCA-MSE-500	76.1	60.3	42.9	16.0	64.9	57.2	---	---	---	---	---	---
GJD-BC	91.4	89.6	85.0	63.1	74.5	**79.9**	**98.5**	95.0	93.6	73.9	**97.7**	**99.0**
SD-FULL	74.3	75.7	73.5	71.5	67.3	7.7	97.1	**96.2**	**95.7**	**95.3**	95.6	67.5
SD-SIMPLE	73.1	75.3	73.1	71.0	68.6	5.7	**97.5**	**96.7**	**96.4**	**96.2**	95.3	63.9
SD-MATCHES	70.3	70.3	67.6	66.7	58.6	4.7	93.9	93.7	94.6	92.3	90.1	44.0

Se usa la siguiente notación para referirse a los métodos y sus variaciones: A-B-C. (i) A describe el nombre del algoritmo de reconocimiento facial: H - Histograma LBP, PCA - generalizado PCA con características LBP, GJD - descriptores Gabor Jets, SD - descriptores SIFT; (ii) B denota la medida de similitud: HI - Intersección de histograma, MSE - error cuadrático medio, XS - Chi-cuadrado, BC - Borda count, y EU - Distancia euclidiana, excepto en el caso de SD y ERCF, que no poseen una medida de distancia explícita, (iii) C describe los parámetros adicionales: número de divisiones en el caso del método basado en LBP, el número de componentes principales en el caso de PCA, el tamaño de la referencia-fijado para el caso GJD, y la versión el caso de SD (FULL, SIMPLE Y MATCHES son diferentes enfoques para encontrar los calces).

Algunos de los resultados de esta comparación se muestran en la Tabla 1. De los experimentos, se puede observar que: Los mejores resultados (~ 98,5%) se obtienen por GJD-BC, seguido por el SD y H-X-80, todas ellas con imágenes de 203x251. El desempeño de la GJD-X-X y los métodos de SD-X dependen en gran medida el tamaño normalizado de las imágenes recortadas. Probablemente debido a los métodos de uso de información sobre la forma y el contorno del rostro, que no aparece en las imágenes de 100x185 píxeles. En la evaluación de robustez frente a oclusiones GJD-BC y H-80-XS alcanzan las tasas de reconocimiento más altas en el caso de 203x251, el 97,7% y 96,7% respectivamente. En las pruebas de iluminación mejores versiones de H-X-X en el caso de 203x251 alcanzan un mejor rendimiento (92,8% vs 79%) que los reportados en la trabajo original [9], probablemente debido a las particiones de la imagen son diferente al que se utilizaron en esta implementación. Con respecto al uno computacional, uno de los requisitos impuesto es la operación en tiempo real. Por otro lado, la memoria requerida por los diferentes métodos es muy importante en algunas aplicaciones, ya que la memoria es un recurso limitado. La Tabla 2 muestra los costos de cómputo y memoria de los diferentes métodos en comparación. Si tenemos en cuenta que en muchas aplicaciones, el tamaño de la galería está en el rango de 10 a 100 personas, los métodos más rápidos son los H-X-X.

En base a los resultados obtenidos en este trabajo se eligió el método de reconocimiento que será utilizado en la implementación del sistema propuesto. Los resultados obtenidos son validados posteriormente en otros trabajos (Ejemplo [43]). Además se realizaran nuevas evaluaciones de los métodos mencionados en la sección 3.2 usando el ambiente virtual implementado en el Capítulo 5.

Tabla 2. Costos computacionales y de memoria.

FET: Tiempo de extracción de características o *Feature Extraction Time*. MT: Tiempo de *Matching*. PT: Tiempo de procesamiento. DM: Memoria usada por la base de datos. MM: Memoria usada por el modelo. TM: Memoria Total. Tiempo medido en milisegundos, memoria medida en Kbytes. Son consideradas DB de 1, 10, 100 and 1,000 rostros. El tamaño de la imagen es de 100x185 pixeles

Método	**FET**	**MT**	**PT** (FET+MT)				**DM**	**MM**	**TM** (DM+MM)			
			1	**10**	**100**	**1000**			**1**	**10**	**100**	**1000**
H-X-10	15	0.11	15	16	26	120	11	0	11	110	1100	11000
H-X-40	15	0.29	15	18	44	305	41	0	41	410	4100	41000
H-X-80	15	0.42	15	19	57	435	80	0	80	800	8000	80000
PCA-MSE-200	170	0.02	170	170	172	190	0,8	137800	137801	137808	137878	138585
PCA-MSE-500	360	0.02	360	360	362	380	2	137800	137802	137820	137996	139757
GJD-BC	50	0.25	50	53	75	300	33	1240	1273	1572	4559	34427
SD-X	4.7	1.03	6	15	108	1036	428	0	428	4284	42845	428451

Capítulo 5
Ambiente Virtual para Evaluación de Sistemas de Detección y Reconocimiento de Rostros

Un componente muy importante en el desarrollo de metodologías de reconocimiento de rostros es la disponibilidad de bases de datos adecuadas y de metodologías de evaluación. Por ejemplo, la muy conocida base de datos de FERET [77] [78] es una de las bases de datos más utilizada (incluye un protocolo de pruebas). Esta base de datos ha sido muy importante en el desarrollo de algoritmos de reconocimiento de rostros, en ambientes controlados, en los últimos años. Algunas bases de datos relativamente nuevas, como LFW (*Labeled Faces in the Wild* [61]) y FRGC (Face Recognition Grand Challenge [81][30]), entre otras, intentan proporcionar condiciones reales para la evaluación de los nuevos algoritmos de reconocimiento. En aplicaciones tales como HRI (*Human Robot Interaction*) y vigilancia, el uso de contexto espacio-temporal y o de mecanismos de visión activa en el proceso de reconocimiento de rostros puede aumentar en gran medida el rendimiento de los sistemas. Pero los enfoques de reconocimiento de rostros que incluyen estos mecanismos dinámicos no se puede validar correctamente con bases de datos actuales (se pueden ver ejemplos de las bases de datos en [31]). Incluso el uso de bases de datos de vídeo no permite probar el uso de esas ideas. Por ejemplo, en un vídeo grabado no es posible cambiar de forma activa el punto de vista del observador. El uso de un simulador que creara animaciones podría permitir lograr esto (cambios punto de vista), sin embargo, este simulador no sería capaz de generar rostros y fondos que parecieran los suficientemente reales.

Por otro lado, el uso combinado de una herramienta de simulación con rostros reales e imágenes de fondo capturadas en condiciones reales, podría permitir lograr el objetivo de proporcionar una herramienta para probar los sistemas de reconocimiento de rostros en condiciones no controladas. En este caso, más que proporcionar una base de datos y un procedimiento de evaluación, la idea sería crear un entorno de evaluación que proporcione una visualización realista de los rostros, mecanismos de visión activa y una metodología de evaluación. El objetivo principal de este capítulo es describir esta herramienta de evaluación. La herramienta proporciona un entorno simulado con personas ubicadas en diferentes posiciones y orientaciones dentro de un mapa virtual. Las imágenes de rostros son

previamente adquiridas con diferentes ángulos en *yaw* y *pitch*, y en condiciones variables de iluminación interior y exterior. Dentro de este entorno, un agente u observador, que tiene la capacidad de detectar y reconocer rostros, puede navegar y observar a las personas, el simulador de forma automática genera las imágenes de rostros (con información de fondo real) a diferentes distancias, ángulos (*yaw*, *pitch* y *roll*) y con el iluminación de interior o exterior. Durante el proceso de reconocimiento, el agente puede desplazarse por el mapa y por lo mismo cambiar su punto de vista para mejorar los resultados de reconocimiento de rostros. La herramienta de simulación proporciona todas las funcionalidades de la generación de imágenes, además de ayudar en la generación de las trayectorias de movimiento dentro del mapa.

Esta herramienta de prueba podría ser de gran interés para las aplicaciones de HRI relacionadas con el reconocimiento visual de los seres humanos, ya que permite comparar y cuantificar las capacidades de reconocimiento de rostros de robots de servicio bajo condiciones de trabajo exactamente iguales.

5.1 Descripción general

En este capítulo se presenta un entorno virtual para entrenar y probar sistemas de detección de rostros en condiciones no controladas (pose, iluminación, expresión, etc.). Los elementos claves de esta herramienta son la base de datos de rostros reales y las imágenes de fondo que son capturadas bajo condiciones reales, y en ambientes no controlados. Esto permite que cada vez que el agente se posicione para observar un rostro de una persona a una determinada distancia y punto de vista, se generen las imágenes u observaciones correspondientes con rostros reales e imágenes de fondo reales. En la Figura 8 se pueden ver ejemplos de las imágenes generadas.

Lo primero que se hará es definir algunos términos que se usarán en este capítulo: el primer término es USUARIO, que se refiere a la persona que utiliza el simulador para probar algún algoritmo de detección o reconocimiento de rostros; el segundo término es AGENTE que se refiere al ente que se encuentra en el mapa haciendo las observaciones (un ejemplo de AGENTE es un robot).

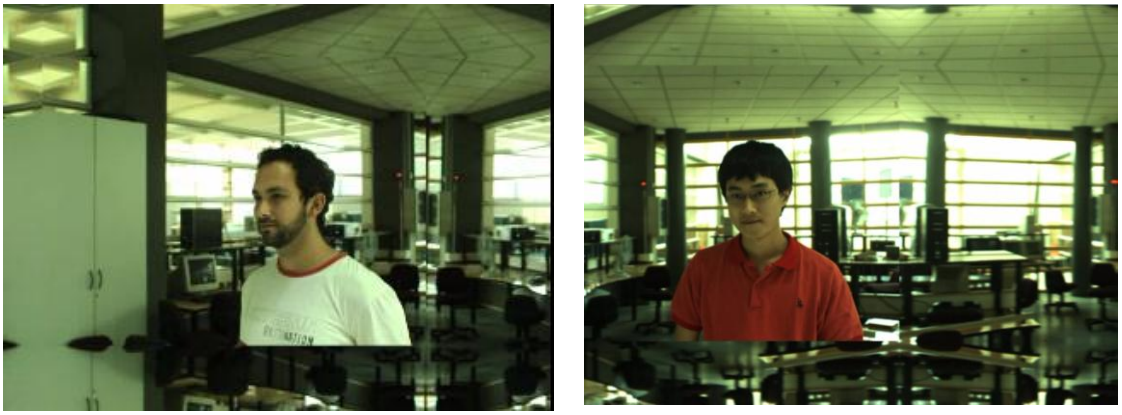


Figura 8**.** Ejemplos de imágenes generadas.

5.1.1 Características del Simulador

Dentro del entorno simulado hay un agente virtual que realiza observaciones, este agente debe ser capaz de detectar y reconocer rostros. En el ambiente virtual se puede navegar y observar las imágenes de rostros reales, a diferentes distancias, ángulos y con la iluminación interior o exterior. Durante el proceso de reconocimiento de rostros el agente puede desplazarse por el mapa, y por lo mismo cambiar su punto de vista y la distancia con respecto a las personas ubicadas en el mapa con el fin de modificar sus observaciones, usando métodos de visión activa, y con ello mejorar los resultados del reconocimiento de rostros. El entorno virtual ofrece todas las funciones que el agente puede tener en un ambiente real (navegación, posicionamiento, imagen del rostro de la composición en diferentes ángulos, etc.). Entre las funcionalidades de movimiento que posee el agente en el simulador se encuentran:

1. $Move(\Delta x, \Delta y)$: Esta función permite al agente desplazarse dentro del mapa virtual de forma relativa a la posición en que se encuentra. En este caso, sea (X_A, Y_A, θ_A), donde (X_A, Y_A) corresponde a la posición actual del agente en el mapa y (θ_A) a la orientación del agente en el mapa. Entonces la posición final del agente sería $(X_A + \Delta x, Y_A + \Delta y, \theta_A)$.
2. $Turn(\Delta\theta)$: Esta función permite al agente girar dentro del mapa virtual de forma relativa a la pose en que se encuentra. En este caso, sea (X_A, Y_A, θ_A), donde (X_A, Y_A) corresponde a la posición actual del agente en el mapa y (θ_A) a la orientación del agente en el mapa. Entonces la pose final del agente sería $(X_A, Y_A, \theta_A + \Delta\theta)$.
3. $SetAgentPosition(x, y, \theta)$: Esta función permite al agente desplazarse dentro del mapa virtual de forma absoluta a una posición deseada. En este caso, sea (X_A, Y_A, θ_A), donde (X_A, Y_A) corresponde a la posición actual del agente en el mapa y (θ_A) a la orientación del agente en el mapa. Entonces la posición final del agente seria (x, y, θ), independiente de la posición inicial del agente.

Además de las funcionalidades de movimiento que el sistema le ofrece al agente, también existen funciones que le permiten facilitar las tareas de navegación dentro del mapa. Entre las funcionalidades de navegación que posee el agente están las siguientes:

1. $GenerateTrayectory(TYPE)$: Esta función permite al agente navegar de forma fácil dentro del mapa ya que el sistema automáticamente genera una trayectoria que recorre a todas las personas dentro del mapa. Se genera una lista de posiciones dentro del mapa virtual de modo que el agente se mueve en forma absoluta entre cada punto de la trayectoria. La forma en que esta trayectoria se genera depende de la variable *TYPE*, más detalles de los diferentes tipos de trayectorias se puede encontrar en sección 5.4.3.
2. $Nextposition()$: Esta función permite al agente moverse de una posición de la trayectoria generada a la siguiente. En cada posición de la trayectoria en que se ubica al agente, se puede utilizar las funciones de movimiento definidas para modificar las observaciones realizadas, y luego utilizar esta función para regresar a la trayectoria.

Otras funcionalidades que posee el simulador están relacionadas a los módulos implementados (ver detalles en sección 5.4), entre las funciones disponibles están:

1. $GenerateImage()$: Esta función genera una imagen tomando en cuenta la posición actual del agente y la posición de las personas en el mapa. Esta relacionada con el módulo Generador de Imágenes, para más detalles ver sección 5.4.2.
2. $StoredInfoFaces(FacesInfo)$: Esta función está relacionada con el módulo de Mapa de Personas (ver sección 5.4.5). El mapa de personas guarda la información de las diferentes personas detectadas dentro del mapa, la idea de este módulo es mantener actualizada la posición de las personas y de esta forma saber cuando a una persona detectada ya ha sido incluida en el mapa. La función almacena a las personas que han sido detectadas en la imagen actual.
3. $SaveMap()$: Esta función permite al usuario guardar la información actual del mapa en una imagen, de esta forma se tiene la posición de agente y de las personas en el mapa en ese instante.
4. $GenerateOcclusion()$: Esta función genera un pilar frente a las personas que ocluye al agente, obligando a utilizar visión activa para modificar las observaciones. Este pilar es localizado en el mapa frente a la persona de forma aleatoria con un radio y un ángulo definido. Más detalle de este módulo en sección 5.4.4.

5.1.2 Simulación

Para realizar una prueba con este simulador lo primero que se debe hacer es crear una instancia de la clase *Simulator* de la siguiente forma:

*Simulator * SIM = new Simulator ("params/ConfigMap.conf");*

El parámetro que se entrega entre los paréntesis es el archivo donde se encuentran todas las configuraciones de parámetros que necesita el simulador para funcionar, entre las cuales se encuentran:

- *MapSizeHeight*: Configura el alto que tendrá el mapa.
- *MapSizeWidth:* Configura el ancho que tendrá el mapa.
- *NumberOfPeople*: Configura el numero de personas que tendrá el mapa.
- *PersonBaseHeight:* Configura la altura base de las personas.
- *DeltaPersonHeight:* Configura el máximo valor que puede tener el ruido que se le agrega a la altura de la persona (mas detalle en sección 5.4.2)
- *Ambient*: Configura si el ambiente de las imágenes es *indoor* o *outdoor*.
- *AgentHeight:* Configura la altura del agente.
- *ActiveVision:* Configura si el agente puede o no usar visión activa para modificar sus observaciones.
- *Occlusion:* Configura si existe o no oclusiones entre el agente y el usuario.
- *TrajectoryType:* Configura el tipo de trayectoria con la cual el agente recorrerá el mapa.

En caso de no entregarle ningún archivo el buscar el archivo de configuración por defecto llamado *Map.conf*. Un ejemplo de este archivo se puede ver en la Figura 9.

```
# Map Related
MapSizeHeight = 4000
MapSizeWidth = 4000
NumberOfPeople = 2
# Person Related
PersonBaseHeight = 165
DeltaPersonHeight = 20
# Simulation Image Related
ActiveVision = 0
Occlusion = 0
ObstacleDistMin = 100
ObstacleDistMax = 100
ObstacleAngleRange = 60
Ambient = 0 # 0: Indoor 1: Outdoor -1: Random
BaseDistance = 200
IMAGE_WIDTH = 1280
IMAGE_HEIGHT = 960
# Agent Related
AgentHeight = 160
# Trajectories Related
TrajectoryType = 5
TrajectoryStep = 2
TrajectoryA = 200
TrajectoryB = 100
TrajectoryC = 200
```

Figura 9**.** Ejemplo de archivo de configuración.

Una vez creada la instancia de la clase Simulador, se deben crear instancias del detector de rostros y del reconocedor de rostros que no se encuentran incluidos en el simulador, y deben ser provistas por el usuario. Por ejemplo:

*FaceDetector * FD = new FaceDetector ();*

*FaceRecognition * FR = new FaceRecognition();*

```
Inicialización:
    SIM = new Simulador();
    ReadParamsFromFile;
    num_recognized_faces=num_false_positives=0;
    Generatetrayectory();
Prueba:
    SetRobotInitialPose();
for (i=0;i<N;i++)
        currentImage =GetImage();
        id=RecognizePerson(currentImage);
        if (id==GetPersonID(i))
                num_recognized_faces+=1;
        else if (id!= NO_IDENTIFICATION)
                num_false_positives+=1;
        end;
        Trayectory.Next();
end;
Reconocimiento:
   RecognizePerson(image)
   while(1)
        Nface=DetectFace(image)
        if((Nface>0){
                if(face.size<MIN_SIZE)
                        result=RecognizeFace(face)
                if(result.confidence<threshold)
                        return(NO_IDENTIFICATION)
                else
                        return(result.id)
                end;
        end;
   end;
```

Figura 10. Pseudo código de una simulación.

Luego de tener definidos ambas instancias se debe generar la trayectoria que recorrerá el agente en el mapa. De la siguiente manera:

SIM->GenerateTrajectory();

Ahora se puede guardar la información de la trayectoria y la ubicación de las personas en el mapa en una imagen de la siguiente manera:

SIM->DrawTrajectory();

SIM->DrawPersons();

SIM->SaveMap();

Una vez que el simulador es creado y configurado se puede generar una imagen en cualquier momento usando el comando:

*IplImage * Imagen = SIM->GeneratedImage();*

En la Figura 10 se puede observar un ejemplo de un pseudo-codigo. En este ejemplo que no utiliza visión activa para modificar las observaciones.

5.2 Construcción de base de datos y Sistema de adquisición

Las imágenes de rostros se adquieren en diferentes ángulos en *yaw* y *pitch* mediante una cámara CCD montada en una estructura giratoria (ver el diagrama de la Figura 12a). La persona que está siendo capturada se encuentra en una posición fija, mientras la cámara, situada a la misma altura que el rostro de la persona y a una distancia fija de 140 cm, gira en el plano axial (la altura de la cámara es ajustable). Un *encoder* colocado en el eje de rotación se utiliza par calcular el ángulo de orientación del rostro. No hay restricciones a la expresión facial de la persona. El sistema es capaz de adquirir imágenes con una resolución de 1°. Sin embargo, en esta primera versión del sistema, las imágenes se toman cada 2°. El proceso de captura tarda unos 120 segundos para cada par (*yaw*,*pitch*), y se utiliza una cámara CCD (DFK 41BU02 modelo) de 1280 x 960 píxeles. En la imagen frontal, el tamaño del rostro es de 200x250 píxeles en promedio.

Las variaciones se obtienen repitiendo el proceso de captura descrito en los ángulos de *pitch* diferentes. En cada caso, la altura de la cámara se mantiene, pero la persona mira a un punto de referencia diferentes en el eje vertical, que se encuentran a una distancia de 160 cm de la persona (ver Figura 12a). Los ángulos de inclinación de -15 °, 0 °, 15 ° en *pitch* son las variaciones típicas rostro humano. Además, para cada ubicación, se capturan imágenes de fondo con el fin de ser capaz de componer las imágenes reales. En la Figura 11 se muestran algunos ejemplos de imágenes tomadas con el dispositivo.

Es importante destacar que el dispositivo de adquisición es portable (no requiere ninguna instalación especial). Por lo tanto, todo el proceso de adquisición se puede realizar en diferentes lugares (ambiente de la calle, el medio ambiente de laboratorio, ambiente de centro comercial, etc.). En este caso se utilizó al menos en dos lugares diferentes para cada persona, una interior (de laboratorio con las ventanas), y una al aire libre (jardines en el interior del campus de la escuela).

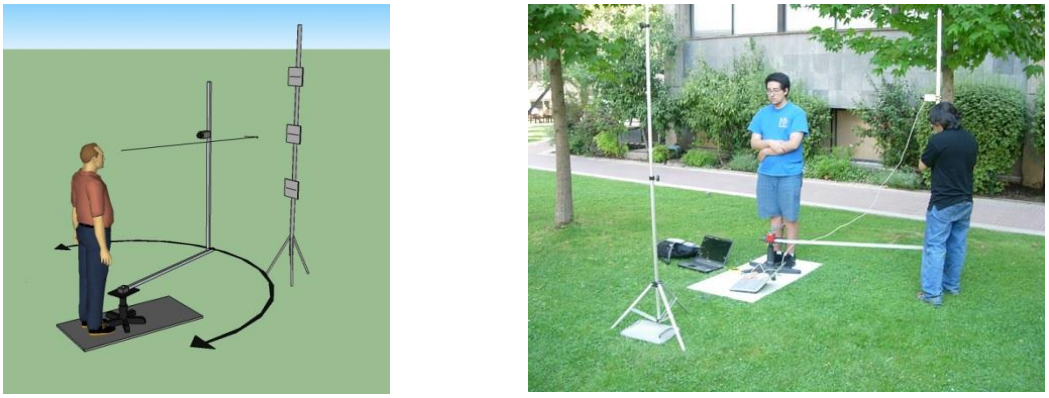
Con las imágenes obtenidas se construyó una base de datos que contiene 80 personas. Para cada persona, se almacenan 726 imágenes (121x3x2). El rango de ángulo de *yaw* es de -120° a 120°, con una resolución de 2 °, lo que da 121 imágenes. Para cada ángulo en *yaw*, se consideran tres ángulos diferentes de *pitch* (-15°, 0°, y 15°). Para cada combinación se toman

las imágenes en interior y exterior. Además, se almacenan en la base de datos las imágenes de fondo correspondiente a los distintos ángulos de orientación de *pitch*, el lugar y la altura de la cámara para cada persona.

Yaw: 50° Pitch: -15 Yaw: 0° Pitch: 15

Yaw: 90° Pitch: -15 Yaw: 30° Pitch: 0

Figura 11**.** Ejemplo de imágenes tomadas usando el sistema en interior/exterior en primera/segunda columna respectivamente.



(a) (b)

Figura 12**.** (a) Diagrama de sistema de adquisición de imágenes. (b) El sistema instalado en exterior.

5.3 Diagrama de bloques del sistema de visión

En la Figura 14 se puede ver la conexión del simulador y del sistema de visión, la idea es que el simulador entrega la imagen que genera al sistema de visión y el sistema de visión le entrega órdenes de actuación para modificar la posición del agente dentro del mapa. El diagrama de bloques de la Figura 13 describe en general el sistema de visión implementado, y esta basado en el diagrama general que se plantea en el Capítulo 4 (ver Figura 6). Las entradas

del sistema corresponden a: (i) la imagen de la cámara visible que en realidad es la observación generada por el simulador; (ii) odometría (en este caso, la odometría del agente virtual); y (iii) el conjunto de datos que provienen de etapas de conocimiento a priori, en este caso una base de datos de rostros o contexto de situación. La salida del sistema es un vector de información para cada objeto de interés dentro de la imagen, que en este caso son rostros. Este vector contendrá la ubicación del rostro (la posición relativa al agente determinada por el sistema), la identidad de la persona (en caso que sea reconocida) la posición del rostro dentro de la imagen, y la pose de la persona. Las fechas indican la dirección del flujo de la información entre los diferentes módulos. En este sistema hay 2 sensores: Odometría del robot (generada por el simulador) y una Cámara Visible (que en realidad es el generador de imágenes). Esta información es entregada a los perceptores para que extraigan información desde estas fuentes, en este caso existen 3 perceptores: el detector de rostros, el detector de ojos y el reconocimiento de rostros. Después toda la información de los perceptores es utilizada en los diferentes módulos: el módulo de coherencia física verifica si la información extraída por los perceptores cumple con las leyes físicas que rigen el ambiente donde opera el robot; el módulo de coherencia espacial verifica que los rostros detectados sean coherentes en altura con las de una persona; el módulo de mapa de personas utiliza la información anterior y la información de la localización para estimar la posición de las personas dentro del mapa y las agrega al mapa si no hayan sido detectados anteriormente; el módulo de visión activa modifica la posición del robot si es necesario para modificar la observación y mejorar el rendimiento de alguno de los perceptores. Toda la información es recibida y verificada por el módulo de filtro de contexto que entrega un vector con la información adquirida (más información de cada módulo en 5.4). El sistema utiliza la información de las últimas detecciones para mantener actualizado los diferentes módulos en caso que sea necesario.

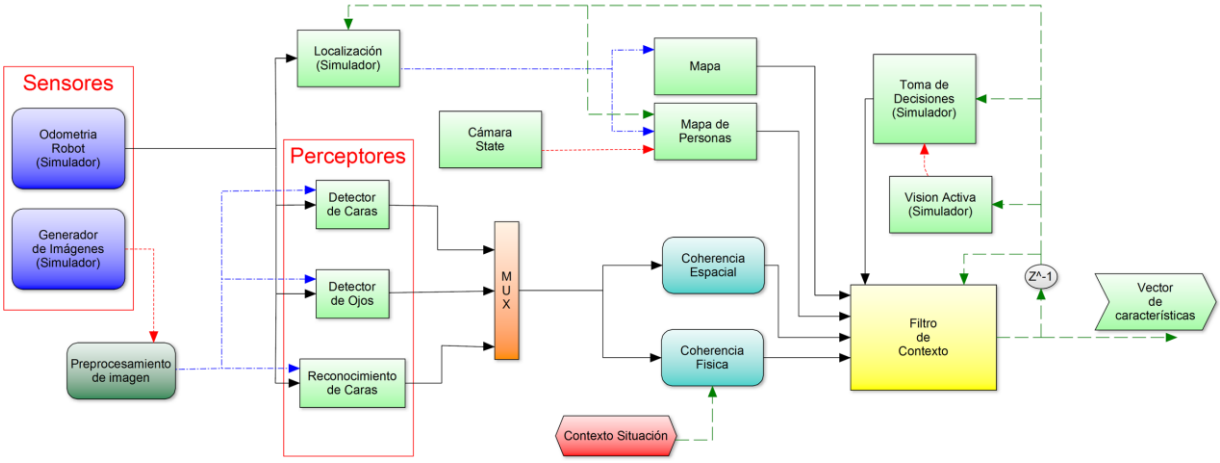


Figura 13: Diagrama general del sistema desarrollado. Las flechas indican dirección del flujo de información.

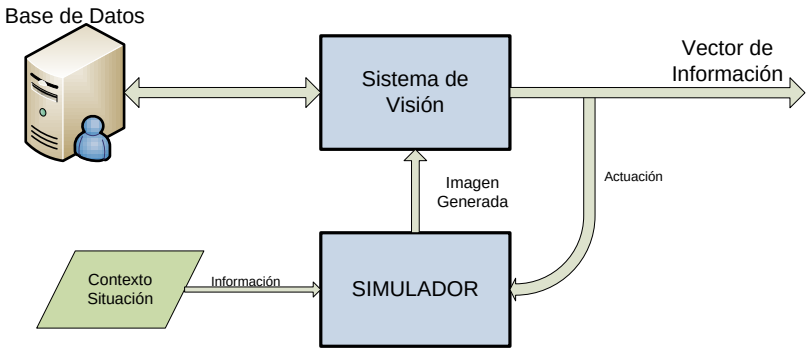


Figura 14: Diagrama de conexión del simulador y el sistema de visión.

5.3.1 Detector de rostros y ojos

Uno de los perceptores más importantes del simulador es el detector de rostros. La detección de rostros se basa en el uso de un método multi-escala de detección de objetos previamente desarrollado por grupo de visión computacional de la Universidad de Chile [115], que utiliza clasificadores de tipo *boosted* en cascada. Más detalles del detector de rostros en sección 4.3.1.

Otro perceptor importante es el detector de ojos que se utiliza principalmente para la alineación de los rostros detectados. El detector de ojos sigue las mismas ideas que tiene el detector de rostros, es decir, tiene sus mismos módulos de procesamiento. La única diferencia es que la búsqueda de los ojos se realiza en la parte superior del área del rostro, es decir, el módulo de extracción ventana extrae ventanas de las regiones donde ya han sido detectadas rostros. Más detalles del detector de ojos en sección 4.3.1.

5.3.2 Reconocedor de rostros

Otro perceptor importante es el de reconocimiento de rostros que utiliza la información de los dos perceptores anteriores (detector de rostros y detector de ojos). Durante los últimos años el problema del reconocimiento de rostros se ha convertido en uno de los temas de investigación más activos entre las aplicaciones de reconocimiento de patrones. El interés en el tema ha sido potenciado por diversas aplicaciones: Vigilancia, interacción humano-robot, seguridad, etc. Aun cuando los sistemas de reconocimiento de rostros han mostrado una evolución importante el problema del reconocimiento de rostros está lejos de ser un problema resuelto, especialmente en ambientes no controlados.

5.4 Módulos del Simulador

5.4.1 Mapa Global

El mapa del simulador contiene una lista con la información de las diferentes personas localizadas dentro del mapa, más la posición del robot, esta información es generada cuando se crea el simulador y se utiliza el archivo de configuración respectivo. Dado un mapa de tamaño $(Size_X, Size_Y)$ que contiene a N personas. Para cada persona se genera (X_i, Y_i, θ_i) que contiene la posición x e y en el mapa y la orientación θ. Además se genera el ángulo de Roll (θ_i^{Roll}), Pitch (θ_i^{Pitch}), altura (h_i) y un identificador único de la persona (ID_i).

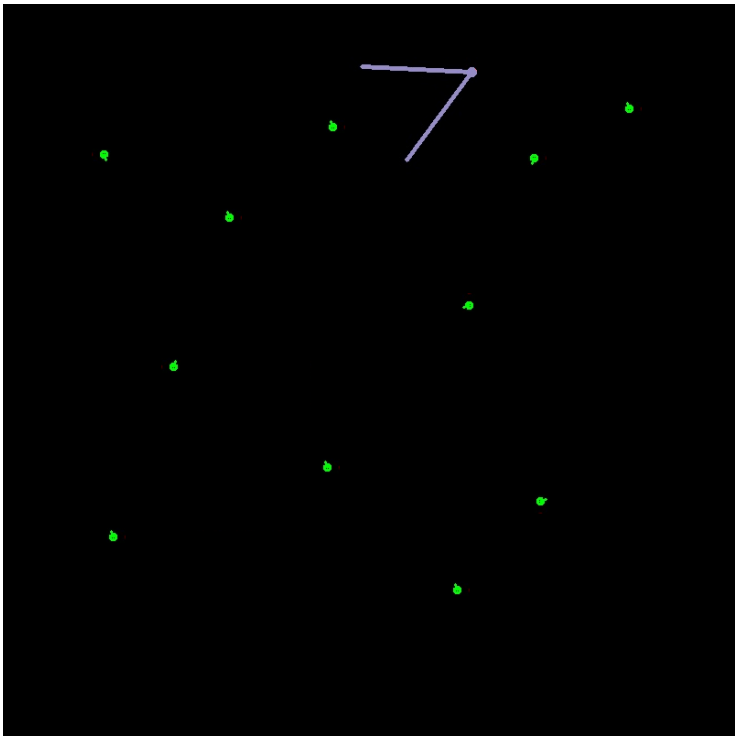


Figura 15**.** Ejemplo de mapa con 11 personas (puntos verdes), más el robot (lila). Las líneas definen el campo de visión del robot.

5.4.2 Generador de imágenes

Para generar las imágenes se toma la posición del robot y se busca a la persona más cercana dentro del campo visual del agente o robot en un radio de 600 [cm]. Para esto se calcula la distancia entre el robot y todas las personas del mapa. Entre las personas que cumplen las restricciones se selecciona a la que se encuentre más cerca del robot (ver pseudo código en Figura 17). Esto se calcula de la siguiente manera:

Sea (X_i, Y_i, θ_i) con $i \in \{1, \dots, N\}$ donde N es el número de personas en el mapa, y sea (X_A, Y_A, θ_A) , donde (X_A, Y_A) corresponde a la posición del agente en el mapa y θ_A a la orientación del agente en el mapa. Sea $D_x = 180 + \tan^{-1}\left(\frac{X_i - X_A}{-Y_i + Y_A}\right) - \theta_A$, el ángulo en que se encuentra desplazada la persona con respecto al eje X del sistema de referencia del agente, es decir el ángulo con que el agente ve a la persona; y $D_i = \sqrt{(X_i - X_A)^2 + (Y_i - Y_A)^2}$ que representa la distancia entre el agente y la persona *i*.

$$\begin{aligned} &\text{Luego, } \min_i(D_i) \text{ con } i \in \{1, \dots, N\} \\ &s.a. \ (1)\ M_i < 600\ [cm] \\ &(2)\ D_x < 60\ [°] \\ &(3)\ D_x > -60\ [°] \end{aligned} \qquad (5.1)$$

Una vez que se encuentra el mínimo se guarda el $indice$ de la persona que cumple las ecuaciones anteriores en la variable i_{Sel}. (i_{min} es igual al índice de la persona que registra el mínimo de la ecuación 5.1).

$$i_{Sel} = i_{min} \quad (5.2)$$

Si no existe ninguna persona que cumpla los requerimientos se genera una imagen que no posee ningún rostro usando las imágenes de fondo de las personas. Si se encuentra una persona, primero se calcula el ángulo $Image_\theta$ con que el agente debería observar a la persona, o sea el ángulo de la persona con respecto al agente. Éste ángulo se calcula de la siguiente manera:

$$Image_\theta = \tan^{-1}\left(\frac{X_{i_{Sel}} - X_A}{-Y_{i_{Sel}} + Y_A}\right) + 90 - \theta_{i_{Sel}} \quad (5.3)$$

Luego se modifica el ángulo $Image_\theta$ para que su valor se encuentre entre $[-\pi, \pi[$.

Ahora se verifica que $Image_\theta$ cumpla con las siguientes restricciones:

$$(1)\ Image_\theta < 120 \quad (5.4)$$

$$(2)\ Image_\theta > -120 \quad (5.5)$$

Estas restricciones están dadas por la base de datos ya que solo existen imágenes en el ángulo de *yaw* de -120° a 120° (con una resolución de 2°)

Después se usa la distancia D_S a la que se encuentra la persona del agente y se calcula el factor de escala SF_S:

$$D_S = \sqrt{\left(X_{i_{Sel}} - X_A\right)^2 + \left(Y_{i_{Sel}} - Y_A\right)^2} \quad (5.6)$$

$$SF_S = \frac{160}{D_S} \quad (5.7)$$

Los 160 cm son debido a que es la distancia a la que fueron tomadas las imágenes en la base de datos. Luego se estima el ángulo de pitch de la imagen utilizando la altura del agente H_A, la altura de la persona H_i y la distancia D_i.

$$\theta_S^{Pitch} = \begin{cases} 0 & Si\ abs\left(H_{i_{Sel}} - H_A\right) < \dfrac{\tan(15) * D_S}{2} \\ 15 & Si\ \left(H_{i_{Sel}} - H_A\right) > \dfrac{\tan(15) * D_S}{2} \\ -15 & \sim \end{cases} \quad (5.8)$$

El valor 15 esta dado las imágenes en la base de datos, ya que los valores en el ángulo *pitch* pueden ser: $\{-15,0,15\}$.

Con la información obtenida en los pasos anteriores se lee la imagen correspondiente de la persona i_{Sel}, que corresponde a la $Image_\theta$ y θ_S^{Pitch}. A esta imagen se le agrega fondo aumentando su tamaño, para luego re-escalar la imagen utilizando el factor de escala calculado SF_S. En la Figura 18 se puede observar una imagen con fondo agregado, esto se realiza para evitar que se generen fondos sin información cuando se re-escale y desplace la imagen. Luego

de agregar el fondo se calculan los desplazamientos que tendrá la persona en la imagen dadas las ubicaciones en el mapa. Se calcula un $Shift_x$ y un $Shift_y$ de la siguiente manera:

$$Shift_x^\theta = 180 + \tan^{-1}\left(\frac{-Y_{i_{Sel}} + Y_A}{X_{i_{Sel}} - X_A}\right) - \theta_A \quad (5.9)$$

$$Shift_y^\theta = \tan^{-1}\left(\frac{H_{i_{Sel}} - H_A}{D_{i_{Sel}}}\right) \quad (5.10)$$

$$Shift_x = \frac{Shift_x^\theta * ImageWidth}{FOV_H} \quad (5.11)$$

$$Shift_y = \frac{Shift_y^\theta * ImageHeight}{FOV_V} \quad (5.12)$$

Donde $Shift_x^\theta$ y $Shift_y^\theta$ son los desplazamientos angulares con respecto al agente; el $FOV_H = 56.3$ y el $FOV_V = 42.3$ son los FOV (*field of view*) de la cámara con que se capturó la base de datos.

Una vez que los desplazamientos $Shift_x$ y $Shift_y$ son calculados se utilizan para desplazar el centro de la imagen en ambos ejes desde el rostro de sujeto *i* a un nuevo centro definido por estos parámetros. Una vez definido el nuevo centro de la imagen se recorta la imagen a un tamaño definido por el usuario. Un ejemplo de la imagen generada con las posiciones de la persona y del agente fijas (se pueden ver mapa en Figura 16) se puede apreciar en la Figura 19. En la Figura 18 se puede ver un ejemplo de la imagen antes de cortar y re-escalar.

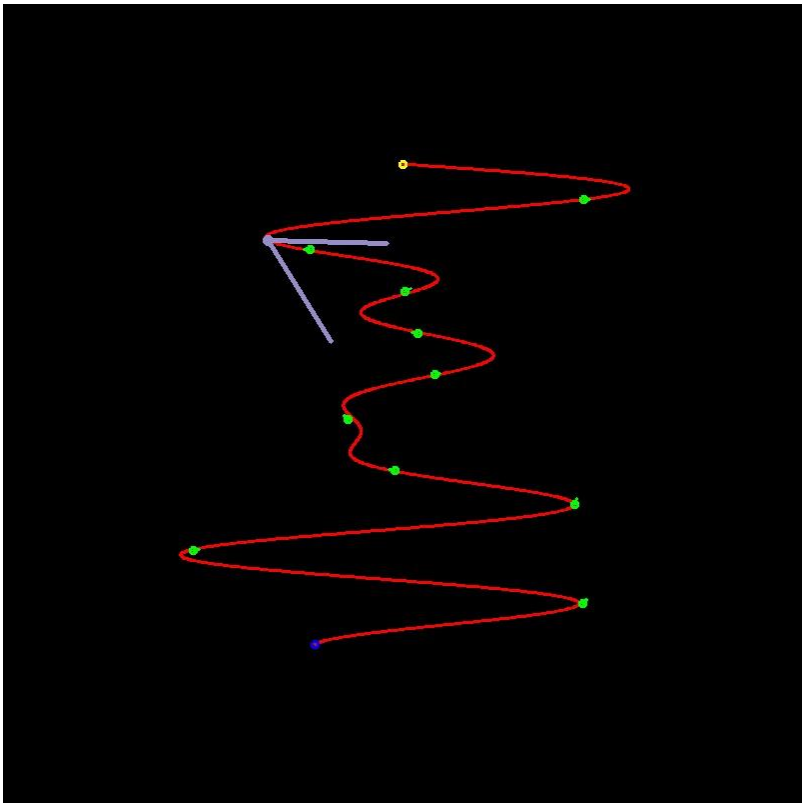


Figura 16. Ejemplo mapa para la imagen generada. Se puede apreciar al agente y a la persona dentro del mapa.

```
GenerateImage() {
        if(ActiveVision)
                ActualPerson = FindClosestPerson();
        if (ActualPerson==-1)
                return GenerateRandomBackground();

        ImageAngle =  CalculateImageAngle(); // Ecuación 5.3
        Distance = CalculateDistance();// Ecuación 5.6
        ScaleFactor = 140/Distance; // Ecuación 5.7

        if (ImageAngle > 120 || ImageAngle< -120)  // Ecuaciones 5.4 y 5.5
                return GenerateRandomBackground();

        SetPersonPitch(ActualPerson,Distance); // Ecuación 5.8
        ShiftX =CalculateShiftX();// Ecuación 5.11
        ShiftY =CalculateShiftY();// Ecuación 5.12

        SourceImage = LoadImage(); //Se lee la imagen original
        AddBackGround(); // Se le agrega fondo
        RotateImage();  //Se rota la image
        ResizeImage();//Se reescala la image
        ShiftImage();//Se desplaza la image

        return FinalImage;
```

Figura 17. Pseudo código de la función que genera las imágenes.

Figura 18**.** Ejemplo imagen leída y con fondo agregado.

Figura 19**.** Ejemplo imagen generada.

5.4.3 Generador de trayectorias

Existen diferentes tipos de navegación dentro del mapa:

- **Navegación libre**: el usuario puede modificar la posición del agente con toda libertad, puede mover al agente sin restricciones y generar las observaciones correspondientes cuando el estime pertinente.

- **Navegación con trayectoria definida**: En este caso el simulador genera una trayectoria por la cuál se moverá el agente, el usuario solo puede mover al agente dentro de esta trayectoria. Existen diferentes tipos de trayectorias las cuales serán detalladas mas adelante.

- **Navegación con trayectoria definida y visión activa**: Esta configuración es una mezcla de las anteriores, o sea existe una trayectoria definida por la cual el agente se podrá mover, pero además el usuario puede modificar la posición del agente de forma libre si así lo requiere para por ejemplo modificar una observación que realizó y mejorar el reconocimiento de rostros.

La idea de las trayectorias definidas es ayudar al usuario a posicionar al agente cerca de la ubicación aproximada de las personas dentro del mapa, y que pueda generar las observaciones correspondientes.

Existen las siguientes trayectorias predefinidas:

- **Frontal**: El simulador genera una trayectoria para cada persona en el mapa. Esta trayectoria se va acercando a la persona. Tiene 3 parámetros: La distancia inicial, la distancia final y el paso (cuanta distancia avanza entre cada iteración). El agente se mueve solo en la trayectoria.

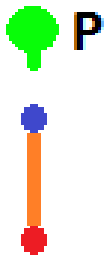


Figura 20**.** Ejemplo movimiento del agente generado con trayectoria frontal.

Figura 21**.** Ejemplo imagen generada con trayectoria *frontal*.

- **Side To Side**: El simulador genera una trayectoria para cada persona en el mapa. La trayectoria es una línea que se mueve en forma perpendicular a una línea imaginaria que sale de la persona. Este movimiento es a una distancia fija y esta distancia es un parámetro. Además se debe especificar la distancia recorrida y el paso (cuanta distancia avanza entre cada iteración). El agente se mueve solo en la trayectoria.

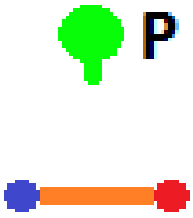


Figura 22**.** Ejemplo movimiento del agente generado con trayectoria Side to Side.

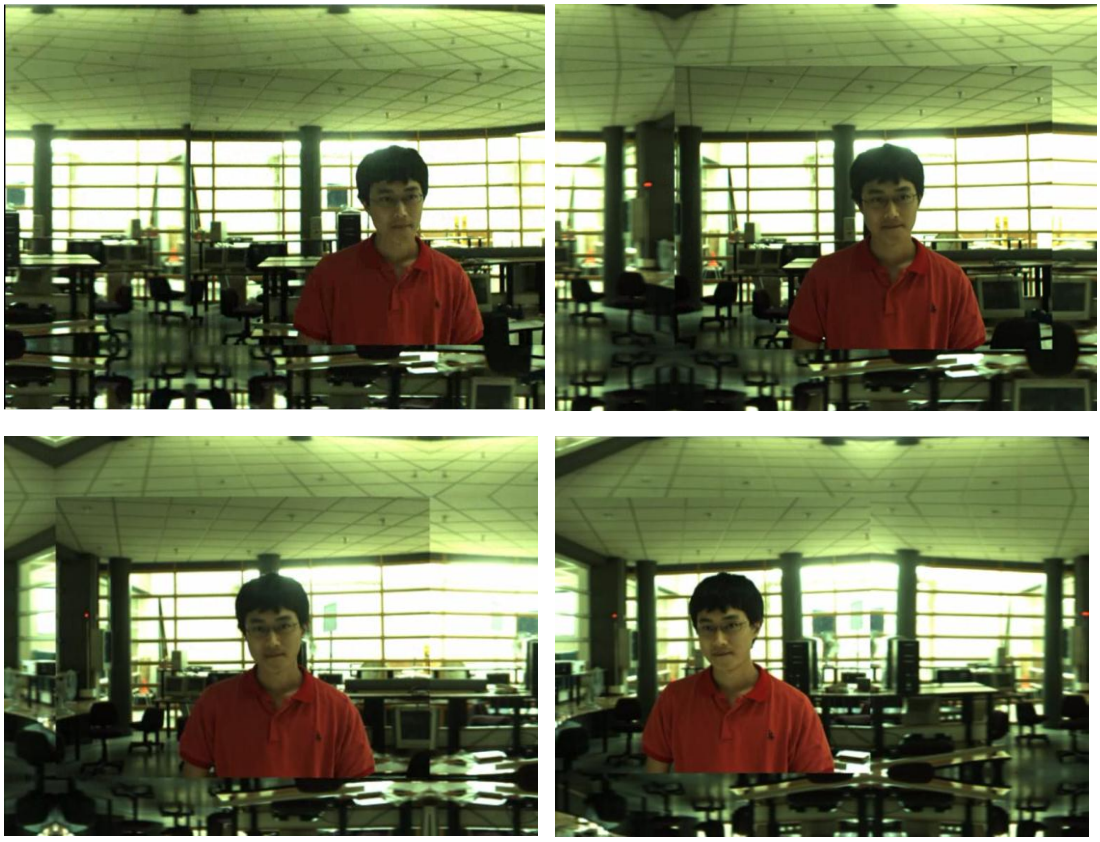


Figura 23**.** Ejemplo imagen generada con trayectoria *Side to Side*.

- **Circular**: El simulador genera una trayectoria para cada persona en el mapa. En esta trayectoria el agente se mueve en torno a la persona a una distancia fija. Los parámetros en este caso son: el radio del movimiento que se encuentra centrado en la persona, el ángulo que define el movimiento y el paso (cuanta distancia avanza entre cada iteración). El agente se mueve solo en la trayectoria.

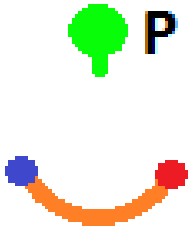


Figura 24**.** Ejemplo movimiento del agente generado con trayectoria circular.

Figura 25**.** Ejemplo imagen generada con trayectoria *Circular*.

- **Strafe**[4]: El simulador genera una trayectoria para cada persona en el mapa. Este movimiento es una mezcla entre 2 movimientos, frontal y Side to side. La idea es que la trayectoria definida se mueve con respecto a una línea imaginaria que sale de la persona. Este movimiento no es a una distancia fija, ya que la trayectoria va acercándose a la persona a medida que va de un lado a otro. El agente se mueve solo en la trayectoria.

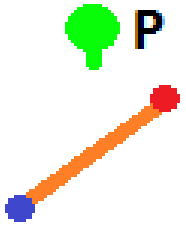


Figura 26**.** Ejemplo movimiento del agente generado con trayectoria Strafe.

[4] Movimiento en el cual el observador se mueve hacia delante-atrás y hacia los lados al mismo tiempo.

Figura 27**.** Ejemplo imagen generada con trayectoria *Stafe*.

- **Random Position**: La idea de este modo es que se posiciona al agente frente a la persona dentro de un cuadrado definido por los parámetros que ingresa el usuario. La localización dentro de este cuadrado es aleatoria. Esto se repite para cada persona dentro del mapa.

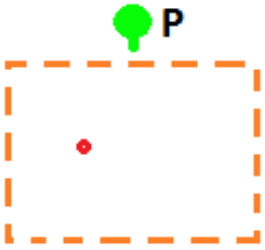


Figura 28**.** Ejemplo área en que el agente es ubicado con trayectoria Random.

Figura 29**.** Ejemplo imagen generada con trayectoria *Random.*

- **Free Trayectory**: Esta trayectoria es la más compleja, ya que la idea es recorrer a todas las personas en el mapa pero no en línea recta. Esta trayectoria tiene diferentes parámetros que están relacionados con la forma de la trayectoria. Primero se define cuál será la trayectoria, dado esto se posiciona a las personas dentro del mapa de modo de que se encuentren cerca de la trayectoria y puedan ser observadas por el agente. Un ejemplo de la trayectoria y las imágenes generadas se puede observar en la Figura 30.

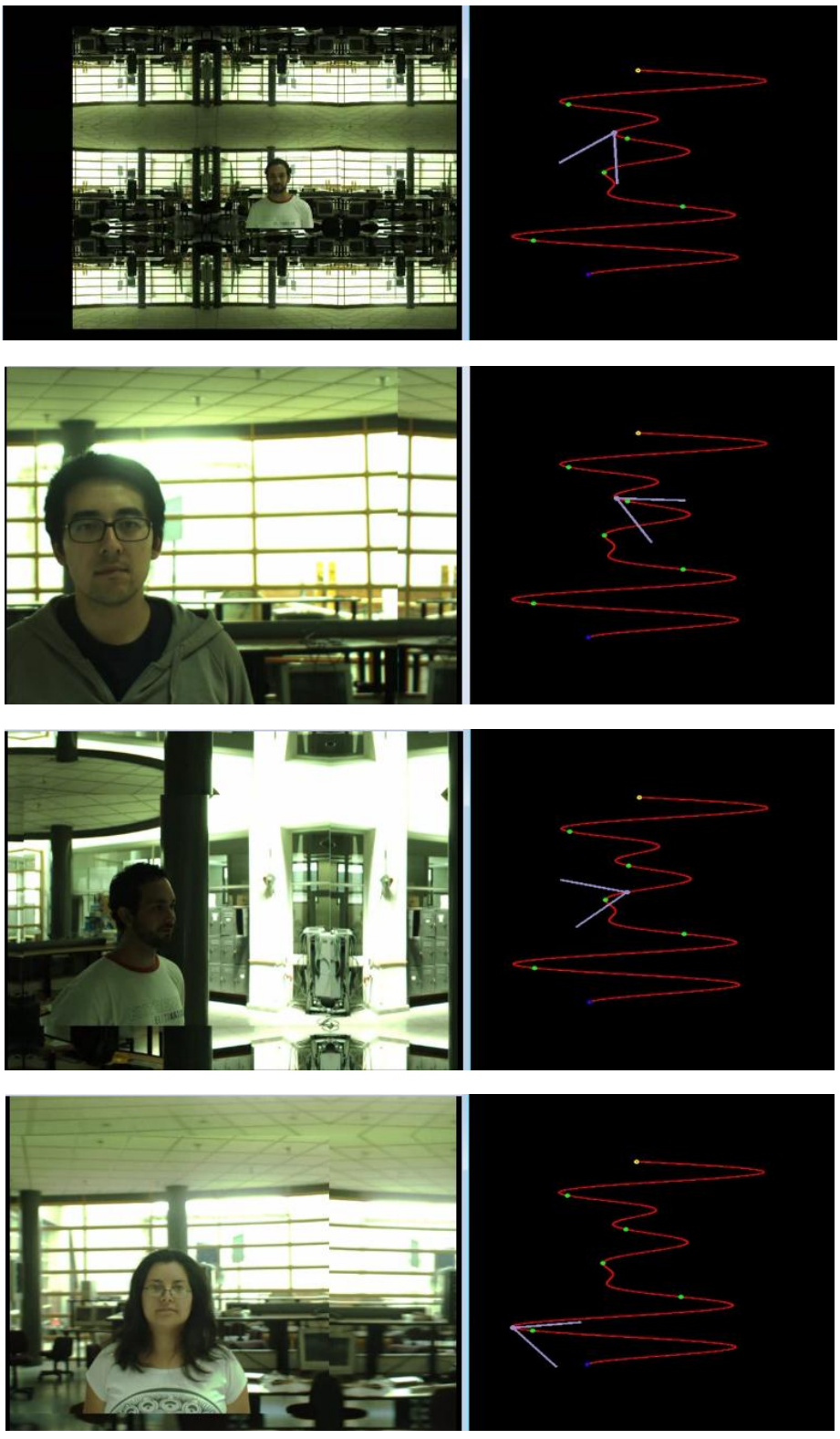


Figura 30**.** Ejemplo imágenes generadas con *Free Trayectory*.

5.4.4 Generación de oclusiones

A cada una de las trayectorias descritas en la sección anterior se les puede agregar oclusión. Esto se hace superponiendo a la imagen un pilar. Este pilar es localizado en el mapa frente a la persona de forma aleatoria con un radio y un ángulo definido. El pilar se posición en la imagen dependiendo si la ubicación del pilar se encuentra dentro del campo visual del agente.

Sea (X_o, Y_o) la ubicación del obstáculo en el mapa, y sea (X_A, Y_A, θ_A) , donde (X_A, Y_A) corresponde a la posición del agente en el mapa y θ_A a la orientación del agente en el mapa.

Se calcula la distancia D_o a la que se encuentra el obstáculo del agente y se calcula el factor de escala SF_o:

$$D_o = \sqrt{(X_o - X_A)^2 + (Y_o - Y_A)^2} \quad (5.13)$$

$$SF_o = \frac{100}{D_o} \quad (5.14)$$

Los 100 cm son debido a que es la distancia a la que fueron tomadas las imágenes del pilar.

Luego se calculan los desplazamientos que tendrá el obstáculo en la imagen dadas su ubicación en el mapa. Se calcula el $Shift_{x_o}$ de la siguiente manera:

$$Shift_{x_o}^{\theta} = 180 + \tan^{-1}\left(\frac{-Y_o + Y_A}{X_o - X_A}\right) - \theta_A \quad (5.15)$$

$$Shift_{x_o} = \frac{Shift_{x_o}^{\theta} * ImageWidth}{FOV_H} \quad (5.16)$$

Donde $Shift_{x_o}^{\theta}$ es el desplazamiento angular del obstáculo con respecto al agente; el $FOV_H = 56.3$ es el FOV (*field of view*) horizontal de la cámara con que se capturó la imagen del pilar.

Usando este desplazamiento ($Shift_{x_o}^{\theta}$) y el cambio de escala (SF_o) se superpone el pilar sobre la imagen generada.

5.4.5 Mapa de personas

El mapa de personas guarda la información de las diferentes personas detectadas dentro del mapa, la idea de este módulo es mantener actualizada la posición de las personas y de esta forma saber cuando a una persona detectada ya ha sido incluida en el mapa. Dado un mapa de tamaño $(Size_X, Size_Y)$. Para cada persona se guarda $(X_i, Y_i, \theta_i, E_i)$ que contiene la posición *x* e *y* en el mapa, la orientación θ_i y E_i *es* el error de estimación. E_i es calculado debido al error asociado a la estimación de la distancia que se encuentra la persona, entre más lejos se encuentre la persona mayor es el error de estimación debido a que se usa el ancho del rostro detectada, y al estar lejos el error de 1 pixel es considerable.

La detección de rostros entrega como salida la ubicación del rostro dentro de la imagen (FD_x, FD_y), más el tamaño del recuadro que enmarca el rostro $(FaceSize_w, FaceSize_H)$. Al usar el detector de rostros, se aplica a distintas ventanas y a distintas escalas de la imagen (ver más detalles en [115]), por lo que pueden ocurrir múltiples detecciones de un mismo rostro.

Esto se utiliza para verificar una detección. Es decir, si muchas ventanas sobrepuestas de tamaños similares son detectadas como rostros, es muy probable que correspondan a un solo rostro. En el caso contrario, en que una ventana aislada haya sido detectada como rostro es poco probable que ésta sea un rostro. Para manejar las detecciones sobrepuestas se realizan distintas técnicas y el resultado depende de cuantas veces el rostro fue detectado a distintas escalas, por lo tanto un rostro que se encuentre a una distancia *x* de la cámara no siempre tendrá el mismo tamaño $(FaceSize_w, FaceSize_H)$.

Esta estimación de la distancia se realiza de la siguiente manera: Sea $(Image_W, Image_H)$ el ancho y el alto de la imagen generada. Se define $ImageFactor$ ya que el cálculo de la distancia depende del tamaño de la imagen:

$$ImageFactor = Image_W / Image_{WBASE} \quad (5.17)$$

$Image_{WBASE}$ es el valor base que se utiliza para calcula el $ImageFactor$, en este caso vale 320.

Luego utilizando la salida del detector de rostros se estima la distancia del rostro detectado en la imagen utilizando el ancho del rostro y el desplazamiento del rostro en la imagen en grados $(Face\theta_x, Face\theta_y)$:

$$D_i^e = \frac{100 * 75 * ImageFactor}{FaceSize_w} \quad (5.18)$$

$$Face\theta_x = \frac{(FD_x - ImageWidth)}{ImageWidth} * \frac{FOV_H}{2} \quad (5.19)$$

$$Face\theta_y = \frac{(FD_y - ImageHeight)}{ImageHeight} * \frac{FOV_V}{2} \quad (5.20)$$

El error E_i se calcula de la siguiente manera:

$$E_i = 50 + \frac{D_i^e}{3} \quad (5.21)$$

Esta ecuación fue calculada estimando la distancia de 50 rostros detectados a diferentes distancias (400 cm, 350 cm, 300 cm, 250 cm, 200 cm, 150 cm y 100 cm) y luego haciendo una regresión lineal con el promedio de los errores en cada posición. En la ecuación 4.18, los números están dados por el tamaño del rostro (75) a una distancia de 100 cm.

El mapa de personas funciona de la siguiente manera. Parte en blanco, dada una nueva detección de rostros se tiene la siguiente información $(D_i^e, Face\theta_x, Face\theta_y, \theta_i, E_i)$ y dado (X_A, Y_A, θ_A), donde (X_A, Y_A) corresponde a la posición del agente y θ_A a la orientación del agente en el mapa. Se proyecta la posición de la persona usando la ubicación del agente, la distancia D_i^e y el ángulo $Face\theta_x$.

$$X_i = D_i^e * \cos(\theta_A + Face\theta_x) + X_A \quad (5.22)$$

$$Y_i = D_i^e * \sin(\theta_A + Face\theta_x) + Y_A \quad (5.23)$$

Una vez que se tiene la posición estimada de la persona busca cual es la persona más cercana, con N_A el número de personas almacenadas:

$$\min_i(D_{ij}) \text{ con } j\epsilon\{1,..,N_A\},\ ID_{Selecionado} = ID_{min}$$
$$s.a.\ (1)\quad D_{ij} < \theta_j$$
$$con\ D_{ij} = \sqrt{(X_i - X_j)^2 + (Y_i - Y_j)^2} \qquad (5.24)$$

Si existe una persona almacenada que cumpla las restricciones se remplazan los datos siempre que se cumpla que:

$$\theta_i < \theta_j \qquad (5.25)$$

De lo contrario se deja a la persona con los datos que están. Si no se encuentra a ninguna persona que cumpla con las restricciones se agrega a la persona actual al mapa.

5.4.6 Filtros de Contexto

El primer filtro que se utiliza es el filtro de contexto de la situación, que está relacionado a la actividad se está realizando, con la escena o ambiente en que se está inmerso, determinan un contexto a nivel de la situación en que se encuentra el agente. En el contexto del mapa del simulador las personas se encuentran de pie, tienen una altura promedio de 1.60 metros que se encuentra definido dentro de los parámetros del simulador, esta información se utiliza para agregar información a otros filtros de contexto. Esta información limita la búsqueda de personas ya que no pueden estar sentadas o acostadas.

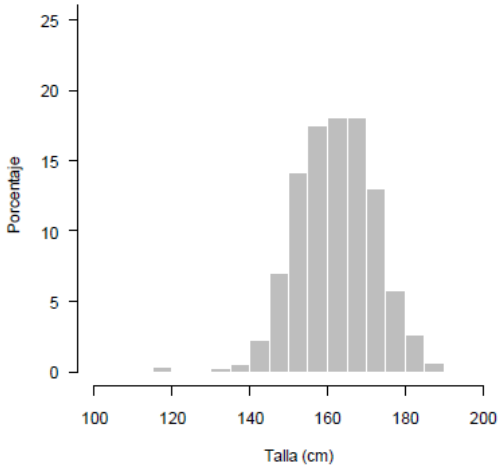


Figura 31. Distribución de talla en la población general. Chile 2009-2010. Fuente: ENS Chile 2009-2010.

En la encuesta nacional de salud ENS Chile 2009-2010 se presentan los promedios nacionales de seis medidas antropométricas, entre ellos la altura promedio en Chile que es de 1.62 metros, con un mínimo de 1.18 metros y un máximo de 2.03 metros. Estas cifras son respecto a los encuestados cuyo rango de edad esta entre los 15 y 70 años. En la Figura 31 se puede observar la distribución de talla de la población chilena, donde la mayor parte de la población se ubica entre los 1.40 y 1.90 de estatura. Usando esta información más la información del filtro de contexto de situación se crearon otros dos filtros de contextos que ayudan a descartar detecciones falsas y así evitar agregar al mapa de personas información errónea. El primer filtro de contexto es el filtro de contexto físico espacial que se refiere a la información determinada a través de las leyes físicas que rigen el ambiente donde opera el agente. El modelo físico del ambiente define un modelo visual, en este caso la existencia de un

piso sobre el cual deben estar las personas, por lo que se descartan personas que no cumplan con estas restricciones. El segundo filtro de contexto es también es un filtro de contexto físico espacial, en este caso la altura de las personas esta acotada usando la información de [28]. Además usando contexto de configuración de objetos (o coherencia espacial), que se refiere a las relaciones espaciales que deben cumplir entre las personas, se puede descartar que 2 personas estén ubicadas en el mismo punto en el mapa, por lo que se supone que es la misma persona.

Estos tres filtros de contexto se utilizan antes de agregar información al mapa de personas, de esta forma se logra descartar información de detecciones falsas.

5.5 Experimentos

5.5.1 Estudio comparativo [5]

En esta sección se presenta un estudio comparativo de los diferentes métodos de reconocimiento presentados en la sección 3.2. La intención de este estudio es validar los resultados obtenidos en otros trabajos relacionados ([114][101]). Con el fin de reconocer los rostros correctamente, el agente debe tener los siguientes módulos disponibles: (i) Detección de Rostros. El agente detecta la posición del rostro dentro de la imagen. (ii) Estimación de la pose del rostro. El agente estima la rotación del rostro en *yaw*, *pitch* y *roll*, (iii) Visión Activa. Utilizando la información acerca del rostro detectado y sus rotaciones, el agente puede decidir cambiar el punto de vista de la observación para modificar la percepción del rostro; (iv) Reconocimiento de rostros. La identidad de la persona que figura en la imagen es determinada. El módulo puede incluir capacidades tales como alineación de rostros o compensación de iluminación.

En el ambiente virtual planteado en este capítulo, los módulos utilizados se encuentran descritos en las secciones 5.3 y 5.4. El simulador puede proporcionar las funciones que en el sistema de reconocimiento facial no se incluyen. En este caso, se proporciona la estimación de los ángulos de rotación del rostro detectado.

Con el fin de obtener una primera validación de la aplicabilidad de esta herramienta de evaluación, se llevan a cabo dos experimentos. En el primer experimento se comparan 3 métodos de detección de rostros. En el segundo experimento se comparan tres métodos de reconocimiento de rostros en diferentes condiciones, con y sin el uso de visión activa.

- **Detección de Rostros**

  Se comparan tres métodos de detección de rostros. En primer lugar, el detector de cascada *Haar* que proporciona *OpenCV* [76] (*HaarTraining* de *OpenCV*). Este detector se basa en el detector en cascada que se describe en [91], que es una extensión del detector de rostros clásico de Viola y Jones [83], y que utiliza el algoritmo de entrenamiento *Gentleboost* y características *Haar*. En segundo lugar se evalúa un detector de rostros propuesto por Verschae et al. [115], que utiliza clasificadores de tipo *boosted* en cascada. En tercer lugar, un detector de rostros propuesto por Verschae en [93], que mejora el rendimiento del segundo detector mediante el uso de un enfoque de

[5] Esta sección esta basada en los artículos [69] y [20].

coarse-to-fine en el entrenamiento de las cascadas. Estos tres detectores son llamados *HaarCascade*, *NestedCascade1*, y *NestedCascade2*, respectivamente.

Se compara la tasa de detección y el número de falsos positivos de los tres detectores usando diferentes parámetros; el ángulo de *yaw* de los rostros observados se selecciona de manera aleatoria en un rango de $\pm\theta^{y}_{max}$, de la misma forma se elige el ángulo de *pitch* en el rango $\pm\theta^{P}_{max}$. Otros parámetros del simulador se mantienen sin cambios ($\Delta x_{\max} = \Delta y_{\max} = \Delta\theta_{\max} = \theta^{r}_{\max} = 0$).

- **Reconocimiento de rostros**

Se comparan tres métodos de reconocimiento de rostros: histogramas de características LBP, Gabor-Jet con características Borda, y los histogramas de características WLD (*Weber Local Descriptor*). Los dos primeros métodos han demostrado un muy buen desempeño en los estudios comparativos anteriores [51][101]. El tercer método se propone en [20], y se basa en las características WLD propuestas en [52]. En todos los casos, los parámetros de los métodos se sintonizan utilizando conjuntos de datos estándar, y no utilizando el simulador para este fin. Una descripción completa de los métodos de reconocimiento se presenta en la sección 3.2.

Usando los resultados presentados en [101], se implementan dos versiones diferentes del método: histogramas de características LBP, uno con la intersección de histograma (HI), y otro con Chi cuadrado (XS) como medida de similitud. En ambos casos, las imágenes de rostros se cortan a un tamaño de 81x150 píxeles y se divide en 40 regiones para calcular los histogramas LBP. Los sistemas se llaman LBP-HI-40 y LBP-XS-40, respectivamente. El método de Gabor implementado utiliza 5 escalas, orientaciones 8, y un tamaño de imagen de 122x225 píxeles, como se describe en [101]. Por último, en el caso del método basado en WLD, después de una amplia experimentación con las bases de datos FERET, BioID y LFW, los parámetros seleccionados fueron los siguientes: intersección de histograma y Chi-cuadrado medidas de similitud, imágenes de tamaño 93x173 píxeles, se divide en 40 regiones para calcular los histogramas WLD, se usan 2 orientaciones dominantes (T = 2), y 26 células en cada orientación (C = 26).

Se compara las tasas de reconocimiento usando diferentes parámetros; el ángulo de *yaw* de los rostros observados se selecciona de manera aleatoria en un rango de $\pm\theta^{y}_{max}$, de la misma forma se elige el ángulo de *pitch* en el rango $\pm\theta^{P}_{max}$. Otros parámetros del simulador se mantienen sin cambios ($\Delta x_{\max} = \Delta y_{\max} = \Delta\theta_{\max} = \theta^{r}_{\max} = 0$).

5.5.2 Evaluación de módulos

Estos experimentos son para evaluar el funcionamiento de los algunos módulos implementados (Detector de rostros, reconocimiento de rostros, módulos de contexto), casi todos se realizan con la trayectoria “Free”, excepto el experimento 8, debido a que es la trayectoria que presenta mayor dificultad y además es la que se acerca mas a lo que se desea realiza con el robot. Debido a que se utilizan variables aleatorias para generar los experimentos, cada experimento se repite 10 veces para que los resultados sean estadísticamente significativos.

5.5.2.1 Definición de experimentos

Los parámetros de los experimentos están definidos en la siguiente sección. Los experimentos que se realizan son los siguientes:

1. **Detección de Personas**: El agente recorre un mapa tratando de detectar la mayor cantidad de rostros posibles. Para esto se genera una trayectoria del agente y la correspondiente ubicación de las personas en el mapa, de la forma descrita anteriormente. El agente guarda información sobre la localización de los rostros en el Mapa de personas (ver sección 5.4.5) para saber que rostros/personas ya fueron detectadas.

2. **Detección de rostros, Reconocimiento de rostros, con construcción offline de DB reconocimiento**: El agente posee una galería o base de datos con los rostros de la arena construida offline, para esto se agrega 1 imagen de cada sujeto en la galería. La imagen utilizada no posee rotaciones en ningún eje. Se genera una trayectoria del agente y la correspondiente ubicación de las personas en el mapa, de la forma descrita anteriormente. El agente NO puede salir de la trayectoria para modificar sus observaciones. El agente debe detectar y reconocer a las N rostros que hay en el mapa.

3. **Detección de rostros, Reconocimiento de rostros, con construcción online de DB reconocimiento**: El agente NO posee una galería con los rostros de la arena y la construirá a medida que va detectando rostros. Para esto se genera una trayectoria del agente y la correspondiente ubicación de las personas en el mapa, de la forma descrita anteriormente. El agente NO puede salir de la trayectoria para modificar sus observaciones. El agente debe detectar y reconocer a los rostros. Se hacen 2 pasadas por la trayectoria, una para hacer la galería y otra para reconocer. Las personas se mueven entre una pasada y otra, para esto las personas se vuelven a posicionar de forma aleatoria, la idea es que no se posea información a priori de las personas en la segunda pasada.

4. **Detección de rostros, Reconocimiento de rostros y Visión activa, con construcción offline de DB reconocimiento**: El agente posee una galería con las personas de la arena construida offline, para esto se agrega 1 imagen de cada sujeto en la galería. La imagen utilizada no posee rotaciones en ningún eje. Para esto se genera una trayectoria del agente y la correspondiente ubicación de las personas en el mapa, de la forma descrita anteriormente. El agente puede salir de la trayectoria para modificar sus observaciones, para esto utiliza visión activa. El agente puede utilizar información del rostro detectado para acercarse/alejarse y o rotar, luego volver a generar otra observación. . El agente debe detectar y reconocer a las N rostros que hay en el mapa.

5. **Detección de rostros, Reconocimiento de rostros y Visión activa, con construcción online de DB reconocimiento**: El agente NO posee una galería con los rostros de la arena y se construirá a medida que va detectando rostros. Para esto se genera una trayectoria del agente y la correspondiente ubicación de las personas en el mapa, de la forma descrita anteriormente. El agente puede salir de la trayectoria para modificar sus observaciones, para esto utiliza visión activa. El agente puede utilizar información del rostro detectado para acercarse/alejarse y o rotar, luego volver a generar otra

observación. El agente debe detectar y reconocer a las N rostros que hay en el mapa. Se hacen 2 pasadas por la trayectoria, una para hacer la galería y otra para reconocer. Las personas se mueven entre una pasada y otra, para esto las personas se vuelven a posicionar de forma aleatoria, la idea es que no se posea información a priori de las personas en la segunda pasada.

6. **Reconocimiento de rostros y Visión activa, con construcción offline de DB reconocimiento**: El agente posee una *galería* con los rostros de la arena, para crearla se agrega 1 imagen de cada sujeto. La imagen utilizada no posee rotaciones en ningún eje. La detección de rostros es perfecta ya que se utiliza el *groundtruth* que es generado por el simulador. Para esto se genera una trayectoria del agente y la correspondiente ubicación de las personas en el mapa, de la forma descrita anteriormente. El agente puede salir de la trayectoria para modificar sus observaciones, para esto utiliza visión activa. El agente puede utilizar información del rostro detectado para acercarse/alejarse y o rotar, luego volver a generar otra observación. El agente solo debe reconocer a las N rostros, dado que la información de la posición de los rostros es dado por el simulador.

7. **Reconocimiento de rostros y Visión activa, con construcción online de DB reconocimiento**: El agente NO posee una galería con los rostros de la arena. La detección de rostros es perfecta ya que se utiliza el *groundtruth*. Se genera una trayectoria del agente y la correspondiente ubicación de las personas en el mapa, de la forma descrita anteriormente. El agente puede salir de la trayectoria para modificar sus observaciones, para esto utiliza visión activa. El agente puede utilizar información del rostro detectado para acercarse/alejarse y o rotar, luego volver a generar otra observación. El agente solo debe reconocer a las N rostros que hay en el mapa, dado que la información de la posición de los rostros es dado por el simulador. Se hacen 2 pasadas por la trayectoria, una para hacer la galería y otra para reconocer. Las personas se mueven entre una pasada y otra, para esto las personas se vuelven a posicionar de forma aleatoria, la idea es que no se posea información a priori de las personas en la segunda pasada.

8. **Detección de Rostros, reconocimiento de rostros. Con y sin visión activa, con construcción offline de DB reconocimiento**: El agente posee una galería con los rostros de la arena, para crearla se agrega 1 imagen de cada sujeto en la galería. La imagen utilizada no posee rotaciones en ningún eje. Se utiliza la trayectoria random, que localiza al agente frente a la persona agregándole un ruido a la posición del agente. El agente puede moverse para modificar sus observaciones si esta usando visión activa. El agente debe reconocer a los rostros. En este caso el número de personas es N=20. El resto de los parámetros definidos son incluidos en los resultados.

5.5.2.2 Parámetros

Cada experimento se realizó con los mismos parámetros para que sean comparables. Es decir se generaron un conjunto de parámetros con los cuales se realizarán los experimentos del 1 al 7. De esta forma se pueden comparar los diferentes métodos. Cada experimento se repitió 10 veces para que los resultados sean estadísticamente significativos, en cada repetición se generó un nuevo conjunto de parámetros que se mantendrá constante en cada experimento. O

sea con un conjunto de parámetros dado se realizan los experimentos del 1 al 7, luego se genera otro conjunto de parámetros y se repiten los experimentos, esto se repitió 10 veces.

El número de personas (N) utilizado en los experimentos 1 al 7 es 10, esto debido a la carga computacional, además se estimó que esa cantidad de personas es suficiente para evaluar los módulos. Para el experimento 8 se utilizan 20 personas debido a que la carga computacional es menor.

El detector de rostros utilizado es el descrito en la sección 4.3.1. El detector de ojos se utilizó para alinear los rostros antes de ser reconocidos. (Ver sección □).

La altura del agente (H_A) es fija e igual a la altura base de las personas (1.60 mts).

Para cada experimento se realizan con 2 conjunto de parámetros diferentes, el primer conjunto de parámetros es el siguiente:

- Ruido en ángulo $R_\theta \in [-5°, 5°]$
- Ruido en desplazamiento: $R_P \in [-50\ cm, 50\ cm]$
- Ruido Altura de personas: $R_H \in [-15\ cm, 15\ cm]$
- Ruido en la trayectoria: $R_H \in [-3°, 3]$

El segundo conjunto de parámetros es de mayor dificultad y es el siguiente:

- Ruido en ángulo $R_\theta \in [-10°, 10°]$
- Ruido en desplazamiento: $R_P \in [-100\ cm, 100\ cm]$
- Ruido Altura de personas: $R_H \in [-15\ cm, 15\ cm]$
- Ruido en la trayectoria: $R_H \in [-3°, 3]$

Además para realizar estos experimentos se definen estos parámetros:

- **Altura de las personas (*H*)**: Hay una altura base (H_B) de 1.60 mts, y se modifica esta altura agregándole una variable aleatoria (R_H) de distribución uniforme. Los valores que toma esta variable aleatoria están acotados entre -15 y +15.

$$H_i = H_B + R_H \quad i \in \{1, .., N\} \tag{5.26}$$

- **Generación de trayectoria (*G*)**: La trayectoria se genera como una suma ponderada de senos, la frecuencia de la función tiene un valor base más un valor aleatorio (R_G) de distribución uniforme. Los valores que toma esta variable aleatoria están acotados entre -3 y +3 (esta variable es adimensional y los valores están definidos para que la trayectoria no se salga del mapa). Esta trayectoria se discretiza con un paso fijo, de esta forma se generan los diferentes puntos de la trayectoria.

$$G(\theta) = A * \left(\sin\left(B_1 * \frac{\theta}{f}\right) + \sin\left(B_2 * \frac{\theta}{f}\right) + \sin\left(B_3 * \frac{\theta}{f}\right)\right) \tag{5.27}$$

$$Con\ \theta = \frac{i * \pi}{T} * PlotScale, \qquad i \in [0, .., Steps] \tag{5.28}$$

$$f = f_B + R_G \tag{5.29}$$

Donde

$$A = \frac{M * MapSize}{N} \quad (5.30)$$

$$Steps = \frac{M * MapSize}{PlotScale} \quad (5.31)$$

La variable $PlotScale$ es para escalar el mapa al mostrarlo en pantalla, $Steps$ indica en cuantos puntos se va a discretizar la trayectoria generada. La amplitud (A) es dependiente del tamaño del mapa. En la ecuación (4.26), $B_1 = 1, B_2 = \sqrt{2}, B_3 = \sqrt{3}$ y están definidos para darle una forma irregular y no periódica a la trayectoria. En la ecuación (4.27), $T = 24$ y esta fijado de modo que asegurar que la trayectoria tenga un número mínimo de pasadas por cero. En la ecuación 4.28, $f_B = 15$ esta fijo de manera de darle una frecuencia base a la trayectoria. En las ecuaciones (4.29) y (4.30), $M = \frac{3}{5}, N = 6$ y están definidos para evitar que la trayectoria salga del mapa. Todos estos valores están definidos de manera empírica.

- **Posición de las personas** ($\boldsymbol{x_i^f, y_i^f, \theta_i^f, R_P, R_\theta}$): La posición de las personas en la trayectoria se fija de la siguiente manera, las *N* personas se distribuyen en la trayectoria en distancias iguales en el eje x, luego se busca el punto de la trayectoria mas cercano en el eje Y, dada esta posición se agrega ruido R_P a la posición de la persona en cada eje. La pose de la persona se fija utilizando la derivada de la trayectoria en el punto mas cercano, a este valor se le agrega un ruido R_θ. R_P *y* R_θ son variables aleatorias de distribución uniforme.

  Sea X_{Init}^G e Y_{Init}^G los puntos iniciales de la trayectoria en el mapa.

$$y_i = Y_{Init}^G + i * PlotScale \text{ con } i \in [0, .., Steps] \quad (5.32)$$

$$x_i = G(y_i) = A * \left(\sin\left(\frac{y_i}{f}\right) + \sin\left(\sqrt{2} * \frac{y_i}{f}\right) + \sin\left(\sqrt{3} * \frac{y_i}{f}\right)\right) \quad (5.33)$$

$$x_i^f = x_i + R_P \quad (5.34)$$

$$y_i^f = y_i + R_P \quad (5.35)$$

$$\theta_i^f = \frac{d(G(\theta))}{d\theta} + R_\theta \quad (5.36)$$

5.6 Resultados y discusión

En esta sección se presentan los resultados de los diferentes experimentos definidos en la sección anterior. Primero se presentan los resultados del estudio comparativo y se discuten los resultados. Luego se presentan los resultados de la evaluación de los diferentes módulos del sistema de visión planteado. En las tablas se muestra en negrita un resumen de los experimentos realizadas. Las tablas completas con cada una de los experimentos se puede observar en el anexo A.

5.6.1 Estudio comparativo (Detección y reconocimiento de rostros)[6]

En la Tabla 3 se presentan los resultados de la comparación de los diferentes métodos de detección de rostros, de los resultados obtenidos se puede concluir lo siguiente:

- *NestedCascade2* tiene un mejor desempeño que los otros detectores de rostros, especialmente porque tiene un número pequeño de falsos positivos, algo que es muy importante en aplicaciones en robots de servicio.
- En la mayoría de los casos las tasas de detección disminuyen al utilizar imágenes *outdoor*, esto es debido a la iluminación es más irregular.
- Para rotaciones pequeñas de menos de 20° en outdoor, *NestedCascade1 y NestedCascade2* funcionan mejor que *HaarCascade*.
- Los métodos de detección funcionan bien cuando las rotaciones de los rostros están acotadas a 40°. Cuando las rotaciones son mayores la tasa de detecciones disminuye considerablemente.
- Se puede apreciar que cuando las rotaciones en *yaw* son pequeñas, de menos de 20°, las rotaciones en *pitch* no afectan significativamente las tasas de detección.

En la Tabla 4 se puede apreciar los resultados de la evaluación de métodos de reconocimiento de rostros, de los resultados se puede concluir:

- En *indoor* los métodos basados en características LBP son más robustos ante rotaciones de los rostros que los métodos Gabor y WLD.
- En *outdoor* el método de Gabor-jets tiene el mejor rendimiento.
- Todos los métodos implementados son robustos ante rotaciones menores a 40°.
- El rendimiento de todos los métodos de reconocimiento disminuye notablemente cuando se usan imágenes *outdoor*.
- El uso de visión activa incrementa el rendimiento de todos los métodos de reconocimiento, especialmente frente a rotaciones fuera del plano.

Tabla 3: Resultados evaluación de métodos de detección de rostros

	$\theta^P_{max} = 0°$															
	$\theta^y_{max} = 0°$				$\theta^y_{max} = 20°$				$\theta^y_{max} = 40°$				$\theta^y_{max} = 60°$			
	Indoor		**Outdoor**		**Indoor**		**Outdoor**		**Indoor**		**Outdoor**		**Indoor**		**Outdoor**	
	DR	**FP**	**DR**	**FP**	**DR**	**FP**	**DR**	**FP**	**DR**	**FP**	**DR**	**FP**	**DR**	**FP**	**DR**	**FP**
HaarCascade	97.4	2	78.3	5	92.1	3	87.0	3	71.1	4	65.2	8	47.4	1	62.6	4
NestedCascade1	89.5	4	100	0	86.8	5	100	0	71.1	6	**100**	**0**	**55.3**	**9**	69.6	3
NestedCascade2	**100**	**0**	100	0	**97.4**	**0**	**100**	**0**	**73.7**	**1**	95.7	0	**50.0**	**2**	**73.9**	**1**
	$\theta^P_{max} = \pm 15°$															
	$\theta^y_{max} = 0°$				$\theta^y_{max} = 20°$				$\theta^y_{max} = 40°$				$\theta^y_{max} = 60°$			
	Indoor		**Outdoor**		**Indoor**		**Outdoor**		**Indoor**		**Outdoor**		**Indoor**		**Outdoor**	
	DR	**FP**	**DR**	**FP**	**DR**	**FP**	**DR**	**FP**	**DR**	**FP**	**DR**	**FP**	**DR**	**FP**	**DR**	**FP**
HaarCascade	94.7	2	95.7	1	84.2	4	69.6	7	65.8	6	78.3	5	60.5	6	65.2	8
NestedCascade1	94.7	2	100	0	84.2	4	100	0	60.5	11	100	0	42.1	12	69.6	4
NestedCascade2	97.4	0	100	0	94.7	0	100	0	63.2	1	95.7	0	52.6	3	82.6	1

[6] Esta sección está basada en los artículos [20] y [69].

Tabla 4: Resultados evaluación de métodos de reconocimiento.

Método		*Indoor*					
	Visión Activa	$\theta^{y}_{\max} = \pm 20$ $\theta^{p}_{\max} = 0$	$\theta^{y}_{\max} = \pm 20$ $\theta^{p}_{\max} = \pm 15$	$\theta^{y}_{\max} = \pm 40$ $\theta^{p}_{\max} = 0$	$\theta^{y}_{\max} = \pm 40$ $\theta^{p}_{\max} = \pm 15$	$\theta^{y}_{\max} = \pm 60$ $\theta^{p}_{\max} = 0$	$\theta^{y}_{\max} = \pm 60$ $\theta^{p}_{\max} = \pm 15$
LBP-HI-40	*No*	86,8	73,7	73,7	50,0	42,1	28,9
	SI	94,7	78,9	97,4	92,1	92,1	84,2
LBP-XS-40	*No*	92,1	73,7	78,9	52,6	44,7	31,6
	SI	**97,4**	81,6	**97,4**	**92,1**	**92,1**	**84,2**
GJD-BC	*No*	92,1	68,4	76,3	26,3	34,2	15,8
	SI	94,7	76,3	86,8	68,4	**92,1**	63,2
WLD-HI-40	*No*	65,8	71,1	50,0	57,9	31,6	36,8
	SI	84,2	**86,8**	86,8	73,7	81,6	65,8
WLD-XS-40	*No*	63,2	68,4	47,4	60,5	28,9	34,2
	SI	84,2	78,9	84,2	73,7	81,6	65,8
	Outdoor						
	Visión Activa	$\theta^{y}_{\max} = \pm 20$ $\theta^{p}_{\max} = 0$	$\theta^{y}_{\max} = \pm 20$ $\theta^{p}_{\max} = \pm 15$	$\theta^{y}_{\max} = \pm 40$ $\theta^{p}_{\max} = 0$	$\theta^{y}_{\max} = \pm 40$ $\theta^{p}_{\max} = \pm 15$	$\theta^{y}_{\max} = \pm 60$ $\theta^{p}_{\max} = 0$	$\theta^{y}_{\max} = \pm 60$ $\theta^{p}_{\max} = \pm 15$
LBP-HI-40	*No*	56,5	47,8	47,8	39,1	26,1	21,7
	SI	60,9	52,2	60,9	43,5	52,2	52,2
LBP-XS-40	*No*	52,2	52,2	34,8	30,4	21,7	17,4
	SI	56,5	52,2	39,1	39,1	43,5	47,8
GJD-BC	*No*	56,5	43,5	34,8	26,1	47,8	39,1
	SI	**69,6**	**65,2**	**65,2**	**52,2**	**69,6**	**60,9**
WLD-HI-40	*No*	34,8	47,8	39,1	21,7	30,4	17,4
	SI	52,2	47,8	47,8	34,8	47,8	30,4
WLD-XS-40	*No*	39,1	34,8	34,8	17,4	30,4	17,4
	SI	52,2	47,8	52,2	39,1	52,2	30,4

5.6.2 Evaluación de módulos

1. Detección de personas

Este experimento es para evaluar la detección de rostros dentro del simulador, y al mismo tiempo evaluar los módulos de filtro de contexto que se utiliza para descartar falsos positivos. En este caso se utilizan 3 filtros de contexto: (1) filtro de contexto de la situación, (2) filtro de contexto físico espacial y (3) filtro de coherencia espacial.

En la Tabla 5 se pueden observar los resultados del primer conjunto de parámetros, se ve que el número de detección de rostros es en promedio de 192.7 de un total de 215.7 en cada recorrido de la trayectoria, de las 192.7 se descartan el número falsas detecciones de personas usando los filtros de contexto (en promedio 27.9), estas personas son descartadas porque no cumplen con las leyes físicas o las alturas determinadas por el sistema están fuera de los parámetros correctos. Finalmente solo se agregan al mapa de personas 15.6 personas en promedio, por lo que el filtro de coherencia espacial funciona descartando detecciones cuando la persona es la misma, y no agregándola al mapa de personas. En este caso el promedio de las personas detectadas es 8.4 de 10, con 7.2 falsos positivos.

Tabla 5: Resumen de resultados del experimento 1. Conjunto de parámetros 1.

Parámetros: Ruido en ángulo: $R_\theta \in [-5°, 5°]$, Ruido en desplazamiento: $R_P \in [-50\ cm, 50\ cm]$, Ruido Altura de personas: $R_H \in [-15\ cm, 15\ cm]$, Ruido en la trayectoria: $R_G \in [-3,3]$ y Número de personas = 10

Parámetros: Ruido en ángulo = , Ruido en desplazamiento = 50 cm				
Rostros	**Personas**			
Número de rostros detectados / Total de rostros o *Ground Truth*	**Número de personas detectadas**	**Número de Personas descartadas usando filtro de contexto**	**Número de personas correctamente detectadas**	**Falsos Positivos**
192.7 / 215.7	**15.6**	**27.9**	**8.4 / 10**	**7.2**

En la Tabla 6 se pueden apreciar los resultados del segundo conjunto de parámetros, se puede observar que el número de detección de rostros es en promedio de 261 en cada recorrido de la trayectoria, de estas detecciones se descartan el número falsas detecciones de personas usando los filtros de contexto (en promedio 22.2), estas personas son descartadas porque no cumplen con las leyes físicas o las alturas determinadas por el sistema están fuera de los parámetros correctos. Luego solo se agregan al mapa de personas 14.6 personas en promedio, por lo que el filtro de coherencia espacial funciona nuevamente. En este caso el promedio de las personas detectadas es 6.4 de 10, con 8.2 falsos positivos. Dado que esta prueba es más difícil el número de rostros y personas detectadas disminuye lo que es coherente.

Tabla 6: Resumen de resultados del experimento 1. Conjunto de parámetros 2.

Parámetros: Ruido en ángulo: $R_\theta \in [-10°, 10°]$, Ruido en desplazamiento: $R_P \in [-100\ cm, 100\ cm]$, Ruido Altura de personas: $R_H \in [-15\ cm, 15\ cm]$, Ruido en la trayectoria: $R_G \in [-3,3]$ y *Número de personas* =10

Parámetros: Ruido en ángulo = 10, Ruido en desplazamiento = 100 cm				
Rostros	**Personas**			
Número de rostros detectados / Total de rostros o *Ground Truth*	**Número de personas detectadas**	**Número de Personas descartadas usando filtro de contexto**	**Número de personas correctamente detectadas**	**Falsos Positivos**
261/275.9	**14.6**	**22.2**	**6.4 / 10**	**8.2**

2. **Detección de rostros, Reconocimiento de rostros, con construcción offline de DB reconocimiento**

Esta prueba es para evaluar el reconocimiento de rostros con detección automática de rostros, tomando en cuenta que la galería se hace de antemano usando rostros sin rotaciones en ningún eje. En este caso se utilizan 3 filtros de contexto: (1) filtro de contexto de la situación, (2) filtro de contexto físico espacial y (3) filtro de coherencia espacial.

En la Tabla 7 se pueden ver los resultados del primer conjunto de parámetros (se muestran 2 ejemplos y el promedio de las 10 pruebas), se ve que el número de detección de rostros es en promedio de 192.7 en cada recorrido de la trayectoria. En este caso el promedio de las personas detectadas correctamente es 8.4 de 10, con 7.2 falsos positivos. Se presentan 2 tasas de reconocimiento, la primera sin el uso del mapa de personas, o sea cada vez que se detecta una persona se intenta reconocer, con esto se tiene una tasa de reconocimiento de 78.41% (el valor de "Reconocimiento Correctas/Total" es diferente para cada iteración, ver Tabla 32 en el anexo A), al usar

el mapa de personas y los filtros de contexto la tasa de reconocimiento sube a un 86.77% lo que muestra que el uso de contexto mejora el desempeño del sistema de visión.

Tabla 7: Resumen de resultados del experimento 2. Conjunto de parámetros 1.

Parámetros: Ruido en ángulo: $R_\theta \in [-5°, 5°]$, Ruido en desplazamiento: $R_P \in [-50\ cm, 50\ cm]$, Ruido Altura de personas: $R_H \in [-15\ cm, 15\ cm]$, Ruido en la trayectoria: $R_G \in [-3,3]$ y *Número de personas* =10

Parámetros: Ruido en ángulo = 5, Ruido en desplazamiento = 50 cm, Galería: 10 personas					
Número de rostros detectados / Total de rostros o *Ground Truth*	**Número de personas correctamente detectadas**	**Sin Mapa de Personas**		**Con Mapa de Personas**	
		Reconocimiento Correctas/Total	**Reconocimiento %**	**Reconocimiento Correctas/Total**	**Reconocimiento %**
288 / 305	9	229/255	89.80%	7/9	77.78%
181 / 201	8	112/161	69.57%	7/8	87.50%
192.7/215.7	**8.4 / 10**	**-----**	**78.41%**	**-----**	**86.77%**

En la Tabla 8 se pueden apreciar los resultados del segundo conjunto de parámetros (se muestran 2 ejemplos y el promedio de las 10 pruebas), se ve que el número de detección de rostros es en promedio de 261 de un total de 275.9 rostros en cada recorrido de la trayectoria. En este caso el promedio de las personas detectadas correctamente es 6.4 de 10, con 8.2 falsos positivos. Nuevamente se presentan 2 tasas de reconocimiento, la primera sin el uso del mapa de personas con lo que se obtiene una tasa de reconocimiento de 79.13% (el valor de "Reconocimiento Correctas/Total" es diferente para cada iteración, ver Tabla 33 en el anexo A), al usar el mapa de personas y los filtros de contexto la tasa de reconocimiento sube a un 87.70% lo que muestra que el uso de contexto puede mejorar el desempeño del sistema de visión. Las tasas de reconocimiento son levemente mejores pero el porcentaje esta calculado en función del número de personas correctamente detectadas que es menor, de todas maneras se puede observar que las tasas de reconocimiento están relacionadas entre una prueba y otra. Los resultados de esta prueba son coherentes dada la mayor dificultad del segundo conjunto de parámetros.

Tabla 8: Resumen de resultados del experimento 2. Conjunto de parámetros 2.

Parámetros: Ruido en ángulo: $R_\theta \in [-10°, 10°]$, Ruido en desplazamiento: $R_P \in [-100\ cm, 100\ cm]$, Ruido Altura de personas: $R_H \in [-15\ cm, 15\ cm]$, Ruido en la trayectoria: $R_G \in [-3,3]$ y *Número de personas* =10

Parámetros: Ruido en ángulo = 10, Ruido en desplazamiento = 100 cm, Galería: 10 personas					
Número de rostros detectados / Total de rostros o *Ground Truth*	**Número de personas correctamente detectadas**	**Sin Mapa de Personas**		**Con Mapa de Personas**	
		Reconocimiento Correctas/Total	**Reconocimiento %**	**Reconocimiento Correctas/Total**	**Reconocimiento %**
365 / 381	8	234/330	70.91%	7/8	87.50%
256 / 281	7	166/227	73.13%	6/7	85.71%
261/275.9	**6.4 / 10**	**-----**	**79.13%**	**-----**	**87.70%**

3. **Detección de rostros, Reconocimiento de rostros, con construcción online de DB reconocimiento**

Esta prueba es para evaluar el reconocimiento de rostros con detección automática de rostros, pero a diferencia del experimento anterior la galería se hace online usando los rostros detectados por el sistema. En este caso se vuelven a utilizar 3 filtros de contexto: (1) filtro de contexto de la situación, (2) filtro de contexto físico espacial y (3) filtro de coherencia espacial. Además se utiliza el mapa de personas.

En la Tabla 9 se pueden observar los resultados del primer conjunto de parámetros (se muestran 2 ejemplos y el promedio de las 10 pruebas), se ve que el número de detección de rostros es en promedio de 192.7 en cada recorrido de la trayectoria. En la primera pasada se crea la base de datos agregando en promedio 15.6 personas. En este caso el promedio de las personas detectadas correctamente es 8.4 de 10, con 7.2 falsos positivos. La tasa de reconocimiento es de 70.20%, se puede apreciar que los resultados son peores que en la prueba anterior debido a que la base de datos fue creada con imágenes que pueden presentar rotaciones, por lo que el reconocimiento se hace más difícil.

Tabla 9: Resumen de resultados del experimento 3. Conjunto de parámetros 1.

Parámetros: Ruido en ángulo: $R_\theta \in [-5°, 5°]$, Ruido en desplazamiento: $R_P \in [-50\ cm, 50\ cm]$, Ruido Altura de personas: $R_H \in [-15\ cm, 15\ cm]$, Ruido en la trayectoria: $R_G \in [-3,3]$ y *Número de personas* =10

Parámetros: Ruido en ángulo = 5, Ruido en desplazamiento = 50 cm				
Número de rostros detectados / Total de rostros o *Ground Truth*	**Número de personas Almacenadas en DB**	**Número de personas correctamente detectadas**	**Reconocimiento Correctas/Total**	**Reconocimiento %**
169 / 188	17	8	5/8	62.50%
293 / 328	14	8	6/8	75.00%
192.7/215.7	**14.8 / 10**	**8.4**	**----**	**70.20%**

En la Tabla 10 se pueden apreciar los resultados del segundo conjunto de parámetros (se muestran 2 ejemplos y el promedio de las 10 pruebas), se ve que el número de detección de rostros es en promedio de 261 en cada recorrido de la trayectoria. En la primera pasada se crea la base de datos agregando en promedio 14.6 personas. En este caso el promedio de las personas detectadas correctamente es 6.4 de 10, con 8.2 falsos positivos. En esta prueba la tasa de reconocimiento es de 73.50%, se puede apreciar que los resultados son peores que en la prueba 2 con los mismos parámetros debido a que la base de datos fue creada con imágenes que pueden presentar rotaciones, por lo que el reconocimiento se hace más difícil. La tasa de reconocimiento es levemente mejor pero el porcentaje está calculado en función del número de personas correctamente detectadas que es menor.

Tabla 10: Resumen de resultados del experimento 3. Conjunto de parámetros 2.

Parámetros: Ruido en ángulo: $R_\theta \in [-10°, 10°]$, Ruido en desplazamiento: $R_P \in [-100\ cm, 100\ cm]$, Ruido Altura de personas: $R_H \in [-15\ cm, 15\ cm]$, Ruido en la trayectoria: $R_G \in [-3,3]$ y *Número de personas* =10

Parámetros: Ruido en ángulo = 10, Ruido en desplazamiento = 100 cm				
Número de rostros detectados / Total de rostros o *Ground Truth*	**Número de personas Almacenadas en DB**	**Número de personas correctamente detectadas**	**Reconocimiento Correctas/Total**	**Reconocimiento %**
365 / 381	15	8	6/8	75.00%
259 / 267	13	6	4/6	66.67%
261/275.9	**14.8 / 10**	**6.4**	**----**	**73.50%**

4. Detección de rostros, Reconocimiento de rostros y Visión activa, con construcción offline de DB reconocimiento

Esta prueba es para evaluar el reconocimiento de rostros con detección automática de rostros, la base de datos se hace de antemano usando rostros sin rotaciones en ningún eje. En este caso se utilizan 3 filtros de contexto: (1) filtro de contexto de la situación, (2) filtro de contexto físico espacial y (3) filtro de coherencia espacial. Además se utiliza visión actica para modificar las observaciones.

En la Tabla 11 se pueden observar los resultados del primer conjunto de parámetros (se muestran 2 ejemplos y el promedio de las 10 pruebas). Se ve que el número de detección de rostros es en promedio de 192.7 en cada recorrido de la trayectoria. En este caso el promedio de las personas detectadas correctamente es 8.4 de 10, con 7.2 falsos positivos. Se presenta la tasa de reconocimiento obtenida usando el mapa de personas, los filtros de contexto y además se usa visión activa, con esto la tasa de reconocimiento sube de un 86.77% (ver Tabla 7) a un 92.92% lo que muestra que el uso visión activa puede mejorar el desempeño del sistema de visión.

Tabla 11: Resumen de resultados del experimento 4. Conjunto de parámetros 1.

Parámetros: Ruido en ángulo: $R_\theta \in [-5°, 5°]$, Ruido en desplazamiento: $R_P \in [-50\ cm, 50\ cm]$, Ruido Altura de personas: $R_H \in [-15\ cm, 15\ cm]$, Ruido en la trayectoria: $R_G \in [-3,3]$ y *Número de personas* =10

Parámetros: Ruido en ángulo = 5, Ruido en desplazamiento = 50 cm			
Número de rostros detectados / Total de rostros o *Ground Truth*	**Número de personas correctamente detectadas**	**Reconocimiento Correctas/Total**	**Reconocimiento %**
169 / 188	8	7/8	87.50%
293 / 328	8	7/8	87.50%
192.7/215.7	**8.4 / 10**	**----**	**92.92%**

En la Tabla 12 se pueden ver los resultados del segundo conjunto de parámetros (se muestran 2 ejemplos y el promedio de las 10 pruebas). Se ve que el número de detección de rostros es en promedio de 261 en cada recorrido de la trayectoria. En este caso el promedio de las personas detectadas correctamente es 6.4 de 10. Nuevamente se presenta la tasa de reconocimiento de rostros obtenida usando el mapa de personas, los filtros de contexto y además se usa visión activa, con esto la tasa de reconocimiento sube de un 87.70% (ver Tabla 8) a un 92.80% lo que muestra que el uso visión activa puede mejorar el desempeño del sistema de visión. Los resultados de esta prueba son coherentes dada la mayor dificultad del segundo conjunto de parámetros.

Tabla 12: Resumen de resultados del experimento 4. Conjunto de parámetros 2.

Parámetros: Ruido en ángulo: $R_\theta \in [-10°, 10°]$, Ruido en desplazamiento: $R_P \in [-100\ cm, 100\ cm]$, Ruido Altura de personas: $R_H \in [-15\ cm, 15\ cm]$, Ruido en la trayectoria: $R_G \in [-3,3]$ y *Número de personas* =10

Parámetros: Ruido en ángulo = 10, Ruido en desplazamiento = 100 cm			
Número de rostros detectados / Total de rostros o *Ground Truth*	**Número de personas correctamente detectadas**	**Reconocimiento Correctas/Total**	**Reconocimiento %**
294 / 302	7	7/7	100.00%
259 / 267	6	5/6	83.33%
261/275.9	**6.4 / 10**	**----**	**92.80%**

5. Detección de rostros, Reconocimiento de rostros y Visión activa, con construcción online de DB reconocimiento

En esta prueba se evalúa el reconocimiento de rostros con detección automática de rostros, pero a diferencia del experimento anterior la base de datos se hace online usando los rostros detectados por el sistema. En esta prueba se vuelven a utilizar 3 filtros de contexto: (1) filtro de contexto de la situación, (2) filtro de contexto físico espacial y (3) filtro de coherencia espacial. Además se utiliza el mapa de personas y nuevamente visión activa para modificar las observaciones.

En la Tabla 13 se pueden observar los resultados del primer conjunto de parámetros (se muestran 2 ejemplos y el promedio de las 10 pruebas), se ve que el número de detección de rostros es en promedio de 192.7 en cada recorrido de la trayectoria. En la primera pasada se crea la galería o base de datos agregando en promedio 12.2 personas. Luego en la segunda pasada se detectan en promedio 15.6 personas. En este caso el promedio de las personas detectadas correctamente es 8.4 de 10. La tasa de reconocimiento es de 86.90%, se puede apreciar que los resultados son peores que en la prueba anterior debido a que la base de datos fue creada con imágenes que pueden presentar rotaciones, por lo que el reconocimiento se hace mas difícil. Pero es mejor que los resultados que se obtienen sin utilizar visión activa (tasa de reconocimiento de 70.20%, ver Tabla 9). Nuevamente el uso de visión activa mejora el reconocimiento aportando positivamente al sistema de visión.

Tabla 13: Resumen de resultados del experimento 5. Conjunto de parámetros 1.

Parámetros: Ruido en ángulo: $R_\theta \in [-5°, 5°]$, Ruido en desplazamiento: $R_P \in [-50\ cm, 50\ cm]$, Ruido Altura de personas: $R_H \in [-15\ cm, 15\ cm]$, Ruido en la trayectoria: $R_G \in [-3,3]$ y *Número de personas* =10

Parámetros: Ruido en ángulo = 5, Ruido en desplazamiento = 50 cm				
Número de rostros detectados / Total de rostros o *Ground Truth*	**Número de personas Almacenadas en DB**	**Número de personas correctamente detectadas**	**Reconocimiento Correctas/Total**	**Reconocimiento %**
168 / 192	12	9	8/9	88.89%
147 / 165	13	9	7/9	77.78%
192.7/215.7	**12.2 / 10**	**8.4**	**----**	**86.90%**

En la Tabla 14 se pueden ver los resultados del segundo conjunto de parámetros (se muestran 2 ejemplos y el promedio de las 10 pruebas), se ve que el número de detección de rostros es en promedio de 261 en cada recorrido de la trayectoria. En la primera pasada se crea la galería agregando en promedio 12.2 personas. En la segunda iteración se detectan en promedio 15.6 personas. En este caso el promedio de las personas detectadas correctamente es 6.4 de 10. La tasa de reconocimiento es de 89.46%, se puede apreciar que los resultados son mejores que la prueba en la que no se utiliza visión activa (sube de 73.50% a 89.46%. Ver Tabla 10)

Tabla 14: Resumen de resultados del experimento 5. Conjunto de parámetros 2.

Parámetros: Ruido en ángulo: $R_\theta \in [-10°, 10°]$, Ruido en desplazamiento: $R_P \in [-100\ cm, 100\ cm]$, Ruido Altura de personas: $R_H \in [-15\ cm, 15\ cm]$, Ruido en la trayectoria: $R_G \in [-3,3]$ y *Número de personas* =10

Parámetros: Ruido en ángulo = 10, Ruido en desplazamiento = 100 cm				
Número de rostros detectados / Total de rostros o *Ground Truth*	**Número de personas Almacenadas en DB**	**Número de personas correctamente detectadas**	**Reconocimiento Correctas/Total**	**Reconocimiento %**
365 / 381	13	8	7/8	87.50%
256 / 281	12	7	6/7	85.71%
261/275.9	**12.8 / 10**	**6.4**	**----**	**89.46%**

6. Reconocimiento de rostros y Visión activa, con construcción offline de DB reconocimiento

Para evaluar el reconocimiento de rostros con detección perfecta de rostros se utiliza la información con se generó la imagen para saber la posición del rostro (nótese que sólo se utiliza la información de donde se encuentra el rostro en la imagen y no la posición de la persona en el mapa), la galería se hace de antemano usando rostros sin

rotaciones en ningún eje. En este caso se utilizan 3 filtros de contexto: (1) filtro de contexto de la situación, (2) filtro de contexto físico espacial y (3) filtro de coherencia espacial. Además se utiliza visión actica para modificar las observaciones.

En la Tabla 15 se pueden apreciar los resultados del primer conjunto de parámetros (se muestran 2 ejemplos y el promedio de las 10 pruebas). En este caso el promedio de las personas detectadas correctamente es 10, sin falsos positivos. Esto es debido a que se tiene la posición de todos los rostros que son generadas y al usar el mapa de personas éste no tiene problemas en localizar solo una vez a cada persona en el mapa. La tasa de reconocimiento es de un 92%.

Tabla 15: Resumen de resultados del experimento 6. Conjunto de parámetros 1.

Parámetros: Ruido en ángulo: $R_\theta \in [-5°, 5°]$, Ruido en desplazamiento: $R_P \in [-50\ cm, 50\ cm]$, Ruido Altura de personas: $R_H \in [-15\ cm, 15\ cm]$, Ruido en la trayectoria: $R_G \in [-3,3]$ y *Número de personas* =10

Parámetros: Ruido en ángulo = 5, Ruido en desplazamiento = 50 cm		
Número de personas correctamente detectadas	**Reconocimiento Correctas/Total**	**Reconocimiento %**
10	8/10	80.00%
10	9/10	90.00%
10.0	**9.2 / 10**	**92.00%**

En la Tabla 16 se pueden ver los resultados del segundo conjunto de parámetros (se muestran 2 ejemplos y el promedio de las 10 pruebas). En este caso el promedio de las personas detectadas correctamente es nuevamente 10, sin falsos positivos. La tasa de reconocimiento es de un 89%, levemente menor que la anterior debido a la mayor dificultad.

Tabla 16: Resumen de resultados del experimento 6. Conjunto de parámetros 2.

Parámetros: Ruido en ángulo: $R_\theta \in [-10°, 10°]$, Ruido en desplazamiento: $R_P \in [-100\ cm, 100\ cm]$, Ruido Altura de personas: $R_H \in [-15\ cm, 15\ cm]$, Ruido en la trayectoria: $R_G \in [-3,3]$ y *Número de personas* =10

Parámetros: Ruido en ángulo = 10, Ruido en desplazamiento = 100 cm		
Número de personas correctamente detectadas	**Reconocimiento Correctas/Total**	**Reconocimiento %**
10	9/10	90.00%
10	8/10	80.00%
10.0	**8.9 / 10**	**89.00%**

7. **Reconocimiento de rostros y Visión activa, con construcción online de DB reconocimiento**

Esta prueba es para evaluar el reconocimiento de rostros con detección perfecta de rostros, para esto se utiliza la información con se generó la imagen para saber la posición del rostro (nótese que solo se utiliza la información de donde se encuentra el rostro en la imagen y no la posición de la persona en el mapa), pero a diferencia del experimento anterior la galería se hace online usando la información de los rostros generado por el sistema. Aun cuando las posiciones de los rostros son perfectas pueden presentarse rotaciones cuando se hace la galería. En este caso se utilizan 3 filtros de contexto: (1) filtro de contexto de la situación, (2) filtro de contexto físico espacial y (3) filtro de coherencia espacial. Además se utiliza visión actica para modificar las observaciones.

En la Tabla 17 se pueden observar los resultados del primer conjunto de parámetros (se muestran 2 ejemplos y el promedio de las 10 pruebas). Las personas detectadas en cada iteración son 10 sin falsos positivos. La tasa de reconocimiento es de un 90% levemente menor posiblemente debido a que la galería se hizo online y los rostros pueden presentar rotaciones.

Tabla 17: Resumen de resultados del experimento 7. Conjunto de parámetros 1.

Parámetros: Ruido en ángulo: $R_\theta \in [-5°, 5°]$, Ruido en desplazamiento: $R_P \in [-50\ cm, 50\ cm]$, Ruido Altura de personas: $R_H \in [-15\ cm, 15\ cm]$, Ruido en la trayectoria: $R_G \in [-3,3]$ y *Número de personas* =10

Parámetros: Ruido en ángulo = 5, Ruido en desplazamiento = 50 cm			
Número de personas Almacenadas en DB	**Número de personas correctamente detectadas**	**Reconocimiento Correctas/Total**	**Reconocimiento %**
10	10	10/10	100.00%
10	10	9/10	90.00%
10.0	**10.0 / 10**	**----**	**90.00%**

En la Tabla 18 se pueden ver los resultados del segundo conjunto de parámetros (se muestran 2 ejemplos y el promedio de las 10 pruebas). En este caso el promedio de las personas detectadas correctamente es nuevamente 10, sin falsos positivos. La tasa de reconocimiento es de un 86%, levemente menor que la anterior debido a la mayor dificultad. El uso del mapa de personas permite que se detecten todas las personas del mapa y que no existan falsos positivos.

Tabla 18: Resumen de resultados del experimento 7. Conjunto de parámetros 2.

Parámetros: Ruido en ángulo: $R_\theta \in [-10°, 10°]$, Ruido en desplazamiento: $R_P \in [-100\ cm, 100\ cm]$, Ruido Altura de personas: $R_H \in [-15\ cm, 15\ cm]$, Ruido en la trayectoria: $R_G \in [-3,3]$ y *Número de personas* =10

Parámetros: Ruido en ángulo = 10, Ruido en desplazamiento = 100 cm			
Número de personas Almacenadas en DB	**Número de personas correctamente detectadas**	**Reconocimiento Correctas/Total**	**Reconocimiento %**
10	10	8/10	80.00%
10	10	8/10	80.00%
10.0	**10.0 / 10**	**8.6/10**	**86.00%**

8. Detección de Rostros, reconocimiento de rostros. Con y sin visión activa, con construcción offline de DB reconocimiento

Esta prueba es para evaluar el reconocimiento y como se beneficia el sistema de visión al usar solo el módulo de visión activa. Se realizan 20 iteraciones con cada conjunto de parámetros. En esta prueba el agente se localiza frente al sujeto usando la trayectoria random. El agente posee una galería con los rostros de la arena, para crearla se agrega 1 imagen de cada sujeto donde la imagen utilizada no posee rotaciones en ningún eje. En este caso se utiliza visión activa solo si se detecta una persona al principio, o sea se utiliza para mejorar el reconocimiento y no para ayudar a detectar a la persona.

En la Tabla 19 se pueden apreciar los resultados del primer conjunto de parámetros: Distancia Base al sujeto = 200 *cm*., Ruido en desplazamiento: $R_P \in [-50\ cm, 50\ cm]$, Ruido Altura de personas: $R_H \in [-15,15]$ y *Número de personas* = 20 (Se muestran 3 ejemplos y el promedio de las 20 pruebas). Se puede ver que en promedio se detectan 18.35 personas de un total de 20 en cada iteración. Al usar visión activa el reconocimiento mejora de un 90.47% a un 95.66%, lo que vuelve a mostrar que el uso de visión activa mejora considerablemente el desempeño del sistema de visión.

Tabla 19: Resumen de resultados del experimento 8. Conjunto de parámetros 1.

Parámetros: Distancia Base del sujeto = 200 *cm*., Ruido en desplazamiento: $R_P \in [-50\ cm, 50\ cm]$, Ruido Altura de personas: $R_G \in [-15\ cm, 15\ cm]$ y *Número de personas* = 20

Sin Visión Activa			**Con Visión Activa**		
Número de personas correctamente detectadas	**Reconocimiento Correctas/Total detectadas**	**Reconocimiento %**	**Número de personas correctamente detectadas**	**Reconocimiento Correctas/Total detectadas**	**Reconocimiento %**
19	17/19	89.47%	19	18/19	94.74%
18	16/18	88.89%	18	17/18	94.44%
17	15/17	88.24%	17	16/17	94.12%
18.35 / 20	**----**	**90.47%**	**18.35 / 20**	**----**	**95.66%**

En la Tabla 20 se pueden observar los resultados del segundo conjunto de parámetros: Distancia Base al sujeto = 250 *cm*., Ruido en desplazamiento: $R_P \in [-100\ cm, 100\ cm]$, Ruido Altura de personas: $R_H \in [-15\ cm, 15\ cm]$ y *Número de personas* = 20. (Se muestran 3 ejemplos y el promedio de las 20 pruebas). Se puede ver que en promedio se detectan 17.15 personas en cada iteración que es un poco inferior a la prueba anterior pero es debido a la mayor dificultad de los parámetros. Al usar visión activa el reconocimiento mejora en este caso de un 89.82% a un 95.89%, lo que vuelve a mostrar que el uso de visión activa mejora considerablemente el desempeño del sistema de visión.

Tabla 20: Resumen de resultados del experimento 8. Conjunto de parámetros 2.

Parámetros: Distancia Base del sujeto = 200 *cm*., Ruido en desplazamiento: $R_P \in [-100\ cm, 100\ cm]$, Ruido Altura de personas: $R_H \in [-15\ cm, 15\ cm]$, y *Número de personas* = 20

Sin Visión Activa			**Con Visión Activa**		
Número de personas correctamente detectadas	**Reconocimiento Correctas/Total detectadas**	**Reconocimiento %**	**Número de personas correctamente detectadas**	**Reconocimiento Correctas/Total detectadas**	**Reconocimiento %**
17	15/17	88.24%	17	15/17	88.24%
19	18/19	94.74%	19	18/19	94.74%
19	17/19	89.47%	19	19/19	100.00%
17.15 / 20	**----**	**89.82%**	**17.15 / 20**	**----**	**95.89%**

5.7 Conclusiones

Del estudio comparativo se puede concluir que el mejor método de detección de rostros utiliza clasificadores de tipo *boosted* en cascada, y mejora el rendimiento mediante el uso de un enfoque de *coarse-to-fine* en el entrenamiento de las cascadas. Este método tiene una tasa baja de falsos positivos lo que lo hace muy útil en aplicaciones con robots de servicio. Los métodos presentan robustez ante rotaciones de los rostros *yaw* y *pitch*.

Con respecto a los métodos de reconocimiento, el estudio comparativo muestra que los métodos basados en características LBP son más robustos en *indoor* ante rotaciones de los rostros que los métodos Gabor y WLD. En *outdoor* el método de Gabor-jets tiene el mejor rendimiento. Además todos los métodos implementados son robustos ante rotaciones menores a 40°. El rendimiento de todos los métodos de reconocimiento disminuye notablemente cuando se usan imágenes *outdoor*.

De los experimentos de evaluación de módulos del sistema se puede que los módulos de contexto realizan un aporte significativo en el rendimiento del sistema de visión. En la Tabla 21 se puede apreciar un resumen de los resultados obtenidos. El experimento 1 muestra que el módulo de mapa de persona ayuda a evitar que se agregue a la misma persona más de una vez en la galería de forma eficiente. Además los diferentes filtros de contexto logran eliminar la mayoría de los falsos positivos, permitiendo una detección más precisa de la posición de las personas dentro del mapa. Se puede apreciar que el número de rostros detectados en cada trayectoria es alto (en promedio 192.7), y que solo se agregan al mapa de personas un número reducido, esto se debe a que el filtro de coherencia espacial funciona descartando detecciones cuando la persona es la misma, y ya ha sido agregada al mapa de personas. Además el número de falsas detecciones de personas descartadas usando los filtros de contexto son en promedio 27.9 para el primer conjunto de parámetros y en promedio 22.2 para el segundo conjunto de parámetros, estas personas son descartadas porque no cumplen con las leyes físicas o las alturas determinadas por el sistema están fuera de los parámetros correctos. El experimento 2 muestra que el módulo de reconocimiento funciona como se esperaba, siendo coherente con los resultados de los estudios comparativos. La tasa de reconocimiento mejora cuando se usan los diferentes módulos de contexto. Se tiene un 78.41% de reconocimiento sin el uso de los filtros de contexto, y un 86.77% de reconocimiento con el uso de los módulos de: Mapa de personas, filtro de contexto espacial, filtro de coherencia espacial y el filtro de contexto de situación. De los experimentos 3 y 5 se puede concluir que el uso de visión activa permite, entre otras cosas, que se construya una mejor galería (en caso que la galería se construya *online*), y con esto la tasa de reconocimiento es mejor. En el caso de los experimentos 2 y 4 en que la galería se construye *offline*, el uso de visión activa permite que la tasa de reconocimiento mejore de 86.77% a un 92.92% en promedio. Los experimentos 6 y 7 tienen la intención de analizar solo el reconocimiento, los diferentes módulos de contexto y la visión activa sin que dependan del módulo de detección de rostros. Si la posición de los rostros esta dada por el simulador el mapa de personas funciona de forma perfecta, estimando la posición de las personas y agregando a 10 de 10 personas. La tasa de reconocimiento en el caso de la construcción de la galería *online* es levemente menor que en el caso de la construcción *offline*. En el experimento 8 se utiliza una trayectoria más simple sólo para evaluar como se beneficia el sistema con el uso del módulo de visión activa, en este caso nuevamente se confirma que el uso de visión activa permite mejorar el funcionamiento del módulo de reconocimiento, la tasa de reconocimiento aumenta de un 90.4% a un 95.66% en promedio.

Tabla 21: Resumen de evaluaciones realizadas.
En (*) no se usa Mapa de Personas.

Experimento	**Detector de Caras**	**Visión Activa**	**Galería**	**Personas agregadas a la galería**	**Personas correctamente detectadas[%]**	**Reconocimiento [%] (de todos los sujetos de la escena)**	**Reconocimiento [%] (de los sujetos detectados)**
2	Adaboost	No	Offline	10.0	84.0%	----	78.4% (*)
2	Adaboost	No	Offline	10.0	84.0%	73.0%	86.8%
3	Adaboost	No	Online	14.8	84.0%	59.0%	70.2%
4	Adaboost	Yes	Offline	10.0	84.0%	78.0%	92.9%
5	Adaboost	Yes	Online	12.2	84.0%	73.0%	86.9%
6	Ground Truth	Yes	Offline	10.0	100.0%	92.0%	92.0%
7	Ground Truth	Yes	Online	10.0	100.0%	90.0%	90.0%
8	Adaboost	No	Offline	20.0	91.8%	83.0%	90.5%
8	Adaboost	Yes	Offline	20.0	91.8%	87.5%	95.7%

Capítulo 6
Reconocimiento de Humanos en Ambientes Domésticos usando Información Visual y Térmica[7]

En este capítulo se describe un sistema robusto para la detección y la identificación de los seres humanos en entornos domésticos para ser usado en un robot de servicio. La principal función de este sistema es evaluar el funcionamiento del sistema de visión propuesto en el Capítulo 4 en una aplicación real, evaluando el funcionamiento de los diferentes módulos de contexto implementados y que fueron validados en el Capítulo 5. Además se implementan nuevos módulos de contexto específicos para esta aplicación. Se puede observar un diagrama general del sistema en la Figura 32.

La detección robusta de personas se logra mediante el uso de fuentes de información térmica y visual que se integran para detectar objetos que son *candidatos a humanos*. Estos candidatos son procesados con el fin de verificar la presencia de los seres humanos y su identidad con la información frente a los espectros térmico (*ET*) y visible (*EV*). La detección de rostros se utiliza para verificar la presencia de los seres humanos, y el reconocimiento de rostros para identificarlos. Se emplean visión activa para modificar la postura relativa del robot con respecto a un candidato a persona cuando la identificación no es directamente posible, por ejemplo, si el objeto está demasiado lejos y el robot debe acercarse a él, o el ángulo de visión no es apropiado para la identificación de los humanos por lo que el robot debe encontrar un ángulo de visión mejor. El sistema se utiliza diferentes instancias de contexto para eliminar detecciones falsas y mejorar el desempeño global del sistema de reconocimiento.

En condiciones de iluminación mala o variable, el sistema se basa principalmente en el uso de la información térmica. Sin embargo, en condiciones de buena iluminación, la

[7] Este capítulo esta basado en el artículo [18]. Además de los resultados presentados en [18] se presenta más detalle de los módulos de contexto utilizados.

información *ET* y *EV* se complementan entre sí. Por ejemplo, la información visual permite un mejor análisis de las texturas y una detección más robusta de los ojos (que se utiliza para la alineación del rostro antes de la identificación). La información térmica permite una fácil diferenciación de los cuerpos y rostros humanos en fondos complejos.

Es importante mencionar que en el sistema implementado, se utilizan los métodos más avanzados para la detección y el reconocimiento de rostros en los espectros visibles y térmicos. En el caso de detección de rostros en el espectro térmico, se utilizan por primera vez para resolver este problema los clasificadores del tipo "*boosted cascade*".

En los diferentes módulos implementados se utilizan los siguientes tipos de contexto: Contexto de bajo nivel en el análisis de los diferentes blobs; Contexto físico espacial en el módulo de Integración y análisis de blobs; contexto de coherencia espacial en los módulos de detección de cuerpo humano y el módulo de Integración y análisis de blobs; Contexto de la situación con el uso del mapa de personas e información térmica en la mayoría de los módulos.

6.1 Trabajo relacionado

Las actividades de investigación en robótica relacionadas con los robots de servicio doméstico han aumentado considerablemente en los últimos años. Algunos de los principales impulsores de este fenómeno son: el uso de los robots domésticos para mejorar la calidad de vida de las personas mayores, las aplicaciones de cuidado de niños, entretenimiento y educación, y la prestación de servicios específicos, tales como limpieza. Además de las iniciativas interesantes como la RoboCup@Home [96], cuyo objetivo es proporcionar pruebas de referencia y metodologías para evaluar las habilidades y el desempeño de los robots de servicio doméstico en la realidad, usando configuración de entorno del hogar no estándar. Con esto se espera que se acelere y enfoque el progreso tecnológico y científico en el campo de los robots de servicio doméstico [121].

La detección robusta y la identificación de los seres humanos por robots en entornos domésticos es un problema abierto. Por ejemplo, en la RoboCup@Home, incluso los mejores equipos no son capaces de lograr la detección robusta humanos y la identificación de los mismos en las competencias diseñadas para probar este tipo de habilidades (por ejemplo, "Who's Who?" [96]).

Existen varios enfoques diferentes de reconocimiento facial que se han desarrollado en los últimos años [117][125][89][31], que van desde los métodos que usan *Eigenspace* (por ejemplo, *Eigenfaces* [68]), hasta sofisticados sistemas basados en imágenes de alta resolución y modelos en 3D. Varios métodos han sido desarrollados para el reconocimiento de rostros con imágenes térmicas, y la mayoría de estos métodos se basan en el mismo tipo de métodos utilizados en las imágenes visibles [19][41][43][101][58][123][119][63][62]. En [43] hay una comparación de los métodos de reconocimiento de rostros en imágenes térmicas (infrarrojos de onda larga, de 8-12 μm). El estudio considera los anteriormente mencionados requisitos HRI de operación en línea y tiempo real, una imagen por persona y ambientes sin restricciones, y se centra en los tres métodos que obtuvieron los mejores resultados en el espectro visible [101]: patrones binarios locales (LBP) histogramas, los descriptores *Gabor Jet*, y los descriptores basados en la transformada invariante a escala (SIFT). En términos generales, los resultados presentados en [43] indican que los métodos basados en LBP son capaces de obtener altas tasas de reconocimiento y presentar requerimientos computacionales y de memoria que son adecuados para el uso del HRI. Por esta razón histogramas LBP se

utilizan para implementar el espectro visible y térmico los módulos de reconocimiento facial utilizados en el sistema propuesto.

Por lo tanto, uno de los principales aportes de este capítulo es la propuesta de un sistema robusto para la detección y la identificación de seres humanos en ambientes domésticos, usando métodos de detección de rostros y los métodos de reconocimiento que trabajan en espectros visibles y térmicos. Además del uso de contexto para mejorar el funcionamiento general del sistema.

6.2 Sistema de detección e identificación de personas

6.2.1 Descripción general del sistema

El diseño del sistema propuesto para la detección y la identificación de seres humanos tiene en cuenta las ventajas de uso de la información térmica y visual, y considera que ambas cámaras tienen un FOV (*field-of-view*) y profundidad de campo parecidos. El sistema propuesto cumple los cuatro requisitos mencionados en la Sección 2.1.1. La Figura 32 presenta el esquema del sistema propuesto. Los módulos principales se pueden agrupar en las siguientes categorías: Detección de Piel, Detección del Cuerpo Humano, Detección de Personas (módulo de Integración de Blobs y el Análisis de Detecciones), Detección de Rostros, Reconocimiento de Rostros, y Toma de Decisiones. Cada uno de estos módulos trabaja con imágenes visuales (I_V en la Figura 32), con imágenes térmicas (I_T en la Figura 32), o con información extraída de ambas fuentes.

Primero, la detección de piel y cuerpo humano se utilizan para detectar grupos de *blobs* de piel en el espectro visible y térmico (*SetBlobsPielTérmica* y *SetBlobsPielVisible*), y un conjunto de *blobs* de cuerpo en espectro térmico (*SetBlobsCuerpo*). Luego, en el módulo de *integración y análisis de Blobs*, la información contenida en estos conjuntos se integra y se analiza usando el módulo de detección de rostros. El módulo de *Integración y Análisis de Blobs* genera candidatos a persona, rostro y rostro frontal que se analizan en el módulo de *toma de decisiones*. Este último módulo lleva a cabo la tarea fundamental de orientar la búsqueda de los seres humanos, así como mejorar los puntos de vista con el fin de detectar y reconocer los seres humanos con gran precisión, mientras que al mismo tiempo reducir al mínimo los movimientos del robot. Entre otras tareas el módulo de *Toma de Decisiones* genera las órdenes de movimiento al cuerpo del robot y la cabeza, y controla la interacción con los seres humanos, además es el encargado de interactuar con el módulo de reconocimiento de rostros y navegar por el medio ambiente. La Tabla 22 enumera los sub-módulos y los métodos utilizados en cada módulo. Los módulos se detallan en las siguientes secciones.

La Figura 33 presenta un ejemplo de una imagen visible y térmica, así como la salida de algunos módulos. Es importante señalar que en todos los módulos se trabajan las imágenes en escala de grises (ya sea térmica o visual), con la excepción del módulo detección de piel visual que funciona en imágenes RGB.

Tabla 22: Lista de módulos y métodos.

Nombre del módulo	Sub módulo	Salida	Método
Detección de piel	*Detección de piel EV*	SetBlobsPielVisible	Dynamically updated Skindiff algorithm [98]
	Detección de piel ET	SetBlobsPielTérmica	Mixture of Gaussians (MoG)
Detección de Cuerpo Humano		SetBlobsCuerpo	MoG + Heuristics
Integración y análisis de Blobs	-	Candidatos a Persona, Candidatos a Rostros y rostros frontales	Heurísticas
Toma de decisiones	Visión activa, navegación y Comportamientos-Conductas	Posición de personas y rostros	Heurísticas
Detección de Rostros	-	Rostros detectados (Visible y térmica)	Nested Cascades of Boosted Classifiers [115]
Reconocimiento de Rostros	-	ID rostros	Histograms of LBP features [9]

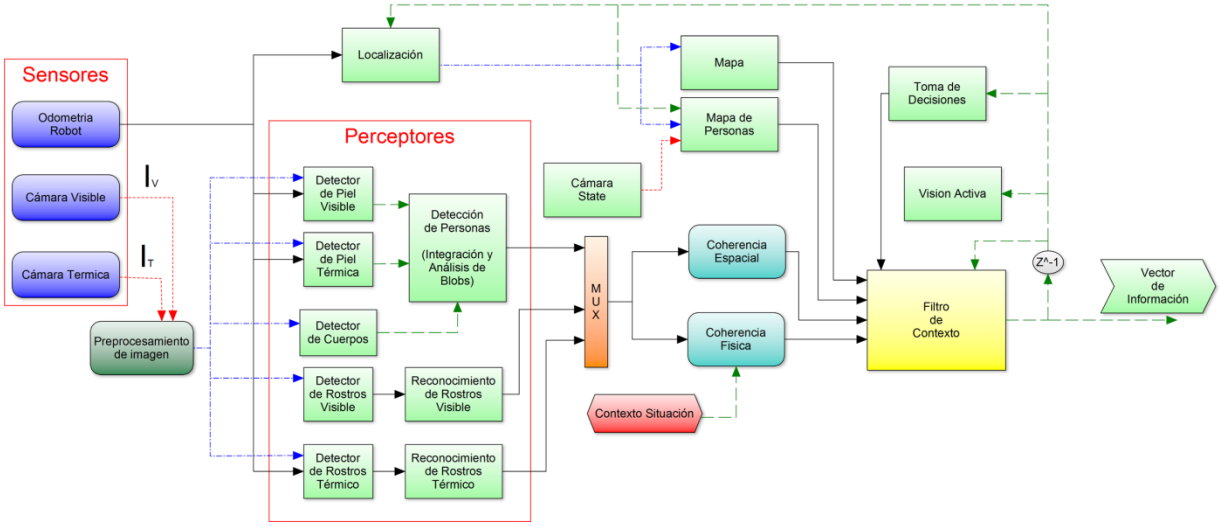
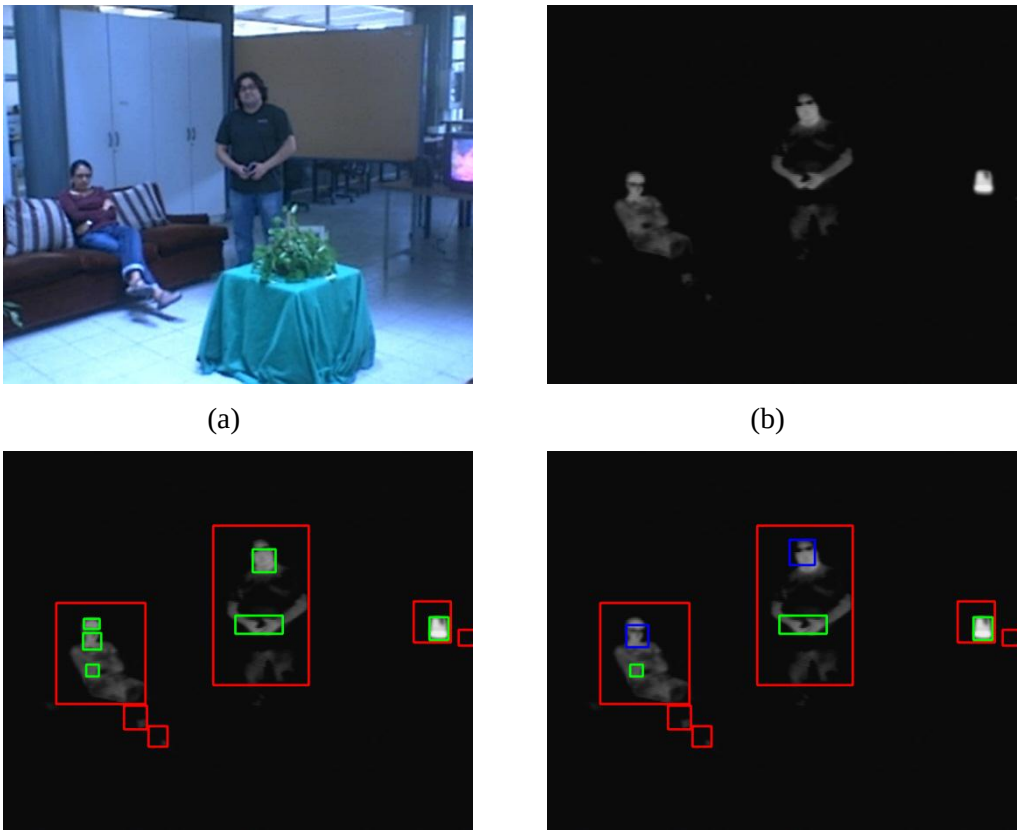


Figura 32. Diagrama de bloques de sistema de detección e identificación de personas. Ver el texto para más información.



(a) (b)

(c) (d)

Figura 33. Salida de los módulos seleccionados.: (a) Imagen Visible. (b) Imagen Térmica. (c) Detección de piel: en rojo el *blobs* de cuerpo humano y en verde los *blobs* de piel térmica. (d) Detección de personas: El rojo los candidatos a personas, en verde los candidatos a rostros, y en azul los rostros frontales. Se pueden observar algunas detecciones falsas.

6.2.2 Perceptores

6.2.2.1 Detección de Piel

El módulo de *Detección de Piel Visible* determina las regiones de las imágenes que contienen piel (en el espectro visible) utilizando el algoritmo de segmentación de piel *Skindiff* [98]. *Skindiff* es un algoritmo de detección rápida de piel que utiliza la información de la vecindad (contexto espacial local y de bajo nivel) para lograr robustez. Tiene dos fases de procesamiento principal, clasificación *pixel-wise* y difusión espacial. La clasificación de píxeles inteligente utiliza un modelo de piel no-paramétrico [53], G_t, implementado usando histogramas, y la difusión espacial toma la información de la vecindad a la hora de clasificar un pixel, y comienza a partir de los píxeles que tienen una probabilidad grande de ser píxeles de piel [98]. El modelo de probabilidad de piel se puede adaptar continuamente con la información de los píxeles de la zona del rostro (píxeles de piel) detectados en las imágenes anteriores:

$$G_t = G_{t-1}\alpha + \hat{G}_{tface(t)}(1-\alpha) \qquad (6.1)$$

donde $\hat{G}_{tface(t)}$ es estimado usando el rostro detectado actual, y $G_o$es el modelo inicial, el cual, puede ser inicializado desde un modelo previamente almacenado, y α una constante. Si el rostro no se detectó en el cuadro anterior, α se establece en uno, de lo contrario

$0 < \alpha < 1$. Aunque nuestra experiencia muestra que la actualización del modelo de piel con los rostros detectados mejora en gran medida los resultados [100], en el presente trabajo el sistema funciona en imágenes individuales y el modelo no se actualiza. El conjunto de *blobs* de piel obtenido de en espectro visible se llama *SetBlobPielVisible*.

El módulo *Detección de Piel Térmica* se basa en un modelo de probabilidad paramétrica de la distribución de la temperatura de piel. Modelos del tipo *Mezcla de gaussianas* (*Mixture of Gaussian* o MoG) para las distribuciones de los modelos de piel ($P[x_q \mid skin]$) y no piel ($P[x_q \mid non-skin]$). Un clasificador Bayes determina cuales píxeles son de piel si cumplen la siguiente relación:

$$r(x_q) = \frac{P[x_q|skin]}{P[x_q|non-skin]} > u_q \qquad (6.2)$$

con x_q la temperatura observada del píxel x, y un umbral u_q previamente fijado. Entonces, una operación de difusión se aplica al grupo de los píxeles de piel térmica en los *blobs* piel térmicas. El conjunto de blobs de piel térmica se llama *SetBlobPielTérmica*. Todos los parámetros del modelo de probabilidad de piel se obtuvieron utilizando una base de datos de entrenamiento.

6.2.2.2 Detección de cuerpos humanos

El módulo de detección de *blobs* Humanos está a cargo de la detección de cuerpos humanos y de partes del cuerpo humano utilizando la información térmica. Este módulo utiliza contexto físico espacial y de configuración de objetos. La detección también incluye partes del cuerpo cubiertas por la ropa. La misma razón de probabilidad $r(x_q)$ que se utiliza para la detección de piel térmica, pero el umbral se ha adaptado para tener en cuenta los cambios en la temperatura del medio ambiente y los cambios en la respuesta de la cámara. Esto se hace mediante la aplicación de un mapeo lineal de los valores de $r(x_q)$ -- tomando los valores máximos y mínimos observados y mapeándolos a 0 y 1 -- y mediante el uso de un umbral de decisión fijo en este rango. Esto permite la adaptación a las situaciones donde hay una gran diferencia en la temperatura de los cuerpos (ropa) y los rostros, la temperatura del cuerpo y la ropa puede variar mucho de verano a invierno. Después de que cada píxel se ha clasificado, una operación de apertura morfológica se aplica para rellenar los agujeros que aparecen en el cuerpo, aunque rara vez aparecen en las regiones del rostro. El conjunto de *blobs* (térmicos) de cuerpo humano se llama *SetBlobsCuerpo*.

Es importante aclarar que la detección térmica piel permite segmentar las regiones de piel que no están cubiertas por la ropa, como los brazos, rostro, manos y piernas, mientras que la detección (térmica) de cuerpo humano permite la detección de grandes partes del cuerpo, lo que podría en parte estar cubiertos por el pelo o la ropa.

6.2.2.3 Detector de rostros

La detección de rostros se basa en el uso de un método multi-escala de detección de objetos (ver el diagrama de bloques en la Figura 7) previamente desarrollado por grupo de

visión computacional de la Universidad de Chile[8] [115], que utiliza clasificadores de tipo *boosted* en cascada. El mismo trabajo se utiliza para construir detectores de rostros capaces de detectar los rostros en los espectros visible y térmico. Una descripción mas completa del método se incluye en □.

Para el entrenamiento y la validación de los detectores de rostros utilizados en este capítulo las siguientes bases de datos fueron utilizadas (El proceso de entrenamiento se describe en [115]):

- Espectro Visible. Entrenamiento: 5,000 imágenes de rostros frontales y 3,500 imágenes de no-rostros. Validación: 5,000 imágenes de rostros frontales y 1,500 imágenes de no-rostros. Las imágenes fueron obtenidas desde diferentes fuentes, y todas ellas fueron tomadas bajo condiciones reales, incluyendo variaciones de iluminación, fondo, raza, etc.
- Espectro Térmico: Entrenamiento: 20,000 imágenes de rostros frontales (generadas desde 800 imágenes de rostros) y 15,000 imágenes de no-rostros (generadas desde 200 imágenes de no-rostros). Validación: 20,000 imágenes de rostros frontales y 15,000 imágenes de no-rostros. Las imágenes térmicas fueron obtenidas utilizando una cámara similar a la utilizada en este trabajo.

6.2.2.4 Reconocedor de rostros

- Reconocimiento de rostros en imágenes de espectro visible.

En [101] se presenta un estudio comparativo del estado de arte en métodos de reconocimiento facial que son más adecuados para trabajar en ambientes sin restricciones utilizando información visible que son planteados en esta tesis. Los métodos se han seleccionado teniendo en cuenta su rendimiento en pruebas realizadas. Para detalles de los métodos ver sección 3.2. El detector de ojos utilizado se describe en [115].

- Reconocimiento de rostros en imágenes térmicas .

En [43] se presenta un estudio comparativo de los métodos de reconocimiento de rostros para aplicaciones con imágenes térmicas HRI. Los resultados obtenidos en este estudio también muestran que los histogramas LBP son robustos y eficientes en el reconocimiento de rostros en imágenes *ET*. Por lo tanto, el método de histogramas LBP fue seleccionado para ser implementado para el reconocimiento de rostros en imágenes térmicas en este sistema.

Como en el caso de las imágenes *EV*, las imágenes se dividen en 80 regiones, y se utiliza como medida de similitud la intersección de histograma. Sin embargo, los rostros no están alineados antes del reconocimiento, sobre todo porque la detección de ojos en imágenes térmicas es más inexacto que en las imágenes *EV* [58]. Las imágenes no alineadas se pueden utilizar porque el algoritmo de histogramas LBP pueden manejar bastante bien el problema de alineamiento [101], que fue un motivo adicional para seleccionar este método. En el caso del módulo de reconocimiento en imágenes *EV*, los rostros fueron alineados utilizando un detector de ojos porque se necesita una alineación

[8] Grupo de Visión Computacional. Departamento de Ingeniería Eléctrica. Universidad de Chile. http://vision.die.uchile.cl

más exacta con el fin de eliminar el fondo que podrían afectar considerablemente el proceso de reconocimiento en ambientes restringidos. El problema del fondo complejo no se presenta en las imágenes *ET*.

6.2.2.5 Integración y Análisis de Blobs (Detección de Personas)

Los candidatos a persona, los candidatos a rostro, y los rostros frontales se determinan mediante la integración de la información contenida en los conjuntos de cuerpo y *blobs* de piel, y utilizando el módulo de detección de rostros. Se utiliza el siguiente procedimiento:

(i) En primer lugar, todas los *blobs* de cuerpo son seleccionados como candidatos persona.

(ii) A continuación, se aplica la detección de rostros dentro de cada *blob* de cuerpo, en este caso se utiliza contexto de configuración de objetos ya que tiene que existir una coherencia espacial en la escena (es lógico realizar la búsqueda de rostros donde hay cuerpos). Todos los rostros frontales detectados que se encuentran dentro de un *blob* de cuerpo son marcados como rostro frontal (se utiliza el detector de rostros frontales descrito anteriormente), y el *blob* de cuerpo correspondiente se marca como que contiene un rostro. Los *blobs* de cuerpo que no contengan un rostro son marcados como candidatos a personas ya que puede existir un rostro que no este frontal a la cámara.

(iii) Por último, todas los *blobs* de piel que no tienen intersección con un rostro frontal detectada son marcados como candidatos a rostros.

6.2.2.6 Toma de decisiones y Visión Activa

El módulo de toma de decisiones está a cargo de la búsqueda activa de los seres humanos y los rostros, además de la interacción Humano-Robot. Dado un mapa *M* y un conjunto de posiciones del mapa $P_i, i = 1,...,N$ para ser visitado, el procedimiento seguido por el robot de búsqueda de candidatos a humanos dentro del mapa es la siguiente:

i. El robot se mueve a P_1.

ii. El robot busca activamente candidatos humano mirando en tres direcciones (0, 45 y -45 grados), moviendo su cabeza. Las imágenes *EV* y *ET* obtenidas son analizadas por el módulo de detección de piel, el módulo de detección de cuerpo humano, y el módulo de análisis e integración de *blobs*, y los dos conjuntos disjuntos se obtienen son: B (cuerpos) y *F* (rostros). *B* contiene los objetos detectados que fueron clasificados como candidatos a persona y no contienen rostros frontales detectados, y *F* contiene los rostros frontales detectados y los candidatos a rostros frontales (que se encuentra en cualquiera de las imágenes *EV* o *ET*). Se selecciona el objeto más grande del conjunto *B*. Este objeto se llama *MB*, y se corresponde con el candidato del cuerpo principal.

iii. Para cada elemento de *F* y de *MB*, se calcula la distancia desde el robot al objeto utilizando la información del rostro en el caso de los elementos pertenecientes al conjunto *F*, y el ancho del *blob* en el caso del objeto *MB*. El ancho de *MB* se considera que es una buena estimación del ancho del torso de una persona. La posición de cada persona/objeto representado por los elementos de *F* o *MB* es almacenada en un mapa de personas (ver más detalle del mapa de personas en sección 5.4.5) si esta posición se encuentra dentro de un rango aceptable dentro del mapa. Si el ancho de MB es pequeño, como para las lámparas, computadoras y otros objetos, su posición se

considera fuera del mapa, y se descarta. En este caso se usa contexto de la situación debido a que se conoce el ambiente en donde esta inmerso el robot y con ello se pueden descartar objetos que se encuentren fuera de este. Además se utiliza contexto físico espacial ya que las personas detectadas deben cumplir con las reglas físicas que las rigen, la persona no puede estar bajo el piso o flotando en el aire, usando esta información se eliminan falsos positivos. Dada la estimación de distancia del objeto al robot y dado el ángulo de la cabeza del robot se estima la proyección del objeto, si el objeto se encuentra a una altura bajo 10 cm y sobre 200 cm en el eje z se descarta.

Sea h_{F_i} la altura en el eje z del elemento F_i. Este objeto se acepta si cumple que:

$$h_{F_i} > 10cm \; y \; h_{F_i} < 200 \; cm \qquad (6.3)$$

iv. El robot trata de llegar a cada persona no reconocida en el mapa, empezando con la persona más cercana. Para ello, el robot se acerca a la persona y le habla, preguntando a la persona si puede mirar al rostro del robot. Si el robot no puede detectar un rostro utiliza visión activa para realizar una búsqueda local, moviendo la cabeza hacia cuatro direcciones diferentes (arriba-derecha, arriba a la izquierda, abajo a la derecha, abajo a la izquierda), y trata de detectar los rostros en cada una de estas posiciones. Si el robot no puede detectar un rostro, el objeto es eliminado del mapa suponiendo que fue un falso positivo, luego el robot continúa visitando los candidatos restantes. Si una persona es detectada se inicia el proceso de reconocimiento, si el rostro es reconocido o se establece como desconocido la posición de la persona en el mapa se marca como visitado para evitar ir nuevamente al mismo lugar. Después de todos las personas en el mapa han sido visitadas, el robot se mueve al siguiente punto P_{i+1p}, y se repiten los pasos del (ii) al (iv).

6.3 Bender: Robot Social

El sistema propuesto para la detección e identificación de humanos ha sido incorporado en Bender, un robot de servicio. Una de las características más relevantes de Bender es su habilidad para interactuar con humanos usando modos de interacción parecidos a los humanos (Rostros, gestos, voz, expresiones faciales, etc.).

6.3.1 Componentes de Hardware

Los componentes principales del robot son: (ver Figura 34)

- Torso. El pecho del robot incorpora un Tablet PC como la plataforma de procesamiento principal, un HP 2710p, potenciado con un 1,2 GHz Intel Core 2 Duo con 2 GB DDR II 667 MHz, con Windows XP Tablet PC Edition. El Tablet incluye conectividad 802.11bg. La pantalla del Tablet PC permite: (i) la visualización de la información relevante para el usuario (un navegador web, imágenes, videos, etc), y (ii) ingreso de datos gracias a la capacidad de pantalla táctil.

- Cabeza. La cabeza del robot incorpora una cámara CCD (Philips ToUcam III - SPC900NC), movimiento pan-tilt de toda la cabeza, y la capacidad de expresar emociones. Esto se logra mediante varios servomotores que se mueven la boca, las cejas y antenas (que representan orejas); y LEDs RGB colocados alrededor de cada ojo. Los movimientos de la cabeza y expresiones se controlan mediante hardware dedicado

(Usando un PIC18F4550), que se comunica con el Tablet PC vía USB. La cámara está conectada al PC usando un hub USB. El peso de la cabeza es de aproximadamente 1,6 kg.

- Visión Térmica. La cámara térmica es una cámara FLIR 320 TAU térmica [34], con una sensibilidad en el rango 7,5 -13,5 micras (rango de la longitud de onda térmica) y una resolución de 324x256 píxeles. Tiene una frecuencia de imagen máxima de 30 Hz (NTSC) y 25 Hz (PAL), la sensibilidad es menor que 75 mK, y el rango de escena es de -40 ° C a +600 ° C. Se utiliza un lente de 9 mm con 48° x 37° FOV. La cámara se coloca en la cabeza del robot y está calibrado de forma manual (de contraste y brillo) en cada sesión a fin de mejorar el contraste entre las partes del cuerpo humano y otros objetos y facilitar la detección de personas.

- Visión 3D. El robot posee un Kinect, que le permite medir profundidad segmentar objetos que se encuentran en una mesa o superficie para la detección y reconocimiento de objetos.

- Brazos. . Los dos brazos están diseñados para permitir que el robot pueda manipular objetos. Ellos son lo suficientemente fuertes como para levantar un gran vaso de agua o una taza de café. Cada brazo tiene seis grados de libertad, dos en los hombros, dos en el codo, una para la muñeca, y uno para la pinza. Los actuadores son ocho servomotores (6 RX-64 y 2 RX-28). Los brazos son controlados directamente desde el PC Tablet a través de USB. El peso del brazo es de aproximadamente 1,8 kg.

- Plataforma móvil. Todas las estructuras descritas están montadas sobre una plataforma móvil. La plataforma es un Pioneer 3-AT, que tiene 4 ruedas, es compacta y ofrece una excelente movilidad, y está conectado a un láser Hokuyo URG-04LX para detectar obstáculos. Esta plataforma está dotada de un microprocesador Hitachi H8. Un notebook (DELL AlienWare) se coloca en la parte superior de la plataforma móvil con la tarea de ejecutar el software de navegación. Este notebook está conectado a la PC Tablet en el pecho por Ethernet ya que ambos computadores se encuentran conectados a un router.

4.2. Arquitectura de Software

Los principales componentes de la arquitectura de software se muestran en la Figura 35. La síntesis y análisis de voz se lleva a cabo en el Tablet PC HP 2710p (con Windows XP Tablet PC Edition); las tareas de visión (reconocimiento general de objetos, rostros, manos y el reconocimiento de gestos) se llevan a cabo en un notebook DELL AlienWare con Windows 7, mientras que la navegación y los módulos de *Mapping* se ejecutan en un DELL ALienWare (Linux), y los módulos de control de bajo nivel se ejecutan en un hardware dedicado (la cabeza y control del brazo). Los computadores se conectan mediante URBI (ver Figura 35). Todos los módulos se ejecutan en el HP 2710p son controlados a través de URBI con UObjects. La navegación y los módulos de *Mapping* se implementan utilizando el kit de herramientas de navegación ROS [97], que proporciona localización, simulación, evasión de obstáculos, entre otras funcionalidades. ROS tiene la ventaja de ser de código abierto y la prestación de apoyo específico para el hardware a nuestra disposición: el módulo navegación envía los comandos de movimiento a la base móvil y lee la información de odometría, además el módulo lee los datos del láser Hokuyo. ROS permite a las interfaces con otros programas, lo que permite procesos de alto nivel (conductas) para enviar comandos.

Los diferentes módulos de software se explican en [100].

Figura 34. Imagen del Robot Bender.

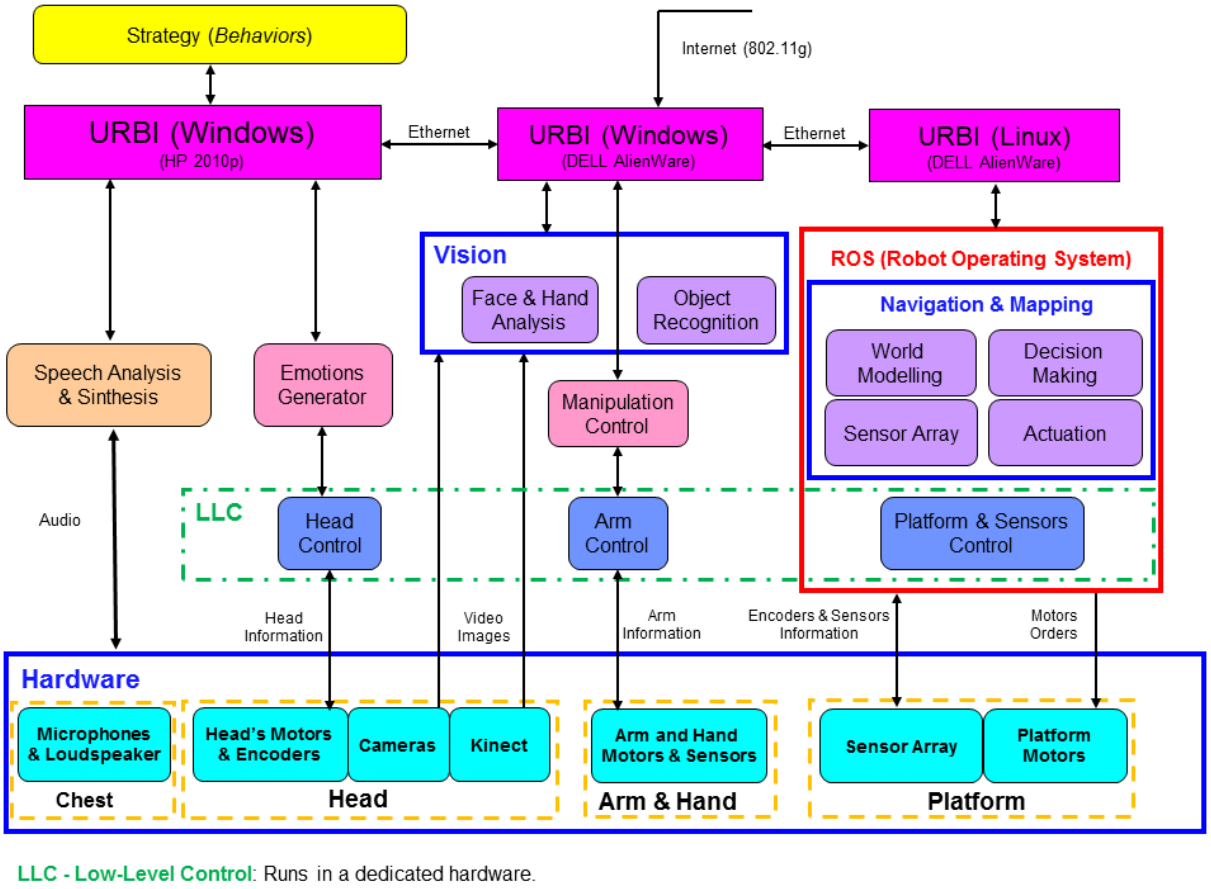


Figura 35. Organización Modular del Software de Bender.

6.4 Resultados

En este capítulo se presentó un sistema con la capacidad de detectar e identificar a los seres humanos en entornos domésticos reales (usando como referencia una prueba de RoboCup@Home). Además, la capacidad de los módulos individuales para la detección de cuerpos y rostros humanos se lleva a cabo utilizando las bases de datos de imagen creadas en entornos domésticos con condiciones de iluminación variable. Los resultados de la detección se presentan en términos de tasa de detección (DR) y número de falsos positivos (FP). La detección se considera correcta si la ventana de detección y la ventana de real se superponen entre sí en al menos el 50% de sus áreas, de lo contrario, la detección es considerada como un falso positivo. En este sistema están incluidos módulos de contexto que fueron validados en el capítulo anterior.

6.4.1 Evaluación en base de datos

Se generaron bases de datos de imágenes (separada en cuatro conjuntos) con el fin de evaluar los diferentes módulos, y para comparar el uso de la información visual y térmica en la detección de cuerpos y rostros humanos en ambientes domésticos. La base de datos fue capturada utilizando una cámara Philips ToUcam III - SPC900NC y una cámara FLIR TAU cámara térmica 320. Estas dos cámaras son las mismas utilizadas en nuestro robot doméstico Bender (ver descripción en la Sección 6.3.1). Las imágenes *EV* y *ET* fueron capturados al mismo tiempo, con las cámaras a dos centímetros de distancia, y a una altura de 120/160 cm dependiendo del experimento. La Figura 36 presenta algunos ejemplos de imágenes capturadas con esta configuración de la cámara.

Los siguientes conjuntos de la base de datos fueron creados[9]:

- *Illumination Database*. Objetivo: comprobar hasta qué punto el tipo de iluminación afecta a los diferentes módulos cuando se utilizan imágenes EV y ET, y para probar los módulos de detección de piel y la forma en que se ven afectados por las condiciones de iluminación. Configuración: 16 personas, 5 imágenes por iluminación y por persona. La distancia entre la cámara y la persona observada se varió de 1 metro a 3 metros con un paso de 50 cm. Tres subgrupos fueron creados, todos capturados a las 9.00 PM, cada uno correspondiente a las condiciones de iluminación diferente:
    - *Indoor light*, corresponde a las condiciones de iluminación estándar de las habitaciones interiores, luz incandescente yo naturales.
    - *Lamp light*, se colocó una lámpara de proyector a un metro detrás de la ubicación de la cámara.
    - *Night light*, las imágenes fueron tomadas sin ningún tipo de iluminación artificial interior, sólo se utilizó la luz natural.
- *Rotation Database*. Objetivo: evaluar si la cantidad de la iluminación afecta a la detección de rostros y personas. Esto se hace para comprobar si es posible detectar *blobs* de persona incluso si la cámara se encuentra a la espalda de ésta, y para evaluar la rotación máxima en la detección de piel visible funciona. Configuración: 18 personas, 5 imágenes con dos tipos de iluminación por persona (150 imágenes). Se consideraron 15 diferentes ángulos de rotación en *YAW*: 0, 5, 10, 15, 20, 25, 30, 45, 60, 75, 90, 105, 120, 135, 180 grados (sentido horario). Las imágenes fueron capturadas

[9] Las bases de datos estarán pronto disponibles en http://vision.die.uchile.cl

con la cámara colocada a una distancia fija de dos metros del sujeto. El tipo de iluminación es luz interior, y una lámpara situada con una orientación de 90 grados con respecto al sujeto.

- *Distance Database*. Objetivo: evaluar si los detectores de rostros EV y ET se ven afectados por el tamaño del rostro en la imagen, es decir, para caracterizar la respuesta de los detectores a diferentes distancias. Configuración: 18 personas, 11 imágenes por persona y tomada con medidas de 50 cm, pasando de 1 metro a 6 metros.

- *Arena Database*. Objetivo: Evaluar las capacidades de detección de los diferentes módulos en tiempo real los ambientes domésticos para detectar cuerpos y rostros humanos. Descripción: 101 imágenes que contiene 171 personas y 104 rostros, de las cuales 37 son rostros frontales. Esta base de datos se construyó en un ambiente en el hogar con personas involucradas en situaciones de la vida real. Esta base de datos contiene personas en las siguientes actividades: caminando, sentados y hablando entre ellos, así como los personas en el suelo. A diferencia de las otras dos bases de datos, el número de seres humanos por la imagen es variable (de 0 a 4). Esta base de datos también contiene dispositivos que generan calor tales como calentadores, que se incluyen para la evaluación de la capacidad de los módulos de detección térmica en discriminar entre este tipo de dispositivo y los cuerpos humanos.

(a) (b)

(c) (d)

Figura 36. Ejemplos de las imágenes de prueba: (a) Indoor Light, Set Illumination, (b) Lamp Light, Set Illumination, (c) Indoor Light, Set Rotation, (d) Set Arena.

6.4.2 Evaluación de módulos

Se realizan tres tipos de evaluaciones: Detección de Piel (Tabla 23 y Tabla 24), Detección de Cuerpo-Humano (Tabla 25 y Tabla 26) y Detección de Rostros (Tabla 27). Para estas evaluaciones se utilizan las bases de datos que se describen en la sección 6.4.1.

6.4.2.1 Detección de Piel

La detección de piel llevada a cabo por los detectores de piel *ET* y *EV* se evalúa con los dos conjuntos de la base de datos (iluminación y rotación) descritos en la sección 6.4.1. Los resultados obtenidos, se muestran en las Tabla 23 y Tabla 24, se puede apreciar que el detector de piel *ET* no depende de las condiciones de iluminación, mientras que el detector de piel *EV* es altamente dependiente de las condiciones de iluminación, dando buenos resultados sólo cuando hay suficiente luz (y la imagen no se encuentra saturada), y fondos que no contengan colores similares a piel. El detector de piel *ET* funciona mejor a los 2 metros distancia de los sujetos, con un menor número de detecciones (tanto verdaderos positivos y falsos positivos) cuando el sujeto está muy lejos de la cámara, y un menor número de detecciones (verdaderos y falsos positivos) cuando el sujeto está más cerca de la cámara, como se puede observar en la Tabla 23. El detector de piel *EV* no muestra resultados diferentes para diferentes distancias (*Illumination Database*), y tiene un rendimiento muy malo en todos los subconjuntos de la base de datos. La principal razón es la gran cantidad colores parecidos a piel en el fondo de las imágenes de la base de datos, que es un problema común en la mayoría de los entornos domésticos que contengan madera y otros materiales color piel (ver Figura 36 para un ejemplo). Es importante mencionar que el detector de piel *ET* es capaz de trabajar con luz de noche (casi no hay iluminación), que es una característica muy importante para un robot doméstico.

Se puede observar en la Tabla 24 que el detector de piel *ET* es muy robusto, y responde bastante bien a las diferentes rotaciones, y es capaz de localizar áreas de piel, incluso si la persona no está mirando a la cámara. El detector de piel *EV* obtiene alrededor de tres veces más falsos positivos que el detector de piel *ET*, y la detección de partes de piel disminuye con el ángulo de rotación. Al igual que en *Illumination Database*, en el conjunto de la rotación el detector de piel *ET* no varía con las condiciones de iluminación y el detector de piel *EV* tiene una tasa de detección inferior cuando se utiliza una lámpara (fuerte).

Tabla 23: Detección de piel visual y térmica en base de datos Illumination. DR: Detection Rate (sobre 16 Sujetos); FP: Número de falsos positivos.

	Indoor Light				**Lamp Light**				**Night Light**			
Distancia	**Térmica**		**Visible**		**Térmica**		**Visible**		**Térmica**		**Visible**	
[mt]	**DR %**	**FP**	**DR %**	**FP**	**DR %**	**FP**	**DR %**	**FP**	**DR %**	**FP**	**DR %**	**FP**
1	96,77	50	6,45	212	96,77	49	51,61	133	90,32	43	0	135
1.5	85,71	25	7,14	232	81,82	30	15,91	151	75	33	0	103
2	80,44	14	6,52	257	78,26	18	0	166	78,26	13	0	289
2.5	68,75	11	4,17	255	72,92	9	2,08	170	68,75	8	0	108
3	63,83	5	14,89	289	66,67	6	2,08	162	64,58	4	0	107

Tabla 24: Detección de piel visual y térmica en base de datos Rotation. DR: Detection Rate (sobre 18 Sujetos); FP: Número de falsos positivos.

	Indoor Light				Lamp Light			
Rotación	**Térmica**		**Visible**		**Térmica**		**Visible**	
[grados]	**DR %**	**FP**	**DR %**	**FP**	**DR %**	**FP**	**DR %**	**FP**
0	96,55	22	65,52	68	98,25	24	12,28	61
5	96,61	22	71,19	66	98,28	22	10,34	63
10	96,61	29	62,71	65	98,28	27	12,07	49
15	96,61	28	59,32	66	98,28	32	8,62	52
20	96,61	27	62,71	66	96,61	29	8,47	49
25	98,28	27	58,62	66	96,67	26	6,67	48
30	98,31	29	55,93	70	98,31	29	1,69	42
45	96,61	29	52,54	75	96,55	25	5,17	47
60	98,21	22	41,07	72	98,25	25	5,26	53
75	89,47	28	36,84	71	87,72	28	7,02	53
90	89,80	36	26,53	71	88,00	30	10,00	42
105	89,13	34	17,39	70	90,91	32	4,55	30
120	94,87	37	12,82	68	94,74	34	2,63	31
135	97,22	35	11,11	63	94,44	41	0,00	26
180	74,19	22	16,13	57	68,75	31	0,00	27

6.4.2.2 Detección de Cuerpos

Se puede observar en las Tabla 25 y Tabla 26 que el detector de cuerpo implementado es muy robusto y capaz de detectar *cuerpos humanos* y las *partes de cuerpo* bajo diferentes condiciones de iluminación y ángulos de vista en las dos bases de datos evaluadas (*Illumination Database & Rotation Database*). Las bases de datos se describen en la sección 6.4.1. La detección de cuerpos en condiciones de luz de noche también funciona bien.

Cabe destacar que el sistema tiene pocos falsos positivos (alrededor de 1,6 por cada imagen), y que los candidatos a cuerpo necesitan ser analizados con más profundidad (por ejemplo, usando un detector de rostros) en caso de tomar decisiones sobre la presencia de los seres humanos.

Tabla 25: Detección térmica de personas en base de datos Illumination. DR: Detection Rate (sobre 16 Sujetos); FP: Número de falsos positivos.

Distancia	**Indoor Light**		**Lamp Light**		**Night Light**	
[mt]	**DR %**	**FP**	**DR %**	**FP**	**DR %**	**FP**
1	100	44	100	47	100	46
1.5	100	57	100	56	100	54
2	100	56	100	56	100	56
2.5	100	56	100	57	100	56
3	100	53	100	54	100	53

Tabla 26: Detección de personas térmica en base de datos Rotation. DR: Detection Rate (sobre 18 Sujetos); FP: Número de falsos positivos.

Rotación	Indoor Light		Lamp Light	
[grados]	DR %	FP	DR %	FP
0	100	30	100	29
5	100	31	100	28
10	100	31	100	27
15	100	32	100	32
20	94,44	32	94,44	33
25	94,44	32	100	32
30	94,44	32	100	32
45	94,44	31	100	29
60	100	30	100	30
75	100	28	100	27
90	100	26	100	27
105	100	27	100	25
120	100	24	100	24
135	100	25	100	24
180	100	27	100	24

6.4.2.3 Detección de Rostros Frontales (Frontal Face Detection)

Como se mencionó anteriormente (sección 6.2.2.3), hemos implementado dos detectores de rostros frontales, uno en imágenes *ET* y uno en imágenes *EV*. La Tabla 27 presenta los resultados de detección obtenidos en la base de datos a distancia (*Distance Database*) que se describe en la sección 6.4.1. Como se puede observar, los detectores funcionan bien cuando los rostros están cerca de la cámara (<3 metros). Es importante destacar que a estas distancias el detector de rostros *ET* tiene una tasa de detección más alto que el detector de rostros *EV*, y que es capaz de detectar más de un rostro al mismo tiempo. Por lo tanto, se confirma que un detector de rostros *ET* basado en el uso de las cascadas de clasificadores *boost*, es capaz de detectar los rostros humanos robustamente.

Para ambos detectores, la tasa de detección se reduce cuando los rostros se alejan de la cámara, pero la disminución es mucho mayor para el detector de rostros *ET*. Esto es causado por la resolución utilizada de la cámara *ET* (324x256) es la mitad de la cámara *EV* (640x480). Por lo tanto, la práctica detector *ET* de rostro frontal no es capaz de detectar los rostros frontal robustamente a más de tres metros de distancia de la cámara. (Este resultado se observa también en la sección 6.4.3).

Tabla 27: Detección de rostros frontales visual y térmica en base de datos Distance. DR: Detection Rate (sobre 16 Sujetos); FP: Número de falsos positivos.

Distancia	Térmica		Visible	
[mt]	DR %	FP	DR %	FP
1	100	0	100	1
1,5	100	0	100	2
2	100	1	88,89	0
2,5	100	1	77,78	1
3	83,33	1	66,67	0
3,5	55,56	1	77,78	1
4	38,89	1	83,33	1
4,5	27,78	0	77,78	0
5	16,67	0	83,33	0
5,5	16,67	0	55,56	0
6	0	0	16,67	0

6.4.3 Detección de Personas en ambientes complejos

La Tabla 28 presenta los resultados de detección de cuerpos, rostros y rostros frontales en la base de datos Arena (*Arena Database* que se describe en la sección 6.4.1) con cinco métodos diferentes:

- Detector *EV* de rostro frontal: Los rostros frontales detectadas en *EV* son la única información que se utiliza para detectar humanos. Este es uno de los métodos estándar que se utiliza para la detección de los seres humanos en una escena desconocida.
- Detector *ET* de rostro frontal: Los rostros frontales detectadas en el espectro térmico son la única información que utilizan para detectar los seres humanos.
- Detector de cuerpo (imágenes *ET*): Los *blobs* de cuerpo detectados, utilizando el detector *ET* humano (ver Sección 6.2.2.2), se utilizan para la detección de los seres humanos.
- Detector *ET* de *blobs* de piel: Los *blobs* de piel que se detectan, utilizando el detector ET de piel (ver sección 6.2.2), se utilizan para la detección de rostros.
- Detector de persona que usa el detector de rostros *EV*: Las personas y los rostros se detectan usando el módulo de *Análisis e Integración de Blobs* descrito en la sección 6.2.2.5. En este caso, se utiliza el detector de rostros *EV*.
- Detector de persona que usa el detector de rostros *ET*: Las personas y los rostros se detectan usando el módulo de *Análisis e Integración de Blobs* descrito en la sección 6.2.2.5. En este caso, se utiliza el detector de rostros *ET*.

Hay que tener en cuenta que en la *Arena Database*, la mayoría de los rostros están muy lejos de la cámara, lo que hace que los resultados del detector de rostros *ET* tengan una tasa de detección muy baja, debido a su baja resolución. (Recordemos que la cámara térmica tiene una resolución de 320x256 píxeles solamente). Esto no sucede con las grandes rostros, y en el "Who's Who?" que es una prueba de referencia (sección 0), el detector *ET* de rostro frontal funciona bien.

Se puede observar en la Tabla 28 que el uso del detector de rostros frontales (ya sea *ET* o *EV*) no es suficiente para la detección de humanos. El detector de rostros frontales *EV* puede detectar los rostros frontales (83,78%) con una tasa muy baja de falsos positivos (sólo el 25 falsos positivos en 101 imágenes), pero se puede detectar sólo 51,91% del número total de rostros (frontal y no-frontal) en un ambiente de hogar.

El *Detector de Cuerpo* puede detectar la mayoría de los humanos, pero tiene un gran número de falsos positivos. No se pueden detectar los rostros. Por otro lado, el detector de *blobs* de piel *ET*, que no es capaz de detectar seres humanos, puede encontrar todas los rostros, pero tiene un gran número de falsos positivos.

El *Detector de Persona* puede reducir el número de falsos positivos considerablemente (a menos de un tercio) cuando se detectan rostros, y puede reducir el número de falsos positivos considerablemente (un quinto) cuando se detectan rostros frontales, sin disminuir la tasa de detección.

Estos resultados demuestran la robustez del sistema propuesto para la detección de personas, rostros y rostros frontales. El detector de persona propuesto puede detectar todos los seres humanos, aproximadamente el 50% de los candidatos a rostros, y ~ 83% de los rostros frontal con un número relativamente bajo de falsos positivos, especialmente en el caso de los rostros frontales. El gran número de falsos positivos al detectar los seres humanos se pueden eliminar basándose en el tamaño de los *blobs*, en otras palabras usando contexto ya que los *blobs* pequeños implican que están ubicados fuera del mapa por lo que pueden ser eliminados (véase, por ejemplo, los *blobs* pequeños de color rojo en la Figura 33(d)). Estos resultados demuestran que el sistema propuesto resuelve adecuadamente el problema de detección de seres humanos en el ámbito doméstico para el uso en un robot de servicio.

Tabla 28: Resultados de la detección de personas, Detección de rostros y detección de rostros frontales en la base de datos Arena. Hay 101 imágenes que contienen en total 171 personas, 104 rostros y 37 rostros frontales. DR: Detection Rate; FP: Número de falsos positivos (para la base de datos completa).

	Detección de Humanos		**Detección de Rostros**		**Detección de Rostros Frontales**	
Método	**DR %**	**FP**	**DR %**	**FP**	**DR %**	**FP**
Detector de Rostros frontales Visible	35,67	25	51,92	25	83,78	25
Detector de Rostros frontales Térmico	3,51	5	5,77	5	16,22	5
Detector de Cuerpos Humanos	99,42	241	-	-	-	-
Detector de blobs de piel Térmica	-	-	100,00	499	-	-
Detector de Personas usando Detector de Rostros Visible	99,42	241	51,92	7	83,78	5
Detector de Personas usando Detector de Rostros Térmica	99,42	241	7,69	1	16,22	1

6.4.4 ‘Who is Who?’ Benchmark

‘Who is Who?’ es una de las pruebas más importantes que se realiza en las competiciones de RoboCup@Home. El objetivo principal es para poner a prueba la capacidad de los robots domésticos "para detectar y reconocer a las personas de forma autónoma en un entorno desconocido" [96]. Para llevar a cabo esta tarea, se espera que "sin necesidad de calibración manual, un robot tendrá que presentarse ante un grupo de personas, pregunte por sus nombres, memorizar y reconocer a la gente cuando las encuentre de nuevo". La prueba se centra "en la detección y reconocimiento de personas, detección y reconocimiento de rostros, seguridad de la navegación y la interacción humano-robot con personas desconocidas" [96]. Básicamente, la prueba es la siguiente: el robot entra en la arena a través de la puerta y se detiene junto a ella. Dos personas entran por la puerta y que se presentan al robot, uno por uno. El robot le pregunta por sus nombres y los memoriza. Cuando un operador le da la orden, el robot va a la habitación y empieza a buscar personas. En la sala, hay otras dos personas que son desconocidas para el robot. Uno de ellos está sentado y el otro está de pie. También hay una persona de pie en la sala que conoce el robot. Cuando el robot encuentra una persona, tiene que acercarse y decir que ha encontrado a una persona. Entonces tiene que reconocer a la persona diciendo su nombre o si la persona es desconocida. La distancia desde el robot a la persona no debe exceder de un metro.

Los siguientes sistemas de detección e identificación de personas se pusieron a prueba en esta prueba:

- *Full Visible*: Sólo el detector de rostros *EV* se utiliza para encontrar a la gente en la arena. Un sistema de reconocimiento de rostros que trabaja en las imágenes *EV* se utiliza para identificarlos. Este es el método estándar usado por la mayoría de los equipos participantes en la RoboCup@Home.
- *Full Térmico*: Se utiliza el detector de personas con la detección de rostros *ET*, que se presenta en las secciones anteriores, para encontrar a la gente en la arena. Un sistema de reconocimiento facial que trabajan en las imágenes *ET* se utiliza para identificarlos.
- *Hibrido Térmico-Visible*: Se utiliza el detector de personas con la detección de rostros *ET*, que se presenta en las secciones anteriores, para encontrar a la gente en la arena. Un sistema de reconocimiento de rostros que trabaja en las imágenes *EV* se utiliza para identificarlos.

La Tabla 29 presenta los resultados de la evaluación de 'Whos’s Who?'. En cada escena hay cinco personas, de los cuales tres eran conocidos por el robot. Dos personas se encontraban con sus rostros mirando hacia el exterior de la arena. Así, el detector de rostros frontales no es capaz de detectar esos dos rostros. El experimento se realizó tres veces para cada método antes mencionado, y en cada experimento las personas conocidas por el robot son las mismas, y su posición en la arena no ha cambiado. Los detectores y módulos de reconocimiento utilizadas corresponden a las descritas en la Sección 6.2 y son caracterizados en la misma sección.

(a) (b)

Figura 37: Ejemplo de prueba ‘Who is Who?’. Configuración de un experimento.

(a) Vista desde arriba. (b) Imagen observada por el robot.

Los resultados obtenidos muestran que el sistema *Full Visible* funciona bien en la detección de rostros (que detectó el 88,9% de los rostros frontales), y el método de reconocimiento de rostro EV tiene una tasa de reconocimiento alta (en promedio se reconoce correctamente 2.7 de las 3 personas detectadas), y se equivoca con sólo 1 de 9 casos. Sin embargo, no se puede detectar a las personas cuyos rostros no están mirando a una dirección donde el robot las puede detectar.

Los resultados obtenidos para el sistema *Full Térmico* muestran que es muy bueno para la detección de personas (que detectó el 86,6% de las personas con 0 falsos positivos), y dio buenos resultados en la detección y verificación de los rostros frontales de las imágenes *ET* (que detecta 88,9% de los rostros frontal presentes), y tiene una tasa buena de reconocimiento (en promedio se reconoce correctamente 3.3 de 4 personas detectadas), pero reconoce erróneamente en 3 de 12 casos.

El sistema *Hibrido Térmico-Visible* tiene un rendimiento similar al *Full Térmico* para la detección de personas y de rostros, pero mejora la tasa de reconocimiento (en promedio se reconoce correctamente 3.7 de 4 personas detectadas), y reconoce erróneamente en 2 de 12 casos.

Los resultados obtenidos muestran que el uso conjunto de información *ET* y *EV* permite altas tasas de detección de personas y altas tasas de reconocimiento al mismo tiempo.

Uno de los principales problemas observados con la cámara térmica es que su respuesta depende de la cantidad de tiempo que la cámara se ha encendido, que afecta considerablemente la tasa de reconocimiento, ya que las imágenes se saturan y en consecuencia los límites deben modificarse. Otro problema es que la temperatura corporal de las personas puede variar ampliamente (por ejemplo, la transpiración). Estos problemas pueden ser tratados en trabajos futuros para mejorar el rendimiento general.

Tabla 29: Evaluación de test 'Who is Who?': De las 5 personas presentes en las imagines, 2 de ellas se encuentran en una posición en donde el detector de rostros frontales no puede encontrarlas. (Ver Figura 37). Todos los métodos se corren 3 veces cada uno. El mejor promedio es mostrado en negrita. Ver el texto para mayores detalles. DR: Detection Rate; FP: Número de falsos positivos (en la evaluación).

Método		**Detección de Rostros**		**Detección de Personas**		**Reconocimiento de Rostros**		
		DR % | TP	**FP**	**TP**	**FP**	**Correcto**	**Incorrecto**	**No encontrado**
Full Visible	Test 1	66.7% | 2	0	40% | 2	0	2	0	3
	Test 2	100% | 3	0	60% | 3	0	2	1	2
	Test 3	100% | 3	0	60% | 3	0	3	0	2
	Prom.	**88.9% | 2.7**	0	53.3% | 2.7	**0**	2.7	**0.3**	2.7
Full Térmico	Test 1	100% |3	0	80% | 4	0	4	0	1
	Test 2	100% | 3	0	100% | 5	0	3	2	0
	Test 3	66.67%| 2	0	80% | 4	0	3	1	1
	Prom.	**88.9% | 2.7**	**0**	**86.6% | 4.3**	**0**	3.3	1	**0.7**
Híbrido Térmico-Visible	Test 1	66.7% | 2	0	80% | 4	0	3	1	1
	Test 2	100% | 3	0	80% | 4	0	4	0	1
	Test 3	100%| 3	0	100% | 5	0	4	1	0
	Prom.	**88.9% | 2.7**	**0**	**86.6% | 4.3**	**0**	**3.7**	0.7	**0.7**

6.5 Análisis de resultados

El desarrollo de robots para ambientes domésticos es una tarea difícil. Uno de los problemas más básicos es encontrar la manera de detectar e identificar los seres humanos de una forma robusta y sin falsas detecciones. En esta sección, se propone un sistema para resolver este problema basado en la integración de diferentes fuentes de información. Las fuentes de información son térmicas y visuales, y se utilizan para la detección de los seres humanos, la localización de sus rostros, y reconocimiento robusto. Se realiza un análisis integrado para detectar objetos candidatos a personas, y luego son procesados con el fin de verificar la presencia de los seres humanos y su identidad. Se utiliza un detector de rostros para verificar la presencia de los seres humanos, y un sistema de reconocimiento de rostro para identificarlos. En el caso que la identificación directa no es posible, se emplea un mecanismo de visión activa para mejorar la relación de pose de un objeto candidato a persona. Se explican las características de los diferentes módulos propuestos, y el sistema propuesto se valida utilizando bases de datos con imágenes reales en ambientes domésticos. Además se realizan pruebas de la RoboCup@Home para validar los resultados obtenidos.

Los resultados presentados demuestran que el sistema propuesto resuelve apropiadamente el problema de la detección e identificación de personas en entornos domésticos. La detección de piel y personas en espectro térmico (*ET*) es robusta en condiciones de iluminación variable y diferentes ángulos de observación, y permite la detección de cuerpos humanos y partes del cuerpo a una distancia apropiada para aplicaciones domésticas (~ 6 metros). Los experimentos también confirman que un detector de rostros *ET*, basado en el uso de una cascada de clasificadores *boost*, es capaz de detectar los rostros humanos de forma robusta.

El uso de una cámara térmica permite a los robots trabajar bajo condiciones de iluminación difíciles (baja iluminación, iluminación desigual, iluminación de diferentes fuentes), y para detectar la ubicación de los seres humanos que están lejos de la cámara con

mayor precisión, mientras que el uso de una cámara de espectro visible permite que el trabajo con imágenes no calibradas, en entornos con muchos objetos calientes, con objetos que tienen texturas o la apariencia de textura y una amplia gama de objetos, debido a la disponibilidad de un mayor número de bases de datos para el entrenamiento de detectores o clasificadores.

El fondo de las imágenes puede resultar un poco molesto cuando se trata de detectar los rostros u objetos con cámaras normales, este problema se puede resolver fácilmente mediante el uso de una cámara térmica. Además, ya que la información dada por el sistema térmico es complementaria a la información proporcionada por el sistema en espectro visible, las detecciones falsas generadas por el sistema térmico pueden ser eliminadas por el sistema EV y viceversa. El problema de la detección de piel puede ser resuelto con mayor facilidad en el espectro térmico que en el espectro visible.

Es importante mencionar que la detección e identificación humana propuesta es capaz de trabajar con iluminación nocturna (casi no hay iluminación), gracias al uso de imágenes térmicas. Esta es una característica muy importante para el uso general robots domésticos, que deben ser capaces de hacerse cargo de las tareas del hogar (por ejemplo, vigilancia, cuidado de ancianos), tanto durante el día y la noche.

El uso de contexto nuevamente permite una mejora importante en el funcionamiento general del sistema, ya que se logran eliminar detecciones falsas, limitar el área de búsqueda dentro de la imagen, etc.

Como se mencionó anteriormente, uno de los principales problemas observados en el uso de cámaras térmicas es que su respuesta depende del tiempo que la cámara está encendida, lo que afecta la tasa de reconocimiento, ya que las imágenes se saturan y los umbrales deben ser modificados.

Capítulo 7
Conclusiones

En la actualidad un problema fundamental para los sistemas robóticos que basan su sistema sensorial en la utilización de cámaras de video y sistemas de visión computacional es detectar y reconocer objetos de interés en ambientes no controlados. Por otro lado, el análisis del rostro juega un papel muy importante en la construcción de un sistema de Interacción Humano-Robot (HRI o *Human Robot Interaction*) que permita a los humanos interactuar con sistemas robóticos de un modo natural. En este trabajo de tesis se diseñó e implementó un sistema de visión que opera en ambientes no controlados que es capaz de detectar y reconocer rostros humanos en forma robusta, utilizando métodos de visión activa e integrando diferentes tipos de contexto.

Se planteó una metodología con que se construye el sistema de visión propuesto en forma general y se define cuales son los módulos principales que lo componen. Para lo anterior, se definió que tipos de contexto (contexto de configuración, contexto de situación, contexto de coherencia espacial, entre otros) y los módulos principales que se utilizan. Entre los cuales están los módulos de detección y reconocimiento de rostros y un módulo de visión activa que permite realizar modificaciones a las observaciones para así mejorar el rendimiento del sistema de reconocimiento.

Se desarrolló un simulador (Capítulo 5) para la validación de sistema general y en particular evaluar el funcionamiento de los diferentes módulos planteados. Este simulador es una poderosa herramienta que genera las observaciones de un agente dentro de un mapa virtual con personas. Para esto se construyó una base de datos, que esta compuesta más de 700 imágenes de cada sujeto, en donde se varían: (1) rotaciones del rostro en *yaw* y *pitch*, (2) iluminación y (3) fondos no homogéneos.

Del estudio comparativo que se realizó utilizando el simulador se puede concluir que el mejor método de detección de rostros utiliza clasificadores de tipo *boosted* en cascada, y mejora el rendimiento mediante el uso de un enfoque de *coarse-to-fine* en el entrenamiento de las cascadas. El método presenta robustez ante rotaciones de los rostros en *yaw* y *pitch*, y una baja tasa de falsos positivos. Con respecto a los métodos de reconocimiento, el estudio comparativo muestra que los métodos basados en características LBP son más robustos ante rotaciones de los rostros que los métodos Gabor y WLD en ambientes *indoor*.

De la evaluación de módulos se puede concluir que los módulos de contexto realizan un aporte significativo en el rendimiento del sistema de visión. Los diferentes filtros de contexto logran eliminar la mayoría de los falsos positivos, permitiendo una detección más precisa de la posición de las personas dentro del mapa. Solo se agregan al Mapa de personas un número reducido de sujetos, debido a que el filtro de coherencia espacial descarta detecciones de la misma persona, cuando esta ya ha sido agregada al mapa de personas. Además el número falsas detecciones de personas descartadas usando los diferentes filtros de contexto son en promedio 27.9 para el primer conjunto de parámetros y en promedio 22.2 para el segundo conjunto de parámetros. Estas detecciones son descartadas porque no cumplen con las leyes físicas o las alturas determinadas por el sistema están fuera de los parámetros correctos. El módulo de reconocimiento de rostros funciona como se esperaba, siendo coherente con los resultados de los estudios comparativos. La tasa de reconocimiento mejora cuando se usan los diferentes módulos de contexto. Se tiene un 78.41% de reconocimiento de rostros sin el uso de los filtros de contexto, y un 86.77% de reconocimiento de rostros con el uso de los módulos de: Mapa de personas, filtro de contexto espacial, filtro de coherencia espacial y el filtro de contexto de situación. El uso de visión activa permite, entre otras cosas, que se construya una mejor base de datos (en caso que la base de datos se construye *online*), y con esto la tasa de reconocimiento mejora. Si la base de datos se construye *offline*, el uso de visión activa permite que la tasa de reconocimiento mejore de 86.77% a un 92.92% en promedio.

Se propuso un sistema robusto para la detección y la identificación de seres humanos en entornos domésticos para ser usado en un robot de servicio (Capítulo 6), de tal forma de poder evaluar en una aplicación real el funcionamiento del sistema de visión propuesto en el Capítulo 4. En esta aplicación se le agregan al sistema nuevos perceptores, módulos, y un sensor extra (una cámara térmica). El uso de la cámara térmica permite al robot de servicio: (1) trabajar bajo condiciones de iluminación difíciles (como por ejemplo baja iluminación) y (2) detectar la ubicación de personas que estén lejos de la cámara o que no están mirando directamente a ella. La cámara térmica es un complemento al uso de la cámara de espectro visible. Las detecciones falsas generadas por el sistema térmico pueden ser eliminadas por el sistema de espectro visible (*EV)* y viceversa.

Se muestra (Capítulo 6) que el sistema propuesto resuelve apropiadamente el problema de la detección e identificación de personas en entornos domésticos. La detección de piel y personas en espectro térmico (*ET*) es robusta en condiciones de iluminación variable y diferentes ángulos de observación, y permite la detección de cuerpos humanos y partes del cuerpo a una distancia apropiada para aplicaciones domésticas (~ 6 metros). Los experimentos también confirman que un detector de rostros *ET*, basado en el uso de una cascada de clasificadores *boost*, es capaz de detectar los rostros humanos de forma robusta. Uno de los principales problemas de trata de detectar los rostros u objetos con cámaras normales es el fondo complejo de las imágenes que puede resultar un poco molesto ya que los objetos de interés se confunden, este problema se pudo resolver fácilmente mediante el uso de la cámara térmica. Por la razón anterior, el problema de la detección de piel puede ser resuelto con mayor facilidad en el espectro térmico que en el espectro visible. El uso e integración de información proveniente de diferentes tipos de contexto mejoró el rendimiento del sistema de detección y reconocimiento de rostros, sin aumentar significativamente en el tiempo de procesamiento requerido. La visión activa fue otro factor importante para mejorar el desempeño del sistema en general, en todos los experimentos influyó de forma positiva, específicamente mejoró la construcción de la base de datos cuando se realiza online y en el reconocimiento de rostros mejorando la alineación.

En general, se propuso una metodología general para la construcción de sistemas de reconocimiento que usen contexto y visión activa para mejorar el desempeño. Esta metodología fue validada en un simulador y evaluada en el sistema de reconocimiento de humanos en ambientes domésticos usando información visual y térmica. La integración de diferentes instancias de contexto ayudo a mejorar el rendimiento del sistema de reconocimiento. El uso de un módulo de visión activa, que retroalimenta el módulo de actuación utilizando la información obtenida desde el sistema de visión, permite modificar las observaciones realizadas y con ello mejorar el rendimiento del sistema de reconocimiento de rostros.

Como trabajo a futuro, en el sistema planteado se pueden utilizar otras instancias de contexto, como por ejemplo contexto temporal y usar un módulo de tracking de personas que mantenga actualizado la posición en el mapa cada vez que las detecte. Se puede incluir un módulo que analice la escena para extraer información de la posible ubicación de las personas y reducir el espacio de búsqueda. Un módulo que seria útil desarrollar es un estimador de pose, para poder saber hacia donde esta mirando la persona y modificar la observación para que el rostro tenga una rotación menor en *yaw*.

Al simulador se le pueden ampliar las funcionalidades, permitiendo realizar evaluaciones de otros tipos de clasificadores, por ejemplo clasificadores de género, edad, raza, etc. Agregar funciones que le permitan generar imágenes con más de una persona, o manejar otro tipo de oclusiones. Aumentar el tamaño de la base de datos para realizar evaluaciones más extensas de los diferentes métodos.

Con respecto al sistema de visión integrado en el robot de servicio Bender, en el futuro se pueden agregar más sensores para mejorar el rendimiento y complementar información, por ejemplo un Kinect. Además seria útil integrar el uso de contexto para otras tareas que realice como el seguimiento de personas, la búsqueda de objetos, etc.

Bibliografía

[1] A. F. Abate, M. Nappi, D. Riccio, G. Sabatino, 2D and 3D face recognition: A survey, Pattern Recognition Letters, Vol. 28, pp. 1885-1906, 2007.

[2] A. Johnston, H. Hill, N. Carman, "Recognizing Faces: Effects of Lighting Direction, Inversion and Brightness Reversal", Cognition, Vol. 40, pp. 1- 19, 1992.

[3] A. M. Martinez, "Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class", IEEE Trans. Pattern Anal. Mach. Intell., Vol. 24, no. 6, pp. 748–763, Jun. 2002.

[4] A. Pentland, B. Moghaddam, T. Starner, "View-Based and modular eigenspaces for face recognition", in Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 84–91, 1994.

[5] A. Torralba, K. Murphy, W. Freeman, M. Rubin. "Context-based vision system for place and object recognition". In Intl. Conf. Computer Vision, 2003.

[6] A. Torralba, P. Sinha. "On Statistical Context Priming for Object Detection". ICCV, Eighth International Conference on Computer Vision (ICCV'01), July 7-14, Vancouver, British Columbia, Canada, 2001.

[7] A. Torralba. "Contextual priming for object detection". Intl. J. Computer Vision, 53(2):153–167 (2003).

[8] A.S. Mian, M. Bennamoun, and R. Owens, An Efficient Multimodal 2D-3D Hybrid Approach to Automatic Face Recognition, IEEE Trans. on Patt. Analysis and Machine Intell., Vol. 29, No. 11, pp. 1927-1943, Nov. 2007.

[9] Ahonen, T., Hadid, A. Pietikäinen, M., "Face description with local binary patterns: application to face recognition". IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 12, pp. 2037-2041, Dec., 2006.

[10] B.A. Wandell. "Foundations of Vision", Sinauer Associates (1995).

[11] Bellotto, N., Huosheng Hu, Multisensor-Based Human Detection and Tracking for Mobile Service Robots Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on Feb. 2009, Vol 39, Issue:1, 167 – 181, 2009.

[12] Bertozzi, M.; Broggi, A.; Fascioli, A.; Graf, T.; Meinecke, M.-M.; , Pedestrian detection for driver assistance using multiresolution infrared vision, IEEE Trans. on Vehicular Technology, vol.53, no.6, pp. 1666- 1678, Nov. 2004.

[13] Binelli, E., Broggi, A., Fascioli, A., Ghidoni, S., Grisleri, P., Graf, T., Meinecke, M. "A modular tracking system for far infrared pedestrian recognition," 2005 IEEE Intelligent Vehicles Symposium, pp. 759- 764, 6-8 June 2005.

[14] Böhme H.J., Wilhelma T., Keya J., Schauera C., Schrötera C., Großa H-M. and Hempelb T., An approach to multi-modal human–machine interaction for intelligent service robots, Robotics and Autonomous Systems, Volume 44, Issue 1, Pages 83-96, 31 July 2003 .

[15] C. Xiang, X. A. Fan, T. H. Lee, "Face recognition using recursive Fisher linear discriminant", IEEE Trans. Image Process., Vol. 15, no. 8, pp. 2097–2105, Aug. 2006.

[16] Call for Papers Special Issue in Real-World Face Recognition, IEEE Trans. on PAMI : http://www.eecs.northwestern.edu/~ganghua/ghweb/CFP_TPAMI_FR.htm

[17] Carolina Galleguillos, Serge Belongie. Context Based Object Categorization: A Critical Survey. Technical Report, Dept. of Computer Science and Engineering, University of California, San Diego, 2008.

[18] Correa, M., Hermosilla, G., Verschae, R., and Ruiz-del-Solar, J. (2012). Human Detection and Identification by Robots using Thermal and Visual Information in Domestic Environments, Journal of Intelligent and Robotic Systems, Vol. 66, No.1-2, pp. 223-243, 2012. (ISI)

[19] Correa, M., Ruiz-del-Solar, J., Bernuy, F. “Face Recognition for Human-Robot Interaction Applications: A Comparative Study”. Lecture Notes in Computer Science (RoboCup Symposium 2008) July 15 – 18, Suzhou, China, 2008.

[20] Correa. M., Javier Ruiz-del-Solar, Parra-Tsunekawa, I. (2010). A Virtual Environment for realistic Testing and Training of Face Detection and Recognition Systems, 19th IEEE Int. Symposium in Robot and Human Interactive Communication – Ro-Man 2010, Sept. 12-15, 2010, Viareggio, Italy.

[21] D. L. Swets, J.Weng, “Using discriminant eigenfeatures for image retrieval”, IEEE Trans. Pattern Anal. Mach. Intell., Vol. 18, no. 8, pp. 831–836, Aug. 1996.

[22] D. Lowe, Distinctive “Image Features from Scale-Invariant Keypoints”. Int. Journal of Computer Vision, 60 (2): 91-110, Nov. 2004.

[23] D.H. Hubel. “Eye, Brain, and Vision”, Scientific American Library, New York (1995).

[24] Dalal N. Triggs B., "Histograms of oriented gradients for human detection," in CVPR, vol. 1, June 2005, pp. 886-893 vol. 1.

[25] Dalal N., Triggs B., Schmid C., "Human detection using oriented histograms of flow and appearance," in ECCV (2), 2006, pp. 428-441.

[26] Dollar P., Wojek C., Schiele B., Perona P., "Pedestrian detection: A benchmark," in CVPR, June 2009, pp. 304-311.

[27] E. Nowak, F. Jurie, “Learning Visual Similarity Measures for Comparing Never Seen Objects”, Proc. IEEE Conf. on Computer Vision and Pattern Recognition - CVPR '07, Minnesota, USA, pp.1-8, 17-22, June 2007.

[28] Encuesta nacional de salud ENS Chile 2009-2010.

[29] Enzweiler M., Gavrila D., "Monocular pedestrian detection: Survey and experiments," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 31, no. 12, pp. 2179 -2195, dec. 2009.

[30] Face Recognition Grand Challenge, Official website. Available in http://www.frvt.org/FRGC/

[31] Face Recognition Homepage. Available in January 2008 in: http://www.face-rec.org/

[32] Faces in Real-Life Images Workshop, Oct. 17th 2009, ECCV 2008: http://hal.inria.fr/REALFACES2008/en

[33] Felzenszwalb P. F., Girshick R. B., Mcallester D., "Cascade object detection with deformable part models," in Proc. of IEEE Int'l Conference on Computer Vision and Pattern Recognition, 2010.

[34] FLIR TAU 320 thermal camera. Information available on Dec. 2010 in http://www.flir.com/cvs/cores/uncooled/products/tau/

[35] G. Guo, S.Z. Li, K. Chan, Face Recognition by Support Vector Machines, Proc. of the IEEE International Conference on Automatic Face and Gesture Recognition, 26-30 March 2000, Grenoble, France, pp. 196-201.

[36] G.B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments. University of Massachusetts, Amherst, Technical Report 07-49, Oct. 2007.

[37] Gavrila D., Munder S., "Multi-cue pedestrian detection and tracking from a moving vehicle," Int. Journal of Computer Vision, vol. 73, pp. 41-59, 2007.

[38] Guerrero, P., Ruiz-del-Solar, J., Palma-Amestoy, R. “Spatiotemporal Context in Robot Vision: Detection of Static Objects in the RoboCup Four Legged League”, Proc. 1st Int. Workshop on Robot Vision, in 2nd Int. Conf. on Computer Vision Theory and Appl. – VISAPP 2007, pp. 136 – 148, March 8 – 11, Barcelona, Spain (2007).

[39] H. D. Ellis, “Introduction to Aspects of Face Processing: Ten Questions in Need of Answers”, in Aspects of Face Processing, H. Ellis, M. Jeeves, F. Newcombe, A. Young, eds., Dordrecht, Nijhoff, pp. 3-13, 1986.

[40] Heo, J., Abidi, B., Kong, S.G., and Abidi, M. 2003. Performance Comparison of Visual and Thermal Signatures for Face Recognition. Biometric Consortium Conference, Arlington, VA.

[41] Hermosilla, G., Loncomilla, P., Ruiz-del-Solar, J. Thermal Face Recognition using Local Interest Points and Descriptors for HRI Applications. Lecture Notes in Computer Science (RoboCup Symposium 2010), 2010, (in press).

[42] Hermosilla, G., Ruiz-del-Solar, J., Verschae, R., and Correa, M. (2012). A Comparative Study of Thermal Face Recognition Methods in Unconstrained Environments, Pattern Recognition, Vol. 45, No. 7, pp. 2445-2459. (ISI)

[43] Hermosilla, G., Ruiz-del-Solar, J., Verschae, R., Correa, M. Face Recognition using Thermal Infrared Images for Human-Robot Interaction Applications: A Comparative Study, 6th IEEE Latin American Robotics Symposium – LARS 2009, Oct. 29 - 30, 2009, Valparaíso, Chile (CD Proceedings).

[44] Hjelmas E., Kee Low B., Face Detection: A Survey, Computer Vision and Image Understanding, Volume 83, Issue 3, Pages 236-274, September 2001.

[45] I. Biederman, P. Kalocsai, "Neural and Psychophysical Analysis of Object and Face Recognition", in Face Recognition: From Theory to Applications, H. Wechsler, P. J. Phillips, V. Bruce, F.F. Soulie and T.S. Huang, eds., Berlin: Springer-Verlag, pp. 3-25, 1998.

[46] iRobot Official Website. Available on Dec. 2010 in http://store.irobot.com/home/index.jsp

[47] J. Ruiz-del-Solar, C. Devia, P. Loncomilla, F. Concha, "Offline Signature Verification using Local Interest Points and Descriptors", Lecture Notes in Computer Science 5197 (CIARP 2008), pp. 22-29, 2008.

[48] J. Ruiz-del-Solar, J. Quinteros, Illumination Compensation and Normalization in Eigenspace-based Face Recognition: A comparative study of different pre-processing approaches, Pattern Recognition Letters, Vol. 29, No. 14, pp. 1966-1979, 2008.

[49] J. Ruiz-del-Solar, P. Loncomilla, C. Devia, "A New Approach for Fingerprint Verification based on Wide Baseline Matching using Local Interest Points and Descriptors", Lecture Notes in Computer Science 4872 (PSIVT 2007), pp. 586-599, 2007.

[50] J. Ruiz-del-Solar, P. Navarrete, "Eigenspace-based face recognition: a comparative study of different approaches", IEEE Transactions on Systems, Man and Cybernetics, Part C, Vol. 35, Issue 3, pp. 315-325, August 2005.

[51] J. Zou, Q. Ji, G. Nagy, "A Comparative Study of Local Matching Approach for Face Recognition", IEEE Transactions on Image Processing, Vol. 16, Issue 10, pp. 2617-2628, Oct. 2007.

[52] Jie Chen, Shiguang Shan, Chu He, Guoying Zhao, Matti Pietikäinen, Xilin Chen, Wen Gao, "WLD: A Robust Local Image Descriptor," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 9, pp. 1705-1720, Aug. 2010, doi:10.1109/TPAMI.2009.155

[53] Jones M., and Rehg, J. M. , Statistical Color Models with Application to Skin Detection, International Journal of Computer Vision, pages 81-96, volume 46, number 1, 2002.

[54] K. Etemad, R. Chellappa, "Discriminant analysis for recognition of human face images", J. Opt. Soc. Amer., Vol. 14, pp. 1724–1733, 1997.

[55] K. Murphy, A. Torralba, W. Freeman. "Using the forest to see the trees: a graphical model relating features, objects and scenes". In Advances in Neural Info. Proc. Systems (2003).

[56] K. Tan, S. Chen, "Adaptively weighted sub-pattern PCA for face recognition", Neurocomputing, Vol. 64, pp. 505–511, 2005.

[57] Kemelmacher-Shlizerman, I.; Basri, R., "3D Face Reconstruction from a Single Image Using a Single Reference Face Shape," Pattern Analysis and Machine Intelligence, IEEE Transactions on , vol.33, no.2, pp.394-405, Feb. 2011.

[58] Kong S., Heo J., Abidi B., Paik J., and Abidi M., Recent Advances in Visual and Infrared Face Recognition - A Review, the Journal of Computer Vision and Image Understanding, Vol. 97, No. 1, pp. 103-135, June 2005.

[59] L. Wiskott, J.-M. Fellous, N. Krueuger, C. von der Malsburg, Face Recognition by Elastic Bunch Graph Matching, Chapter 11 in Intelligent Biometric Techniques in Fingerprint and Face Recognition, eds. L.C. Jain et al., CRC Press, 1999, pp. 355-396.

[60] L. Wiskott, J.-M. Fellous, N. Krueuger, C. von der Malsburg, Face Recognition by Elastic Bunch Graph Matching, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 19, No. 7, 1997, pp. 775-779.

[61] Labeled Faces in the Wild database, Results webpage. Available in http://vis-www.cs.umass.edu/lfw/results.html

[62] Li, S., Chu, R., Ao, M., Zhang, L., and He, R., "Highly Accurate and Fast Face Recognition Using Near Infrared Images", Lecture Notes in Computer Science 3832, 151-158, 2005.

[63] Li, S., Chu, R., Liao, Sh., Zhang, L., Illumination Invariant Face Recognition Using Near-Infrared Images, IEEE Trans. on Pattern Analysis and Machine Intell., vol.29, no.4, pp.627-639, April 2007.

[64] Liyuan Li, Ying Ting Koh, Shuzhi Sam Ge, Weimin Huang, Stereo-based human detection for mobile service robots. Control, Automation, Robotics and Vision Conference, 2004. ICARCV 2004, pp 74 – 79, Vol. 1 , Dec. 2004.

[65] M. Jones (2009). Face Recognition: Where We Are and Where To Go From Here, Mitsubishi Electric Research Laboratories Technical Report TR2009-023, June 2009.

[66] M. Kirby, L. Sirovich, "Application of the Karhunen-Loeve Procedure for the Characterization of Human Faces", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 12, No. 1, pp. 103-108, January 1990.

[67] M. S. Bartlett, J. R. Movellan, T. J. Sejnowski, "Face recognition by independent component analysis", IEEE Trans. Neural Netw., Vol. 13, no. 6, pp. 1450–1464, Jun. 2002.

[68] M. Turk, A. Pentland, "Eigenfaces for recognition", J. Cogn. Neurosci., Vol. 13, no. 1, pp. 71–86, 1991.

[69] Mauricio Correa, Javier Ruiz-del-Solar, Parra-Tsunekawa, I., Rodrigo Verschae (2011). A Realistic Simulation Tool for Testing Face Recognition Systems under Real-World Conditions. Lecture Notes in Computer Science 6556 (RoboCup Symposium 2010), pp. 13–24.

[70] Medionia G., R.J. Françoisa A., Siddiquia M., Kima K. and Yoonb H., Robust real-time vision for a personal service robot, Computer Vision and Image Understanding, Volume 108, Issues 1-2, Pages 196-203, Special Issue on Vision for Human-Computer Interaction , October-November 2007.

[71] Meis, U., Oberlander, M., Ritter, W. "Reinforcing the reliability of pedestrian detection in far-infrared sensing," 2004 IEEE Intelligent Vehicles Symposium, vol., no., pp. 779- 783, 14-17, June 2004.

[72] Mudaly, S.S., Novel computer-based infrared pedestrian data-acquisition system, Electronics Letters , vol.15, no.13, pp.371-372, June 21 1979.

[73] Munder and D. Gavrila, "An experimental study on pedestrian classification," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, pp. 1863-1868, 2006.

[74] N. Vaswani, R. Chellappa, "Principal components null space analysis for image and video classification", Image Processing, IEEE Transactions on, Vol. 15, Issue 7, pp. 1816-1830. July 2006.

[75] Nanda, H.; Davis, L.; , "Probabilistic template based pedestrian detection in infrared videos," 2002 IEEE Intelligent Vehicle Symposium, pp. 15- 20 vol.1, 17-21 June 2002.

[76] OpenCV - Open Computer Vision Library. Main page is http://opencvlibrary.sourceforge.net/

[77] P. J. Phillips, H. Moon, S. A. Rizvi, P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms", IEEE Trans. Pattern Anal. Mach. Intell., Vol. 22, no. 10, pp. 1090–1104, Oct. 2000.

[78] P. J. Phillips, H. Wechsler, J. Huang and P. Rauss, The FERET database and evaluation procedure for face recognition algorithms, Image and Vision Computing J., Vol. 16, no. 5, pp. 295-306, 1998.

[79] P. J. Phillips, P. J. Grother, R. J. Micheals, D. M. Blackburn, E. Tabassi, J. M. Bone, "Face recognition vendor test 2002: Evaluation report", National Institute of Standards and Technology - Interagency Reports - NISTIR N° 6965, 2003.

[80] P. N. Belhumeur, J. P. Hespanha, D. J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection", IEEE Trans. Pattern Anal. Mach. Intell., Vol. 19, no. 7, pp. 711–720, Jul. 1997.

[81] P. Phillips, P. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, Overview of the Face Recognition Grand Challenge, Proc. of the IEEE Conf. Computer Vision and Pattern Recognition – CVPR 2005, Vol. 1, p. 947-954.

[82] P. Sinha, B. Balas, Y. Ostrovsky, R. Russell, “Face Recognition by Humans: 19 Results All Computer Vision Researchers Should Know About”, Proceedings of the IEEE, Vol. 94, No. 11, pp. 1948-1962, November 2006.

[83] P. Viola and M. Jones, Fast and robust classification using asymmetric adaboost and a detector cascade, Advances in Neural Inform. Processing System 14. MIT Press, 2002.

[84] P. Viola M. Jones, Rapid object detection using a boosted cascade of simple features, Proc. IEEE Conf. on Computer Vision and Pattern Recognition – CVPR 2001, Vol. 1, pp. 511 – 518, 2001.

[85] P.K. Kalocsai, W. Zhao, E. Elagin, “Face Similarity Space as Perceived by Humans and Artifical Systems”, in Proceedings, International Conference on Automatic Face and Gesture Recognition, pp. 177-180, April 14-16 , Nara, Japan, 1998.

[86] Paisitkriangkrai S., Shen C., Zhang J., "Fast pedestrian detection using a cascade of boosted covariance features," Circuits and Systems for Video Technology, IEEE Transactions on, vol. 18, no. 8, pp. 1140 -1151, aug. 2008.

[87] Piotr Dollár, Christian Wojek, Bernt Schiele, Pietro Perona. Pedestrian Detection: An Evaluation of the State of the Art, IEEE Trans. on Pattern Analysis and Machine Intelligence, v. 34, n. 4, April 2012.

[88] Queirolo, C.C.; Silva, L.; Bellon, O.R.P.; Segundo, M.P.; , "3D Face Recognition Using Simulated Annealing and the Surface Interpenetration Measure," Pattern Analysis and Machine Intelligence, IEEE Transactions on , vol.32, no.2, pp.206-219, Feb. 2010

[89] R. Chellappa, C.L. Wilson, S. Sirohey, Human and Machine Recognition of Faces: A Survey, Proceedings of the IEEE, Vol. 83, Issue 5, May 1995, pp. 705-740.

[90] R. Gottumukkal, V. K. Asari, “An improved face recognition technique based on modular PCA approach”, Pattern Recognit. Lett., Vol. 25, no. 4, pp. 429–436, 2004.

[91] R. Lienhart, A. Kuranov, V. Pisarevsky, Empirical analysis of detection cascades of boosted classifiers for rapid object detection. Lecture Notes in Computer Science 2781 (DAGM 2003) (Springer 2003) 297-304.

[92] R. Singha, M. Vatsaa, and A. Noore, Integrated multilevel image fusion and match score fusion of visible and infrared face images for robust face recognition, Pattern Recognition, Vol. 41, Issue 3, pp. 880-893, March 2008.

[93] R. Verschae, Object Detection using Nested Cascades of Boosted Classifiers, Ph.D. Thesis, Universidad de Chile, 2010.

[94] R.L. Gregory. “Eye and Brain”, Princeton University Press, 5th edition (1997).

[95] Reis BY, Pagano M, Mandl KD. “Using temporal context to improve biosurveillance”. Proc Natl Acad Sci USA, 2003.

[96] RoboCup@Home Official Website. Available on Dec. 2010 in http://www.ai.rug.nl/robocupathome/

[97] ROS (Robot Operating System). http://www.ros.org

[98] Ruiz-del-Solar J., Verschae R., Robust Skin Segmentation Using Neighborhood Information, In the Eleventh International Conference on Image Processing (ICIP 2004), pp. 207-210, October 24-27, 2004, Singapore, IEEE Press., 2004.

[99] Ruiz-del-Solar, J. Verschae, R. Vallejos, P. and Correa, M. “Face Analysis for Human Computer Interaction Applications”, Proc. 2nd Int. Conf. on Computer Vision Theory and Appl. – VISAPP 2007, Special Sessions, pp. 23 – 30, Barcelona, Spain. March 2007.

[100] Ruiz-del-Solar, J., Correa, M., Lee-Ferng, J., Hevia-Koch, P., Parra, I., Mascaró, M. (2010). UChile HomeBreakers 2010 Team Description Paper, RoboCup Symposium 2010, 19-25 June 2010, Singapore (CD Proceedings).

[101] Ruiz-del-Solar, J., Verschae, R., Correa, M. Recognition of Faces in Unconstrained Environments: A Comparative Study. EURASIP Journal on Advances in Signal Processing (Recent Advances in Biometric Systems: A Signal Processing Perspective), Vol. 2009, Article ID 184617, 19 pages, 2009.

[102] S. Zeki. “Vision of the Brain”, Blackwell Publishing Incorporated (1993).

[103] S.E. Palmer. “Vision Science: Photons to Phenomenology”. The MIT Press (1999).

[104] Santosh K. Divvala, Derek Hoiem, James H. Hays, Alexei A. Efros, Martial Hebert. An Empirical Study of Context in Object Detection. Proceedings of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Junio, 2009.

[105] Satake J. and Miura J., Robust Stereo-based Person Detecting and Tracking for a Person Following Robot, Proc. ICRA 2009 workshop on person detection and tracking, Kobe, Japan, 2009.

[106] Selinger, A. and Socolinsky, D.A. 2001. Appearance-Based Facial Recognition Using Visible and Thermal Imagery: A Comparative Study. Technical Report 02-01, Equinox Corporation.

[107] Singh, Richa (2008) Integrated multilevel image fusion and match score fusion of visible and infrared face images for robust face recognition. Pattern Recognition 41(3).

[108] Strat, T. “Employing contextual information in computer vision”. Proceedings of DARPA Image Understanding Workshop, , Washington, DC, April 1993.

[109] T. Ahonen, A. Hadid, M. Pietikainen, “Face recognition with local binary patterns”, European Conference on Computer Vision – ECCV 2004, Prague, May 11-14, pp. 469–481, 2004.

[110] T. De Bie, N. Cristianini, R. Rosipal, “Eigenproblems in Pattern Recognition”, Handbook of Computational Geometry for Pattern Recognition, Computer Vision, Neurocomputing and Robotics, E. Bayro-Corrochano (editor), Springer-Verlag, Heidelberg, April 2004.

[111] T.-K. Kim, H. Kim, W. Hwang, J. Kittler, “Component-Based LDA face description for image retrieval and MPEG-7 standardisation”, Image Vis. Comput., Vol. 23, pp. 631–642, 2005.

[112] Tao J. Odobez, J.-M., "Fast human detection from videos using covariance features," in Workshop on VS at ECCV, 2008.

[113] Tuzel O., Porikli F., Meer P., "Pedestrian detection via classification on riemannian manifolds," IEEE T-PAMI, vol. 30, no. 10, pp.1713-1727, Oct. 2008.

[114] Verschae R., Ruiz-del-Solar J., Correa M. “Face Recognition in Unconstrained Environments: A Comparative Study” Presented at the "Faces in Real-Life Images: Detection, Alignment, and Recognition" workshop at ECCV 2008, Marseille, France, October 17, 2008.

[115] Verschae, R., Ruiz-del-Solar, J., Correa, M. “A Unified Learning Framework for object Detection and Classification using Nested Cascades of Boosted Classifiers”. Machine Vision and Applications (DOI - 10.1007/s00138-007-0084-0, Springer) (in press), ISI, 2007.

[116] Viola P. A., Jones, M. J., Snow D., "Detecting pedestrians using patterns of motion and appearance," IJCV, vol. 63, no. 2, pp. 153-161,2005.

[117] W. Zhao, R. Chellappa, A. Rosenfeld, P.J. Phillips, “Face Recognition: A Literature Survey”, ACM Computing Surveys, Vol. 35, Issue 4, pp. 399-458, 2003.

[118] W.S. Yambor, “Analysis of PCA-based and Fisher Discriminant-Based Image Recognition Algorithms”, M.S. Thesis, Technical Report CS-00-103, Computer Science Department, Colorado State University, July 2000.

[119] Wilder, J., Phillips, P.J., Cunhong Jiang, Wiener, S., Comparison of visible and infra-red imagery for face recognition, Proc. of the 2nd Int. Conf. on Automatic Face and Gesture Recognition, pp.182-187, 14-16 Oct 1996.

[120] Wilhelm T., Böhme H.-J., and Gross H.-M., A multi-modal system for tracking and analyzing faces on a mobile robot, Robotics and Autonomous Systems, Volume 48, Issue 1, Pages 31-40, European Conference on Mobile Robots (ECMR '03) , 31 August 2004.

[121] Wisspeintner, T.; van der Zant, T.; Iocchi, L.; Schiffer, S., RoboCup@Home: Scientific Competition and Benchmarking for Domestic Service Robots, Interaction Studies, Volume 10, Number 3, pp. 392-426(35) , 2009.

[122] Wojek C., Schiele B., "A performance evaluation of single and multi-feature people detection," in DAGM Symp. on Patt Rec, 2008, pp. 82-91.

[123] Wu S.-Q, Song W., Jiang L.-J., Xie S.-L., Pan F., Yau W.-Y., Ranganath S., “Infrared Face Recognition by Using Blood Perfusion Data”, Lecture Notes in Computer Science 3546, 2005.

[124] X. Geng, Z.-H. Zhou, “Image region selection and ensemble for face recognition”, J. Comput. Sci. Technol., Vol. 21, no. 1, pp. 116–125, 2006.

[125] X. Tan, S. Chen, Z.-H. Zhou, and F. Zhang, Face recognition from a single image per person: A survey, Pattern Recognition, Vol. 39, pp. 1725–1745, 2006.

[126] X. Tan, S. Chen, Z.-H. Zhou, and F. Zhang, Face recognition from a single image per person: A survey, Pattern Recognition, Vol. 39, pp. 1725–1745, 2006.

[127] Y. Freund, R.E. Schapire, A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting, Journal of Computer and System Sciences, Vol. 55, No. 1, 1997, pp. 119-139.

[128] Y. Su, S. Shan, X.Chen, and W. Gao, Patch-Based Gabor Fisher Classifier for Face Recognition, Proc. 18th Int. Conf. Pattern Recognition – ICPR 2006, pp. 528-531, 2006.

[129] Yueming Wang; Jianzhuang Liu; Xiaoou Tang; , "Robust 3D Face Recognition by Local Shape Difference Boosting," Pattern Analysis and Machine Intelligence, IEEE Transactions on , vol.32, no.10, pp.1858-1870, Oct. 2010.

[130] Zhu Q., Yeh M.-C., Cheng K.-T., Avidan S., "Fast human detection using a cascade of histograms of oriented gradients," in Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on, vol. 2, 2006, pp. 1491 - 1498.

Anexo A

En esta sección se incluyen los resultados completos presentados en la sección 5.6, para análisis de resultados por favor ver esa sección. Para ver descripción de los experimentos revisar sección 5.5.2.1.

1. Detección de rostros

Tabla 30: Resultados del experimento 1. Conjunto de parámetros 1.

Parámetros: Ruido en ángulo: $R_\theta \in [-5°, 5°]$, Ruido en desplazamiento: $R_P \in [-50\ cm, 50\ cm]$, Ruido Altura de personas: $R_H \in [-15\ cm, 15\ cm]$, Ruido en la trayectoria: $R_G \in [-3,3]$ y *Número de personas* =10

Parámetros: Ruido en ángulo = 5, Ruido en desplazamiento = 50 cm				
Rostros	**Personas**			
Número de rostros detectados / Total de rostros o *Ground Truth*	**Número de personas detectadas**	**Número de Personas descartadas**	**Número de personas correctamente detectadas**	**Falsos Positivos**
168 / 192	16	20	9	7
147 / 165	15	47	9	6
156 / 173	17	28	8	9
288 / 305	16	20	9	7
181 / 201	11	32	8	3
175 / 204	17	28	9	8
162 / 194	16	20	9	7
188 / 207	14	37	7	7
169 / 188	18	22	8	10
293 / 328	16	25	8	8
192.7 / 215.7	**15.6**	**27.9**	**8.4**	**7.2**

Tabla 31: Resultados del experimento 1. Conjunto de parámetros 2.

Parámetros: Ruido en ángulo: $R_\theta \in [-10°, 10°]$, Ruido en desplazamiento: $R_P \in [-100\ cm, 100\ cm]$, Ruido Altura de personas: $R_H \in [-15\ cm, 15\ cm]$, Ruido en la trayectoria:$R_G \in [-3,3]$ y *Número de personas* =10

Parámetros: Ruido en ángulo = 10, Ruido en desplazamiento = 100 cm				
Rostros	**Personas**			
Número de rostros detectados / Total de rostros o *Ground Truth*	**Número de personas detectadas**	**Número de Personas descartadas**	**Número de personas correctamente detectadas**	**Falsos Positivos**
365 / 381	13	13	8	5
256 / 281	18	35	7	11
237 / 258	13	15	5	8
241 / 254	16	17	6	10
192 / 206	14	28	6	8
255 / 263	17	18	7	10
267 / 288	12	37	5	7
244 / 259	14	22	7	7
294 / 302	13	19	7	6
259 / 267	16	18	6	10
261/275.9	**14.6**	**22.2**	**6.4**	**8.2**

2. **Detección de rostros, Reconocimiento de rostros, con construcción offline de DB reconocimiento**

Tabla 32: Resultados del experimento 2. Conjunto de parámetros 1.

Parámetros: Ruido en ángulo: $R_\theta \in [-5°, 5°]$, Ruido en desplazamiento: $R_P \in [-50\ cm, 50\ cm]$, Ruido Altura de personas: $R_H \in [-15\ cm, 15\ cm]$, Ruido en la trayectoria: $R_G \in [-3,3]$ y *Número de personas* =10

Parámetros: Ruido en ángulo = 5, Ruido en desplazamiento = 50 cm, Base de datos: 10 personas					
Número de rostros detectados / Total de rostros o *Ground Truth*	**Número de personas correctamente detectadas**	**Sin Mapa de Personas**		**Con Mapa de Personas**	
		Reconocimiento Correctas/Total	**Reconocimiento %**	**Reconocimiento Correctas/Total**	**Reconocimiento %**
168 / 192	9	114/149	76.51%	8/9	88.89%
147 / 165	9	100/132	75.76%	9/9	100.00%
156 / 173	8	103/122	84.43%	7/8	87.50%
288 / 305	9	229/255	89.80%	7/9	77.78%
181 / 201	8	112/161	69.57%	7/8	87.50%
175 / 204	9	124/152	81.58%	8/9	88.89%
162 / 194	9	108/141	76.60%	8/9	88.89%
188 / 207	7	102/139	73.38%	6/7	85.71%
169 / 188	8	107/148	72.30%	6/8	75.00%
293 / 328	8	207/246	84.15%	7/8	87.50%
192.7/215.7	**8.4**	**-----**	**78.41%**	**-----**	**86.77%**

Tabla 33: Resultados del experimento 2. Conjunto de parámetros 2.

Parámetros: Ruido en ángulo: $R_\theta \in [-10°, 10°]$, Ruido en desplazamiento: $R_P \in [-100\ cm, 100\ cm]$, Ruido Altura de personas: $R_H \in [-15\ cm, 15\ cm]$, Ruido en la trayectoria: $R_G \in [-3,3]$ y *Número de personas* =10

Parámetros: Ruido en ángulo = 10, Ruido en desplazamiento = 100 cm, Base de datos: 10 personas					
Número de rostros detectados / Total de rostros o *Ground Truth*	**Número de personas correctamente detectadas**	**Sin Mapa de Personas**		**Con Mapa de Personas**	
		Reconocimiento Correctas/Total	**Reconocimiento %**	**Reconocimiento Correctas/Total**	**Reconocimiento %**
365 / 381	8	234/330	70.91%	7/8	87.50%
256 / 281	7	166/227	73.13%	6/7	85.71%
237 / 258	5	166/195	85.13%	5/5	100.00%
241 / 254	6	159/202	78.71%	5/6	83.33%
192 / 206	6	167/207	80.68%	6/6	100.00%
255 / 263	7	173/209	82.78%	5/7	71.43%
267 / 288	5	191/234	81.62%	4/5	80.00%
244 / 259	7	174/204	85.29%	6/7	85.71%
294 / 302	7	201/268	75.00%	7/7	100.00%
259 / 267	6	185/237	78.06%	5/6	83.33%
261/275.9	**6.4**	**-----**	**79.13%**	**-----**	**87.70%**

3. **Detección de rostros, Reconocimiento de rostros, con construcción online de DB reconocimiento**

Tabla 34: Resultados del experimento 3. Conjunto de parámetros 1.

Parámetros: Ruido en ángulo: $R_\theta \in [-5°, 5°]$, Ruido en desplazamiento: $R_P \in [-50\ cm, 50\ cm]$, Ruido Altura de personas: $R_H \in [-15\ cm, 15\ cm]$, Ruido en la trayectoria: $R_G \in [-3,3]$ y *Número de personas* =10

Parámetros: Ruido en ángulo = 5, Ruido en desplazamiento = 50 cm				
Número de rostros detectados / Total de rostros o *Ground Truth*	**Número de personas Almacenadas en DB**	**Número de personas correctamente detectadas**	**Reconocimiento Correctas/Total**	**Reconocimiento %**
168 / 192	14	9	7/9	77.78%
147 / 165	15	9	6/9	66.67%
156 / 173	13	8	6/8	75.00%
288 / 305	17	9	6/9	66.67%
181 / 201	14	8	5/8	62.50%
175 / 204	16	9	7/9	77.78%
162 / 194	15	9	6/9	66.67%
188 / 207	13	7	5/7	71.43%
169 / 188	17	8	5/8	62.50%
293 / 328	14	8	6/8	75.00%
192.7/215.7	**14.8**	**8.4**	**----**	**70.20%**

Tabla 35: Resultados del experimento 3. Conjunto de parámetros 2.

Parámetros: Ruido en ángulo: $R_\theta \in [-10°, 10°]$, Ruido en desplazamiento: $R_P \in [-100\ cm, 100\ cm]$, Ruido Altura de personas: $R_H \in [-15\ cm, 15\ cm]$, Ruido en la trayectoria: $R_G \in [-3,3]$ y *Número de personas* =10

Parámetros: Ruido en ángulo = 10, Ruido en desplazamiento = 100 cm				
Número de rostros detectados / Total de rostros o *Ground Truth*	**Número de personas Almacenadas en DB**	**Número de personas correctamente detectadas**	**Reconocimiento Correctas/Total**	**Reconocimiento %**
365 / 381	15	8	6/8	75.00%
256 / 281	16	7	5/7	71.43%
237 / 258	14	5	4/5	80.00%
241 / 254	16	6	4/6	66.67%
192 / 206	12	6	4/6	66.67%
255 / 263	16	7	5/7	71.43%
267 / 288	17	5	4/5	80.00%
244 / 259	15	7	5/7	71.43%
294 / 302	14	7	6/7	85.71%
259 / 267	13	6	4/6	66.67%
261/275.9	**14.8**	**6.4**	**----**	**73.50%**

4. **Detección de rostros, Reconocimiento de rostros y Visión activa, con construcción offline de DB reconocimiento**

Tabla 36: Resultados del experimento 4. Conjunto de parámetros 1.

Parámetros: Ruido en ángulo: $R_\theta \in [-5°, 5°]$, Ruido en desplazamiento: $R_P \in [-50\ cm, 50\ cm]$, Ruido Altura de personas: $R_H \in [-15\ cm, 15\ cm]$, Ruido en la trayectoria: $R_G \in [-3,3]$ y *Número de personas* =10

Parámetros: Ruido en ángulo = 5, Ruido en desplazamiento = 50 cm			
Número de rostros detectados / Total de rostros o *Ground Truth*	**Número de personas correctamente detectadas**	**Reconocimiento Correctas/Total**	**Reconocimiento %**
168 / 192	9	8/9	88.89%
147 / 165	9	9/9	100.00%
156 / 173	8	8/8	100.00%
288 / 305	9	8/9	88.89%
181 / 201	8	7/8	87.50%
175 / 204	9	8/9	88.89%
162 / 194	9	9/9	100.00%
188 / 207	7	7/7	100.00%
169 / 188	8	7/8	87.50%
293 / 328	8	7/8	87.50%
192.7/215.7	**8.4**	**----**	**92.92%**

Tabla 37: Resultados del experimento 4. Conjunto de parámetros 2.

Parámetros: Ruido en ángulo: $R_\theta \in [-10°, 10°]$, Ruido en desplazamiento: $R_P \in [-100\ cm, 100\ cm]$, Ruido Altura de personas: $R_H \in [-15\ cm, 15\ cm]$, Ruido en la trayectoria: $R_G \in [-3,3]$ y *Número de personas* =10

Parámetros: Ruido en ángulo = 10, Ruido en desplazamiento = 100 cm			
Número de rostros detectados / Total de rostros o *Ground Truth*	**Número de personas correctamente detectadas**	**Reconocimiento Correctas/Total**	**Reconocimiento %**
365 / 381	8	7/8	87.50%
256 / 281	7	6/7	85.71%
237 / 258	5	5/5	100.00%
241 / 254	6	6/6	100.00%
192 / 206	6	6/6	100.00%
255 / 263	7	6/7	85.71%
267 / 288	5	5/5	100.00%
244 / 259	7	6/7	85.71%
294 / 302	7	7/7	100.00%
259 / 267	6	5/6	83.33%
261/275.9	**6.4**	**----**	**92.80%**

5. **Detección de rostros, Reconocimiento de rostros y Visión activa, con construcción online de DB reconocimiento**

Tabla 38: Resultados del experimento 5. Conjunto de parámetros 1.

Parámetros: Ruido en ángulo: $R_\theta \in [-5°, 5°]$, Ruido en desplazamiento: $R_P \in [-50\ cm, 50\ cm]$, Ruido Altura de personas: $R_H \in [-15\ cm, 15\ cm]$, Ruido en la trayectoria: $R_G \in [-3,3]$ y *Número de personas* =10

Parámetros: Ruido en ángulo = 5, Ruido en desplazamiento = 50 cm				
Número de rostros detectados / Total de rostros o *Ground Truth*	**Número de personas Almacenadas en DB**	**Número de personas correctamente detectadas**	**Reconocimiento Correctas/Total**	**Reconocimiento %**
168 / 192	12	9	8/9	88.89%
147 / 165	13	9	7/9	77.78%
156 / 173	12	8	8/8	100.00%
288 / 305	11	9	8/9	88.89%
181 / 201	13	8	7/8	87.50%
175 / 204	12	9	9/9	100.00%
162 / 194	13	9	7/9	77.78%
188 / 207	11	7	6/7	85.71%
169 / 188	12	8	6/8	75.00%
293 / 328	13	8	7/8	87.50%
192.7/215.7	**12.2**	**8.4**	**----**	**86.90%**

Tabla 39: Resultados del experimento 5. Conjunto de parámetros 2.

Parámetros: Ruido en ángulo: $R_\theta \in [-10°, 10°]$, Ruido en desplazamiento: $R_P \in [-100\ cm, 100\ cm]$, Ruido Altura de personas: $R_H \in [-15\ cm, 15\ cm]$, Ruido en la trayectoria: $R_G \in [-3,3]$ y *Número de personas* =10

Parámetros: Ruido en ángulo = 10, Ruido en desplazamiento = 100 cm				
Número de rostros detectados / Total de rostros o *Ground Truth*	**Número de personas Almacenadas en DB**	**Número de personas correctamente detectadas**	**Reconocimiento Correctas/Total**	**Reconocimiento %**
365 / 381	13	8	7/8	87.50%
256 / 281	12	7	6/7	85.71%
237 / 258	14	5	5/5	100.00%
241 / 254	11	6	5/6	83.33%
192 / 206	12	6	5/6	83.33%
255 / 263	13	7	6/7	85.71%
267 / 288	13	5	5/5	100.00%
244 / 259	14	7	6/7	85.71%
294 / 302	12	7	7/7	100.00%
259 / 267	14	6	5/6	83.33%
261/275.9	**12.8**	**6.4**	**----**	**89.46%**

6. **Reconocimiento de rostros y Visión activa, con construcción offline de DB reconocimiento**

Tabla 40: Resultados del experimento 6. Conjunto de parámetros 1.

Parámetros: Ruido en ángulo: $R_\theta \in [-5°, 5°]$, Ruido en desplazamiento: $R_P \in [-50\ cm, 50\ cm]$, Ruido Altura de personas: $R_H \in [-15\ cm, 15\ cm]$, Ruido en la trayectoria: $R_G \in [-3,3]$ y *Número de personas* =10

Parámetros: Ruido en ángulo = 5, Ruido en desplazamiento = 50 cm		
Número de personas correctamente detectadas	**Reconocimiento Correctas/Total**	**Reconocimiento %**
10	9/10	90.00%
10	10/10	100.00%
10	8/10	80.00%
10	9/10	90.00%
10	9/10	90.00%
10	10/10	100.00%
10	9/10	90.00%
10	9/10	90.00%
10	10/10	100.00%
10	9/10	90.00%
10.0	**----**	**92.00%**

Tabla 41: Resultados del experimento 6. Conjunto de parámetros 2.

Parámetros: Ruido en ángulo: $R_\theta \in [-10°, 10°]$, Ruido en desplazamiento: $R_P \in [-100\ cm, 100\ cm]$, Ruido Altura de personas: $R_H \in [-15\ cm, 15\ cm]$, Ruido en la trayectoria: $R_G \in [-3,3]$ y *Número de personas* = 10

Parámetros: Ruido en ángulo = 10, Ruido en desplazamiento = 100 cm		
Número de personas correctamente detectadas	**Reconocimiento Correctas/Total**	**Reconocimiento %**
10	10/10	100.00%
10	9/10	90.00%
10	8/10	80.00%
10	9/10	90.00%
10	9/10	90.00%
10	10/10	100.00%
10	8/10	80.00%
10	9/10	90.00%
10	8/10	80.00%
10	9/10	90.00%
10.0	**----**	**89.00%**

7. **Reconocimiento de rostros y Visión activa, con construcción online de DB reconocimiento**

Tabla 42: Resultados del experimento 7. Conjunto de parámetros 1.

Parámetros: Ruido en ángulo: $R_\theta \in [-5°, 5°]$, Ruido en desplazamiento: $R_P \in [-50\ cm, 50\ cm]$, Ruido Altura de personas: $R_H \in [-15\ cm, 15\ cm]$, Ruido en la trayectoria: $R_G \in [-3,3]$ y *Número de personas* =10

Parámetros: Ruido en ángulo = 5, Ruido en desplazamiento = 50 cm			
Número de personas Almacenadas en DB	**Número de personas correctamente detectadas**	**Reconocimiento Correctas/Total**	**Reconocimiento %**
10	10	9/10	90.00%
10	10	10/10	100.00%
10	10	8/10	80.00%
10	10	9/10	90.00%
10	10	9/10	90.00%
10	10	9/10	90.00%
10	10	9/10	90.00%
10	10	8/10	80.00%
10	10	10/10	100.00%
10	10	9/10	90.00%
10.0	**10.0**	**----**	**90.00%**

Tabla 43: Resultados del experimento 7. Conjunto de parámetros 2.

Parámetros: Ruido en ángulo: $R_\theta \in [-10°, 10°]$, Ruido en desplazamiento: $R_P \in [-100\ cm, 100\ cm]$, Ruido Altura de personas: $R_H \in [-15\ cm, 15\ cm]$, Ruido en la trayectoria: $R_G \in [-3,3]$ y *Número de personas* =10

Parámetros: Ruido en ángulo = 10, Ruido en desplazamiento = 100 cm			
Número de personas Almacenadas en DB	**Número de personas correctamente detectadas**	**Reconocimiento Correctas/Total**	**Reconocimiento %**
10	10	9/10	90.00%
10	10	8/10	80.00%
10	10	8/10	80.00%
10	10	9/10	90.00%
10	10	9/10	90.00%
10	10	10/10	100.00%
10	10	8/10	80.00%
10	10	8/10	80.00%
10	10	8/10	80.00%
10	10	9/10	90.00%
10.0	**10.0**	**----**	**86.00%**

8. **Detección de Rostros, reconocimiento de rostros. Con y sin visión activa, con construcción offline de DB reconocimiento**

Tabla 44: Resultados del experimento 8. Conjunto de parámetros 1.

Parámetros: Distancia Base del sujeto = 200 cm., Ruido en desplazamiento: $R_P \in [-50\ cm, 50\ cm]$, Ruido Altura de personas: $R_H \in [-15\ cm, 15\ cm]$ y *Número de personas* = 20

Sin Visión Activa			**Con Visión Activa**		
Número de personas correctamente detectadas	**Reconocimiento Correctas/Total**	**Reconocimiento %**	**Número de personas correctamente detectadas**	**Reconocimiento Correctas/Total**	**Reconocimiento %**
17	15/17	88.24%	17	16/17	94.12%
17	15/17	88.24%	17	15/17	88.24%
17	16/17	94.12%	17	17/17	100.00%
18	17/18	94.44%	18	17/18	94.44%
20	18/20	90.00%	20	18/20	90.00%
20	18/20	90.00%	20	19/20	95.00%
18	15/18	83.33%	18	17/18	94.44%
19	17/19	89.47%	19	18/19	94.74%
18	16/18	88.89%	18	17/18	94.44%
17	15/17	88.24%	17	16/17	94.12%
17	16/17	94.12%	17	17/17	100.00%
18	15/18	83.33%	18	18/18	100.00%
18	18/18	100.00%	18	17/18	100.00%
19	17/19	89.47%	19	17/19	89.47%
20	18/20	90.00%	20	20/20	100.00%
18	16/18	88.89%	18	17/18	94.44%
20	18/20	90.00%	20	19/20	95.00%
19	18/19	94.74%	19	19/19	100.00%
18	17/18	94.44%	18	18/18	100.00%
19	17/19	89.47%	19	18/19	94.74%
18.35 / 20	**----**	**90.47%**	**18.35 / 20**	**----**	**95.66%**

Tabla 45: Resultados del experimento 8. Conjunto de parámetros 2.

Parámetros: Distancia Base del sujeto = 200 cm., Ruido en desplazamiento: $R_P \in [-100\ cm, 100\ cm]$, Ruido Altura de personas: $R_H \in [-15\ cm, 15\ cm]$, y *Número de personas* = 20

Sin Visión Activa			**Con Visión Activa**		
Número de personas correctamente detectadas	**Reconocimiento Correctas/Total**	**Reconocimiento %**	**Número de personas correctamente detectadas**	**Reconocimiento Correctas/Total**	**Reconocimiento %**
18	14/18	77.78%	18	18/18	100.00%
18	16/18	88.89%	18	17/18	94.44%
18	18/18	100.00%	18	18/18	100.00%
16	15/16	93.75%	16	16/16	100.00%
15	13/15	86.67%	15	14/15	93.33%
17	15/17	88.24%	17	15/17	88.24%
19	18/19	94.74%	19	18/19	94.74%
19	17/19	89.47%	19	19/19	100.00%
16	15/16	93.75%	16	16/16	100.00%
17	15/17	88.24%	17	16/17	94.12%
20	18/20	90.00%	20	19/20	95.00%
15	13/15	86.67%	15	14/15	93.33%
18	16/18	88.89%	18	17/18	94.44%
16	14/16	87.50%	16	15/16	93.75%
16	15/16	93.75%	16	15/16	93.75%
16	14/16	87.50%	16	15/16	93.75%
18	16/18	88.89%	18	18/18	100.00%
17	16/17	94.12%	17	16/17	94.12%
19	16/19	84.21%	19	18/19	94.74%
15	14/15	93.33%	15	15/15	100.00%
17.15 / 20	**----**	**89.82%**	**17.15 / 20**	**----**	**95.89%**