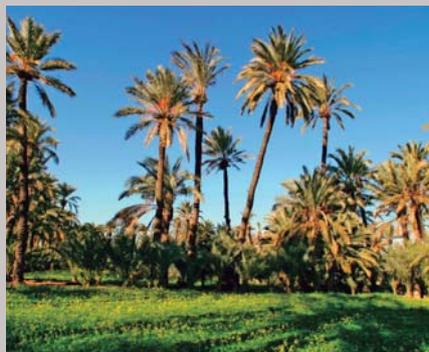
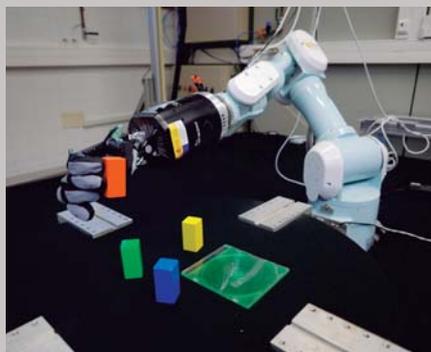
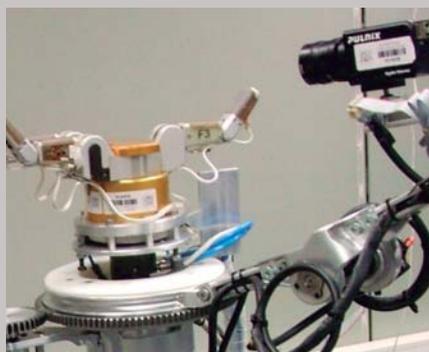


JNR2019

Jornadas Nacionales de Robótica
Spanish National Robotics Conference

LIBRO DE ACTAS

Alicante, SPAIN - June 13th-14th, 2019



JNR2019

Jornadas Nacionales de Robótica **Spanish Robotics Conference** **2019**

- Libro de Actas -

Alicante, 13-14 de junio de 2019

Organizado por:

Universidad de Alicante

Universidad Miguel Hernández de Elche

Comité Español de Automática - Grupo Temático de Robótica



Editores: Fernando Torres Medina · Óscar Reinoso García

ISBN: 978-84-09-12133-5

PREFACIO

A través de estas páginas, queremos dar la bienvenida a todos los asistentes de las Jornadas Nacionales de Robótica de 2019. Son varias ya las ediciones que se vienen realizando de este evento lo que sin duda afianza el carácter de este que poco a poco se consolida como un referente donde investigadores y universitarios relacionados con el mundo de la robótica tienen un punto de encuentro en el que discutir temas relacionados con esta temática. Como organizadores de este evento, durante varios meses hemos intentado que la organización de las diferentes actividades posibilite un marco donde todos los asistentes puedan aprovechar todas y cada una de las actuaciones programadas. A diferencia de otras ediciones, en esta anualidad se ha querido incrementar el tipo de resultados que se presentan en la misma, facilitando en todo momento la difusión de resultados y actividades realizadas en los últimos meses por los diferentes miembros y grupos de investigación. En definitiva, queremos que estas Jornadas propicien un punto de encuentro para todos los investigadores donde compartir avances, resultados y trabajos futuros de la robótica en España. En conjunto, y gracias a todas las actividades propuestas, se presentarán durante las jornadas un total de 36 comunicaciones. Un número muy elevado de trabajos que sin duda alguna muestran el alto interés de las jornadas entre toda la comunidad científica. Estas comunicaciones se organizan en dos sesiones de comunicaciones, en las que cada uno de los trabajos contarán con 5 minutos de exposición y un turno final de discusión.

Una de las señas de identidad de las Jornadas Nacionales de Robótica en sus distintas ediciones consiste en contar con prestigiosos investigadores europeos. En esta ocasión se cuenta con dos conferenciantes de prestigio. En la primera de estas sesiones plenarias organizadas, tendrá lugar la conferencia “Is this my body? Neuro-inspired robotic body perception and action”, impartida por Pablo Lanillos, del Institute of Cognitive Systems (ICS) perteneciente a la Technical University of Munich (TUM). En la segunda de las sesiones plenarias, se impartirá la conferencia “Robot vision for the perception of Objects” a cargo de Markus Vincze del Vision for Robotics (V4R) dentro del Automation and Control Institute (ACIN) de la Technical University of Wien (TUW). Sin duda alguna estos prestigiosos conferenciantes atraerán el interés de todos los asistentes.

Además de las comunicaciones presentadas por parte de los asistentes a las Jornadas, así como de las conferencias invitadas, se quiere contar en todo momento con las principales novedades de las empresas relacionadas con la robótica en España. En este sentido, dentro de una de las actividades se ha programado una sesión en la que diferentes empresas (Universal Robots, CFZ Cobots, Robotnik Automation SLL y ABB) presentarán algunos de los últimos desarrollos relacionados con la robótica colaborativa y la industria 4.0.

Además de estas actividades, tendrá lugar una reunión informativa en la que algunas de las entidades y sociedades españolas relacionadas con la robótica mostrarán las diferentes actuaciones llevadas a cabo por estas (EuRobotics, IROS, European Robotics Forum, IEEE Student Branch, ...).

Dentro del programa planificado se integran además una comida de trabajo, 2 pausas para café y una cena en la que todos los asistentes podrán continuar compartiendo las experiencias, tendencias y trabajos desarrollados en las diferentes actividades.

Por último, quisiéramos agradecer el apoyo recibido para la organización y celebración de estas Jornadas. En primer lugar, a las dos universidades que organizan y soportan las mismas, la Universidad de Alicante y la Universidad Miguel Hernández de Elche, y en especial a los miembros de ambas universidades que han participado muy activamente en su organización. A la Red Nacional de Robótica por su inestimable colaboración y ayuda en la realización de algunas de las actividades propuestas, así como en el soporte prestado. A la sección del IEEE RAS Spanish Chapter, que ha posibilitado que uno de los conferenciantes invitados pueda acudir a las Jornadas. Como patrocinadores se cuenta con el apoyo de diferentes empresas relacionadas con la robótica (ABB, Robotnik y CFZ Cobots). Desde el primer momento que tuvieron conocimiento de estas, se han prestado en todo momento a apoyar las mismas. Finalmente, y quizás más importante a todos los asistentes a las Jornadas, ya que, sin su presencia y actividad, estas carecerían de sentido.

Fernando Torres Medina
Universidad de Alicante

Oscar Reinoso García
Universidad Miguel Hernández de Elche

COMITÉS

Comité organizador

El comité organizador está formado por los siguientes profesores e investigadores de los grupos de investigación Automática, Robótica y visión Artificial (AUROVA) de la Universidad de Alicante (UA), y Automatización, Robótica y Visión por Computador (ARVC) de la Universidad Miguel Hernández (UMH).

Coordinadores:

Oscar Reinoso García (UMH-ARVC).
Fernando Torres Medina (UA-AUROVA).

Miembros:

Mónica Ballesta Galdeano (UMH-ARVC).
Francisco A. Candelas Herías (UA-AUROVA).
Arturo Gil Aparicio (UMH-ARVC).
Pablo Gil Vázquez (UA-AUROVA).
Luis M. Jiménez García (UMH-ARVC).
Luis Payá Castelló (UMH-ARVC).
Adrián Peidró Vidal (UMH-ARVC).
Santiago T. Puente Méndez (UA-AUROVA).
David Valiente García (UMH-ARVC).

Comité científico

Juan Andrade (IRI, CSIC-UPC).
Carlos Balaguer (RoboCity2030).
Antonio Barrientos (UPM).
Enrique Bernabeu (UPV).
Dolores Blanco (UC3M).
Itziar Cabanes (EHU).
Marc Carreras (UDG).
Manuel Ferre (CAR-CSIC).
Cecilia García (CAR-UPM).
Antonio Giménez (CEA-GTRob).
Fernando Gómez (UHU).
Martin Mellado (UPV).
Alberto Ortiz (UIB).
Raul Suarez (IEEE-RAS Sp.Chap.).
Verónica Sáiz (UPV).
Miguel Angel Salichs (HispaRob).
Antonio Sánchez (UPV).
Rafael Sanz (UVIGO).
Pedro Sanz (UJI).
José Andrés Somolinos (AutoMar).
Alberto Vale (IST-Lisbon).
Eduardo Zalama (UVA).

COLABORADORES

Entidades colaboradoras



Universitat d'Alacant
Universidad de Alicante



UNIVERSITAS
Miguel Hernández



RED NACIONAL
DE ROBÓTICA
(DPI2017-90853-REDT)



Empresas colaboradoras



Índice de contenidos

Estudio de descriptores holísticos basados en métodos analíticos y técnicas de deep learning para localización con robots móviles	1
S. Cebollada, L. Payá, A. Peidró, L.M. Jiménez, O. Reinoso	
Avances en BLUE: Robot para la Localización en Entornos No Estructurados	9
Miguel Á. Muñoz-Bañón, Iván del Pino, Pablo Gil, Francisco A. Candelas, Fernando Torres	
Robot de cables para operación en superficies verticales	15
Raúl Gómez Ramos, Eduardo Zalama, Jaime Gómez-García-Bermejo	
Desarrollo y validación de algoritmos de control de una órtesis activa de rodilla para sujetos hemiparéticos basados en la cinemática de la pierna sana	21
Julio Salvador Lora Millán, Eduardo Rocon	
Seguimiento de las esquinas de ojos como primer paso en la estimación de movimientos de cabeza para aplicaciones clínicas	29
Agostina J. Larrazabal, César E. Martínez, Cecilia E. García Cena	
Inferencia Activa para la Percepción y la Acción de Robots Humanoides	36
Guillermo Oliver, Pablo Lanillos, Gordon Cheng	
Análisis de las reacciones psicofisiológicas de los usuarios en terapias de rehabilitación para un jugador y competitivas asistidas por dispositivos robóticos	43
J.M. Catalán, A. Blanco, J.A. Díez, J.V. García, R. Puerto, N. García-Aracil	
Diseño de un exoesqueleto de brazo para la rehabilitación del hombro controlado por una interfaz cerebro-máquina (BCI)	48
J. Belmar, E. Iáñez, M. Ortiz, J.M. Azorín	
Algoritmo de Reconocimiento de Gestos Dinámico para Cirugía Endoscópica Endonasal . .	55
E. Bauzano, M.C. López-Casado, I. Rivas-Blanco, C.J. Pérez-del-Pulgar, V.F. Muñoz	
Controlling Robot Motion by Blinking Eyes: an Experience on Users Training	63
Fernando Gómez-Bravo, Rafael López de Ahumada, Juan A. Castro-García, Raúl Jiménez-Naharro, Alberto J. Molina-Cantero, Juan A. Gómez-Galán, Santiago Berrazueta-Alvarado, Manuel Sanchez-Raya	
Improvement of the Sensory and Autonomous Capability of Robots Through Olfaction: the IRO Project.	71
Javier Monroy, Jose-Raul Ruiz-Sarmiento, Francisco-Angel Moreno, Cipriano Galindo, Javier Gonzalez-Jimenez	
Localización Visual de un Robot Submarino de bajo coste para la inspección de Minas Inundadas	79
Olaya Álvarez-Tuñón, Alberto Jardón, Carlos Balaguer	

Sistema de Navegación y Exploración Autónoma para robots de Búsqueda y Rescate en Incidentes en Minas	85
N. Pérez, A. Jardón, A. Rodríguez, C. Balaguer	
El robot social Mini como plataforma de información y ocio	92
Sergio González-Díaz, Elena Velázquez-Navarro, Fernando Alonso-Martín, Álvaro Castro-González, José Carlos Castillo, María Malfaz, Miguel A. Salichs	
Sensorización de un sistema de eslabón blando actuando como cuello robótico	98
Lisbeth Mena, Concepción A. Monje, Luis Nagua, Jorge Muñoz, Carlos Balaguer	
Banco de pruebas para actuadores basados en aleaciones con memoria de forma	103
Janeth Arias Guadalupe, Dorin Copaci, David Serrano del Cerro, Luis Moreno, Dolores Blanco	
Detección Inteligente y Manipulación Robótica Bimanual para la Cosecha Selectiva de Cultivos de Alto Valor	110
R. Fernández, D. Sepúlveda, E. Navas, M. Armada, P. Gonzalez-de-Santos	
Trajectory Planning for MASAR: a New Modular and Single-Actuator Robot	118
Adrián Peidró, Julio Gallego, Luis Payá, José María Marín, Óscar Reinoso	
Banco de pruebas instrumentado para robots móviles terrestres	126
Jorge De León, José Ángel López, Jaime Del Cerro, Antonio Barrientos	
Algoritmo de Control Basado en Modelo Cinemático y Dinámico de un Dispositivo Robótico para Terapias de Rehabilitación	134
Aitziber Mancisidor, Itziar Cabanes, Asier Zubizarreta, Asier Brull, Eva Portillo	
The Role of HRI within COMOMUIS Research Project	141
Pedro J Sanz, Marcos de la Cruz, Giacomo Lunghi, Carlos Veiga, Raúl Marín, Mario Di Castro	
Proyecto AGRICOBOT: Robot Colaborativo para Transporte Inteligente en Interior de Invernaderos con Soporte en IoT	148
J. C. Moreno, A. Giménez, F. Rodríguez	
Diseño, Desarrollo y Validación de un Sistema Robotizado de Micro-inspección para la Experimentación con <i>C. elegans</i>	155
Jose F. González, Antonio J. Sánchez	
Twisted coiled nylon-based actuators for robotic applications	163
Erika Bengtsdotter, Dorin Copaci, David Serrano del Cerro, Dolores Blanco, Luis Moreno	
Teleoperación de manos robóticas mediante virtualización en Unity 3D para dispositivos móviles	169
S.T. Puente, L. Más, F. Torres, F.A. Candelas	
Desincronización Beta Bihemisférica durante una tarea motora en pacientes crónicos después de un accidente cerebrovascular crónico	175
Santiago Ezquerro, Juan A. Barios, Arturo Bertomeu-Motos, Jorge Diez, Jose M. Sanchez-Aparicio, Luis Donis-Barber, Ramón Ñeco, Eduardo Fernandez, N. Garcia-Aracil	

Desarrollo de Algoritmos de Reconstrucción Cinemática del Brazo para Terapias de Neurorehabilitación Asistidas por Robots de Efecto Final	181
Arturo Bertomeu-Motos, Nicolas Garcia-Aracil	
Towards the Automation of Visual Inspections of Cargo Holds of Large-Tonnage Vessels	189
Alberto Ortiz, Francisco Bonnin-Pascual, Emilio Garcia-Fidalgo, Joan P. Company-Corcoles, Kai Yao	
Sistema Multi-Robot para Cooperación con Equipos de Rescate de Primera Respuesta Humanos y Caninos en Escenarios de Catastrofe	197
Alfonso J García-Cerezo, Anthony Mandow, Jose Antonio Gómez Ruiz, Juan Jesús Fernández Lozano, Jesús M. Gómez de Gabriel, Antonio Jesús Reina Terol, Jorge Luis Martínez Rodríguez, Jesús Morales Rodríguez, María Alcázar Martínez Sánchez, Jesús Miranda Páez, Ricardo Vázquez Marín, Ana Cruz Martín, Javier Serón Barba, Francisco Pastor Martín, Victoria Plaza Leiva, Juan Manuel Gandarias, Manuel Toscano Moreno	
Metodología para el diseño y puesta en funcionamiento de sistemas de manipulación remota en instalaciones con radiación	206
Manuel Ferre Pérez, Sofía Coloma Chacón, Mario di Castro, Miguel A. Sánchez-Urán González	
El robot social Mini como plataforma para el desarrollo de juegos de interacción multimodales	214
Elena Velázquez-Navarro, Sergio González-Díaz, Fernando Alonso-Martín, José Carlos Castillo, Álvaro Castro-González, María Malfaz, Miguel A. Salichs	
Manipulación Diestra de Objetos Desconocidos Usando Puntos de Contacto Virtuales	221
Andrés Montaña, Raúl Suárez	
Visualización 3D de las deformaciones geométricas del cerebro en tiempo real utilizando realidad aumentada	229
Natividad Bermejo-Herrero, Karin Correa-Arana, Ernesto Ávila Navarro, José María Sabater-Navarro	
Detección y seguimiento de personas utilizando sensores LIDAR	234
Claudia Álvarez-Aparicio, Ángel Manuel Guerrero-Higueras, Francisco Javier Rodríguez-Lera, Jonatan Ginés Clavero, Francisco Martín Rico, Vicente Matellán	
Entorno de Vida Asistida Inteligente: Diseño de una arquitectura para la Persona Mayor con la Inclusión de Robots	241
David Loza M., Alexandra Verdugo C., Eduardo Zalama C., Jaime Gómez-García-Bermejo	
Predicción de la Estabilidad en Tareas de Agarre Robótico con Información Táctil	249
Brayan S. Zapata-Impata, Pablo Gil, Fernando Torres	

Estudio de descriptores holísticos basados en métodos analíticos y técnicas de deep learning para localización con robots móviles

S. Cebollada^{a,*}, L. Payá^a, A. Peidró^a, L.M. Jiménez^a, O. Reinoso^a

^aDepartamento de Ingeniería de Sistemas y Automática, Universidad Miguel Hernández de Elche, Av. de la Universidad s/n, 03202, Elche, España.

Resumen

Este trabajo evalúa diversos descriptores holísticos para realizar la tarea de localización, la cual resulta de vital importancia en la robótica móvil. Durante este proceso, la única fuente de información utilizada es una cámara omnidireccional. La información visual recogida por la cámara es procesada para obtener descriptores holísticos. Tras esto, la posición del robot es estimada mediante la comparación entre los descriptores contenidos en el modelo visual y el descriptor calculado durante el proceso de localización. Los descriptores evaluados son (1) los basados en métodos analíticos (HOG y *gist*) y (2) los basados en técnicas de *deep learning* (auto-encoders y Redes Neuronales Convolucionales). La evaluación del proceso de localización se lleva a cabo a través de un dataset de imágenes panorámicas. Dicho dataset proporciona imágenes capturadas en entornos de interior bajo condiciones reales de uso. Los resultados obtenidos muestran que los descriptores basados en *deep learning* también pueden proporcionar soluciones interesantes para resolver la tarea localización.

Palabras Clave:

Robótica Móvil, Imágenes Omnidireccionales, Descriptores Holísticos, Localización, Deep Learning.

1. Introducción

Hoy en día, está muy extendido el uso de información visual para solventar las tareas de robótica móvil. Para estas situaciones, el robot debe de ser capaz de construir un mapa dentro del entorno en el que se encuentra y posteriormente estimar su posición dentro de ese entorno. Estas tareas son comúnmente conocidas como mapeo (o *mapping*) y localización. Entre los diferentes sensores usados en robótica móvil, el uso de cámaras omnidireccionales presenta una solución muy interesante, dado que son capaces de aportar información que cubre un campo de visión de 360 grados.

Los descriptores holísticos (o descriptores de apariencia global) han sido utilizados por varios autores para extraer información característica de las imágenes y así poder solventar las tareas de mapeo y localización. Por ejemplo, Rituerto *et al.* [Rituerto et al. \(2014\)](#) proponen el uso del descriptor *gist* [Leonardis y Bischof \(2000\)](#); [Oliva y Torralba \(2001\)](#) para crear mapas topológicos a partir de imágenes omnidireccionales. [Berenquer et al.](#) [Berenquer et al. \(2015\)](#) proponen la transformada de Radon [Radon \(1917\)](#) como descriptor holístico para desarrollar métodos de localización jerárquica. [Faessler et al.](#) [Faessler et al. \(2016\)](#) presentan un sistema basado en visión para mapear un espacio tridimensional para el uso de cuadricópteros con el objetivo de eliminar el retardo temporal entre el cuadricóptero

y los sistemas externos. [Korrapati y Mezouar](#) [Korrapati y Mezouar \(2017\)](#) introducen el uso de imágenes omnidireccionales mediante descriptores holísticos para crear mapas topológicos y también para detección de cierre de bucle. Más recientemente, [Román et al.](#) [Román et al. \(2018\)](#) evalúan el uso de descriptores holísticos para resolver la tarea de localización ante cambios de iluminación. [Cebollada et al.](#) [Cebollada et al. \(2019\)](#) evalúan el uso de descriptores holísticos para construir mapas jerárquicos a través de algoritmos de clustering y después solventar la localización a través de dichos mapas. Este trabajo, también evalúa diferentes medidas de distancia para obtener una medida de similitud entre imágenes que se asemeje lo máximo posible a la distancia geométrica entre las posiciones de captura de las imágenes correspondientes.

El cálculo de estos descriptores esta basado en métodos analíticos, sin embargo, durante los últimos años, algunos autores han propuesto el uso de técnicas de *deep learning* para crear descriptores holísticos. Por ejemplo, [Xu et al.](#) [Xu et al. \(2016\)](#) proponen el uso de auto-encoders para detectar imágenes histopatológicas de cáncer de mama. Por otro lado, [Xu et al.](#) [Xu et al. \(2019\)](#) usan descriptores basados en CNN (*Convolutional Neural Networks*) para obtener la posición más probable en un mapa de interior a través del algoritmo de Monte Carlo y también resolver el problema del robot secuestrado"; [Payá et al.](#) [Payá et al. \(2018\)](#) proponen también el uso de descriptores holísticos basados en CNN pero en este caso para realizar mapeo jerárquico. Ambos trabajos se sustentan en el uso de la red *places* [Zhou et al. \(2014\)](#). Los descriptores extraídos de dicha red corresponden a los vectores calculados en ciertas capas completamente convolucionales (*fully convolutional layers*).

A través de este trabajo, se lleva a cabo una comparación

*Autor en correspondencia.

Correos electrónicos: sergio.cebollada@umh.es (S. Cebollada), lpaya@umh.es (L. Payá), apeidro@umh.es (A. Peidró), luis.jimenez@umh.es (L.M. Jiménez), o.reinoso@umh.es (O. Reinoso)

entre descriptores holísticos basados en métodos analíticos y métodos basados en *deep learning* para resolver la tarea de localización. La bondad de estos métodos es medida de acuerdo a la precisión (error de localización) y el tiempo de cómputo (tiempo en calcular el descriptor y tiempo requerido para estimar la posición del robot).

El resto de artículo se estructura de la siguiente manera: Sección 2 explica el algoritmo usado para estimar la posición del robot dentro del escenario. Después, en la sección 3, se presentan los descriptores holísticos que han sido evaluados en este trabajo. La sección 4 explica y presenta los datos utilizados así como los experimentos realizados y la discusión de los mismos. Finalmente, en la sección 5 se detallan las conclusiones alcanzadas y las futuras líneas de investigación que se originarán tras este estudio.

2. Localización

La tarea de localización que se utiliza en este trabajo es la basada en el problema conocido como *image retrieval*, el cual consiste en seleccionar de una base de datos la imagen que mayor similitud presenta con respecto a una nueva imagen. Para llevar a cabo este propósito, el robot habrá obtenido previamente información visual del entorno, es decir, calcular descriptores holísticos partiendo de las $N_{Entreno}$ imágenes capturadas desde distintas posiciones del entorno. Esta tarea es conocida como mapeo o *mapping* y es imprescindible realizar este paso antes de poder empezar la tarea de localización. Después del *mapping*, los pasos para llevar a cabo la tarea de localización son los siguientes:

- El robot captura una nueva imagen omnidireccional desde una posición desconocida.
- Dicha imagen se transforma a panorámica (im_{test}) y tras esto, se calcula su descriptor holístico (\vec{d}_{test}).
- Una vez calculado el descriptor, el robot calcula la distancia coseno (elegida en Cebollada et al. (2019) como una de las mejores distancias para el caso de descriptores holísticos presentados) entre el descriptor test (\vec{d}_{test}) y cada uno de los descriptores del modelo visual (\vec{d}_j , donde $j = 1, \dots, N_{Entreno}$).
- Un vector de distancias es obtenido como $\vec{h}_t = \{h_{t1}, \dots, h_{tN_{Entreno}}\}$ donde $h_{tj} = dist\{\vec{d}_{test}, \vec{d}_j\}$.
- El nodo que presenta la mínima distancia ($d_t^m | t = argmin_j h_{tj}$) se corresponde con la posición estimada del robot.

3. Los descriptores holísticos

La localización visual ha sido extensamente solventada tanto usando características locales a lo largo de un conjunto de escenas como utilizando un único descriptor por cada imagen, el cual contiene información de apariencia global. Este segundo método es conocido como descripción holística o descripción de apariencia global y su uso para solventar la localización resulta interesante ya que permite el uso de algoritmos de localización directos. Por ejemplo, Payá et al. Payá et al. (2010) proponen

un método de mapeo a través de éstos descriptores (utilizando firma de Fourier) en entornos de interior y el método de Monte Carlo para solventar el problema de localización.

A grandes rasgos, los pasos para calcular el descriptor holístico son los siguientes: (1) El punto de partida es una imagen panorámica expresada como una matriz bidireccional ($im_j \in \mathbb{R}^{M_x \times M_y}$). (2) Unos determinados cálculos matemáticos son llevados a cabo sobre la matriz para obtener un vector que caracteriza la imagen original ($\vec{d}_j \in \mathbb{R}^{k_1 \times 1}$).

Inicialmente, los descriptores holísticos empleados en visión por computador fueron los basados en métodos analíticos. Si embargo, durante los últimos años, la aparición de técnicas de *deep learning* ha potenciado el uso de descriptores que se basan en estas nuevas técnicas.

3.1. Descriptores basados en métodos analíticos

Estos métodos se basan en el cálculo de gradientes y orientaciones de los diferentes píxeles que conforman la imagen. Su uso es muy frecuente para solventar tareas de robótica móvil a través de información visual. Por ejemplo, Chang et al. Chang et al. (2010) usan el descriptor *gist* para resolver la tarea de localización y navegación. Murillo et al. Murillo et al. (2013) también utilizan *gist* para resolver el problema de localización, pero en este caso, el descriptor era una versión reducida obtenida a través de PCA (Principal Component Analysis). Más recientemente, Román et al. Román et al. (2018) llevó a cabo un estudio en el cual se evaluaron diferentes descriptores holísticos para solventar la localización bajo cambios de iluminación.

Entre los diferentes descriptores que hay, este trabajo propone el uso del descriptor HOG y *gist*, los cuales ya han sido utilizados en trabajos previos Cebollada et al. (2019) y han demostrado presentar resultados muy interesantes para la tarea de localización.

Con respecto al descriptor **HOG**, éste fue presentado por Dalal y Triggs Dalal y Triggs (2005) para resolver la detección de viandantes. En este trabajo, el procedimiento utilizado es el mismo que propuso Leonardis y Bischof Leonardis y Bischof (2000): la imagen se divide en k_1 celdas horizontales. Para cada celda se calcula un histograma de gradientes de orientación (con b barras por histograma). Finalmente, la información de cada uno de los histogramas se reordenan en una única línea para construir el descriptor final $\vec{d} \in \mathbb{R}^{b \cdot k_1 \times 1}$

En cuanto al descriptor *gist*, éste fue introducido por Oliva et al. Oliva y Torralba (2006). En este trabajo, la versión utilizada consiste en lo siguiente: (1) obtener m_2 diferentes resoluciones de la imagen, (2) aplicar filtros de Gabor sobre las m_2 imágenes con m_1 diferentes orientaciones, (3) agrupar los píxeles de cada imagen en k_2 bloques horizontales y (4) reagrupar la información obtenida en una única fila para crear el descriptor final $\vec{d} \in \mathbb{R}^{m_1 \cdot m_2 \cdot k_2 \times 1}$.

3.2. Descriptores basados en técnicas de deep learning

Durante los últimos años, el uso de *deep learning* se ha extendido ampliamente para solventar problemas de visión por computador. En referencia al problema de localización mediante información visual, este trabajo estudia el uso de redes CNN (Convolutional Neural Networks) y el uso de auto-encoders. La idea es obtener vectores que caractericen las imágenes a través de alguna técnica de *deep learning*. Por un lado, estos métodos

pueden resultar interesantes dado que su uso puede centrarse en un tipo de imagen específica (como imágenes en entornos de interior en nuestro caso) y por tanto, disponer de descriptores muy eficientes. Por otro lado, estos métodos requieren de un entrenamiento, lo cual implica un gran procesamiento de datos y por consiguiente un considerable tiempo de cómputo.

Con respecto al uso de las CNNs, estas redes han sido diseñadas para resolver problemas de clasificación de imágenes. En este sentido, (1) se introduce un conjunto de imágenes correctamente etiquetadas a la red para realizar el proceso de aprendizaje y tras esto, (2) la red está ajustada para resolver la clasificación de nuevas imágenes (se introduce una imagen a la entrada y a la salida se obtiene la opción de etiquetado con mayor probabilidad de acierto). Las CNNs están compuestas por varias capas ocultas cuyos parámetros y pesos son ajustados a medida que se van llevando a cabo las iteraciones de entrenamiento. En este trabajo, las salidas generadas por algunas de las capas ocultas son utilizadas para obtener descriptores holísticos. Esta idea ya fue propuesta por otros autores como Mancini *et al.* Mancini *et al.* (2017), quienes usaron la CNN para llevar a cabo una categorización de escenarios junto con el clasificador Naïve Bayes, Payá *et al.* Payá *et al.* (2018) propusieron el uso de un descriptor basado en CNN para crear modelos visuales jerárquicos para robots móviles. La arquitectura de la CNN que se ha utilizado en este trabajo es *places* Zhou *et al.* (2014), la cual ha sido entrenada con alrededor de 2,5 millones de imágenes para llevar a cabo una categorización de 205 posibles tipos de escenas. Para usar esta red, no es preciso realizar un re-entreno, pues la configuración por defecto de las capas proporciona buenas soluciones. La fig. 1 muestra la arquitectura de la CNN *places*, la cual está basada en la arquitectura de la CNN *caffe*. Básicamente, la red consiste en (1) una capa de entrada, (2) varias capas intermedias ocultas y (3) una capa de salida. Dentro de las capas intermedias, (2.1) las primeras capas son empleadas para llevar a cabo un aprendizaje de características (capas que incorporan filtros y la salida de cada capa es utilizada como entrada para la siguiente) y (2.2) las capas finales son entrenadas para llevar a cabo la clasificación (capas que están completamente conectadas y generan vectores que aportan información para clasificar).

En este trabajo se han evaluado la información visual generada a la salida de 5 capas, de las cuales, tres son capas completamente conectadas ('*fc6*', '*fc7*' y '*fc8*') con un tamaño de vector de $[4096 \times 1]$, $[4096 \times 1]$ y $[205 \times 1]$ respectivamente. Además, se han obtenido dos descriptores partiendo de la información generada a la salida de las capas convolucionales 2D ('*conv4*' y '*conv5*'). Estas capas aplican varios filtros sobre la imagen de entrada con la finalidad de activar unas determinadas características de la imagen. Por tanto, la salida de estas capas es un conjunto de imágenes que han sido filtradas de maneras distintas. Finalmente, se calcula el descriptor seleccionando una de los resultados obtenidos tras el filtrado (matriz) y reorganizando esa información en una sola fila (vector). Dado que los tamaños de las matrices resultantes son de $[13 \times 13]$, el tamaño del descriptor generado será $[169 \times 1]$.

En cuanto al uso de los **auto-encoders**, el objetivo de estas redes neuronales es reconstruir la entrada sobre la salida a través de una compresión con representación espacio-latente Hubens (2018). Estas redes en primer lugar comprimen la entrada (codificación) y después reconstruyen la entrada partiendo de la

representación generada (decodificación). La idea consiste en construir una representación que genere descriptores que aporten información útil y un tamaño relativamente pequeño, es decir, entrenar un auto-encoder para extraer la información visual más relevante. Por ejemplo, Gao and Zang Gao y Zhang (2017) usaron auto-encoders para detectar bucles cerrados para la tarea de SLAM (Simultaneous Localization And Mapping).

Para la realización de los experimentos, se han propuesto dos tipos de auto-encoders. Ambos han sido entrenados utilizando los mismos parámetros (coeficiente para el regularizador de pesos, $L_2 = 0,004$; coeficientes que controlan el impacto del regulador de escasez, 4; proporción deseada para entrenar como reacciona una neurona a un ejemplo, 0,15; función de transferencia de decodificación, "*Logistic sigmoid function*"; y máximo número de repeticiones, 1000). Además, ambos auto-encoders han sido entrenados utilizando una tarjeta gráfica GPU (NVIDIA GeForce GTX 108 Ti). Sin embargo, mientras la primera opción (*auto-enc-Frib*) es entrenada con imágenes obtenidas del dataset utilizado para realizar los experimentos (explicado en detalle en sec. 4), la segunda alternativa (*auto-enc-SUN*) ha sido entrenada con imágenes obtenidas de un dataset (SUN 360 DB Xiao *et al.* (2012)) que contiene imágenes panorámicas genéricas (escenarios aleatorios y diferentes a los utilizados para llevar a cabo la localización del robot). El objetivo de esta segunda opción es crear un auto-encoder global basado en imágenes panorámicas de escenarios de interior, el cual proporcione una solución lo bastante buena para obtener descriptores para imágenes panorámicas independientemente del entorno en el que se tengan que aplicar. Esta solución eliminaría la necesidad de tener que realizar un entrenamiento previo al cálculo de los descriptores. Por tanto se solventaría la mayor desventaja presentada en el uso de auto-encoders para obtener descriptores. Para el caso de *auto-enc-Frib*, el dataset de entrenamiento consiste en 519 imágenes panorámicas cuyo tamaño es de $[512 \times 128]$. Además, este auto-encoder es entrenado variando el tamaño de neuronas en la capa oculta y por tanto el tamaño del descriptor resultante también varía en función de dicho parámetro ($[N_{HiddenSize} \times 1]$). En cuanto al tiempo de cómputo para entrenar este tipo de auto-encoder, el ordenador requirió entre 6 minutos y 2,94 horas, dependiendo directamente del número de neuronas en la capa oculta (a mayor número de neuronas, mayor tiempo de cómputo).

4. Experimentos

4.1. Dataset

Los experimentos se han llevado a cabo utilizando el dataset COLD (COsy Localization Database) Pronobis y Caputo (2009), el cual contiene información visual a lo largo de una trayectoria. Este dataset está formado por tres diferentes escenarios de interior en universidades localizadas en tres ciudades diferentes (Friburgo, Saarbrücken y Liubliana), además las imágenes de cada universidad que se presentan fueron capturadas bajo tres condiciones lumínicas distintas (en días soleados, en días nublados y por la noche). Sin embargo, para estos experimentos, solo se considera el escenario de Friburgo bajo condiciones de nublado (condición lumínica que menos afecta las imágenes capturadas). La falta de cambios de iluminación es debido a que este trabajo se centra en estudiar la bondad de los descriptores holísticos para solventar la localización, sin embargo, sí

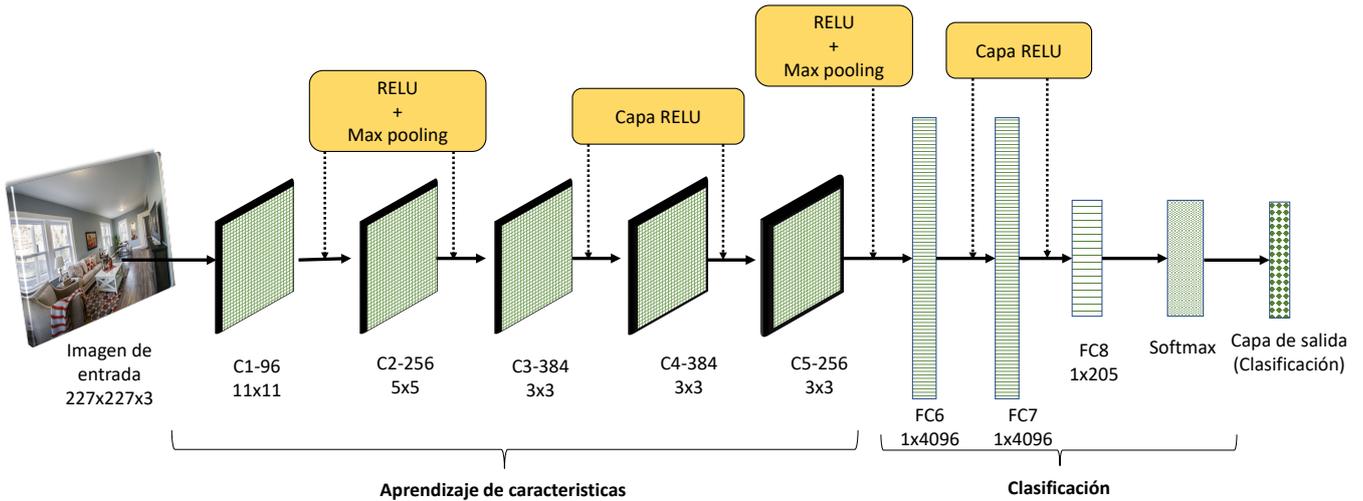


Figura 1: Arquitectura CNN del modelo pre-entrenado de la red 'caffe'.

se plantea realizar el estudio sobre como afecta los cambios de iluminación a la tarea de localización en trabajos futuros. Este dataset incluye cambios en el entorno como gente caminando o cambios de mobiliario y/o objetos. La fig. 2 muestra una imagen ejemplo obtenida del dataset de Friburgo, que fue utilizada para realizar los experimentos. Además, la fig. 3 muestra la vista en ojo de pájaro del dataset utilizado y el camino recorrido por el robot para obtener las imágenes.



Figura 2: Imagen panorámica obtenida de la base de datos COLD.

De todas las imágenes recogidas recorriendo el camino marcado sobre el mapa, éstas fueron repartidas entre dos datasets: el dataset de entrenamiento y el dataset de test. El dataset de entrenamiento está formado por 519 imágenes las cuales presentan una distancia media de entorno 20 cm de separación entre la posición de captura de una imagen y la siguiente. El dataset de test está compuesto por 2595 imágenes y la distancia media de separación es de 4,10 cm. La tabla 1 recoge en detalle la información referente a los datasets.

4.2. Evaluación de la localización

Se consideran dos parámetros para evaluar la eficiencia que presenta cada método de descripción para resolver la tarea de localización: por un lado, (1) el error de localización medio, el cual mide la distancia Euclídea entre la posición estimada por el robot y la posición real (proporcionada por el *ground truth* del dataset, que únicamente es utilizado como medida para evaluación). Por otro lado, (2) el tiempo de cómputo medio, el cual se analiza mediante dos valores, (2.a) el tiempo de cómputo necesario para calcular el descriptor de una imagen y (2.b) el tiempo de cómputo requerido para estimar la posición de la imagen test dentro del mapa.

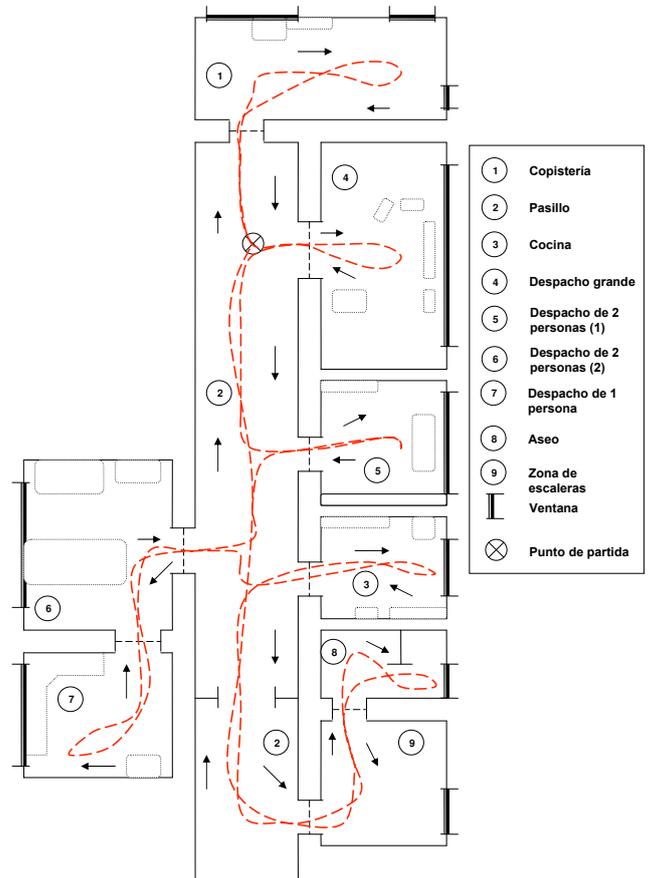


Figura 3: Vista de pájaro del dataset de Friburgo. Original extraído de '<https://www.nada.kth.se/cas/COLD/>'

Las figuras 4, 5, 6, 7 y 8 muestran los resultados obtenidos mediante la utilización de descriptores basados en métodos analíticos (HOG y *gist*) y los basados en técnicas de *deep learning* (auto-encoders y CNN). Estas figuras muestran los datos

Tabla 1: Número de imágenes en cada habitación para los datasets de entrenamiento y test respectivamente creados a partir del dataset de Friburgo.

Nombre	Número de imágenes para entrenamiento	Número de imágenes para test
Copistería	44	223
Pasillo	212	1044
Cocina	51	255
Despacho grande	34	175
Despacho de 2 personas (1)	46	232
Despacho de 2 personas (2)	26	131
Despacho de 1 persona	31	154
Aseo	49	247
Zona de escaleras	26	134
Número total	519	2595

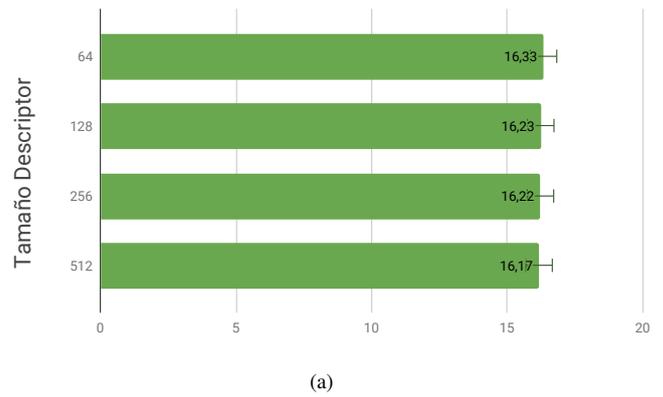
referentes al tamaño del descriptor, error medio de localización (cm), tiempo de cómputo medio para calcular el descriptor (ms) y el tiempo de cómputo medio para estimar la posición de la imagen test (ms).

En cuanto a los resultados obtenidos utilizando descriptores basados en métodos analíticos (ver fig. 4 y 5), para el caso de HOG, el error de localización no decrece significativamente al aumentar el tamaño de descriptor; el tiempo para calcular el descriptor también es prácticamente constante pero el tiempo para estimar la pose crece al aumentar el tamaño del descriptor. Por tanto, se considera como la opción más óptima el descriptor HOG cuyo tamaño es de 64. En lo referente al descriptor *gist*, el error de localización disminuye (en el orden de milímetros) al aumentar el tamaño del descriptor, sin embargo, como es lógico, el tiempo de cálculo aumenta al aumentar dicho valor. Por tanto, para este caso, el tamaño mas pequeño de descriptor también es elegido como la opción más óptima.

Por otro lado, si se analizan los resultados de localización obtenidos mediante los descriptores basados en auto-encoders (ver fig. 6 y 7), para ambos casos (auto-enc-Frib y auto-enc-SUN), las configuraciones con 10 neuronas en la capa intermedia presentan las peores soluciones en cuanto a error de localización. En el caso de auto-enc-Frib, los auto-encoders con $N_{HiddenSize} = [50, 100, 200, 500]$ neuronas proporciona buenas soluciones (error de localización medio entre 7,04 y 7,45 cm), pero para el caso auto-enc-SUN, solo cuando $N_{HiddenSize} = 500$ neuronas se consiguen resultados similares. Si se observa el tiempo de cómputo (tanto para calcular el descriptor como para estimar la posición), de nuevo, a mayor neuronas, mayor tamaño de descriptor y por consiguiente se necesita mayor tiempo de cómputo. Por lo tanto, para el caso de auto-enc-Frib, la opción más óptima es la que el auto-encoder fue entrenado con 100 neuronas (mínimo error de localización y tercer tiempo de cómputo más bajo). En el caso de auto-enc-SUN, a pesar de que la configuración de auto-encoder con 500 neuronas presenta los peores tiempos de cómputo, esta opción es la única que produce buenas soluciones para solventar la tarea de localización.

Finalmente, para los casos de descriptores holísticos basados

Error localización (cm) frente a tamaño de



Tiempo cálculo descriptor (ms) y tiempo estimación

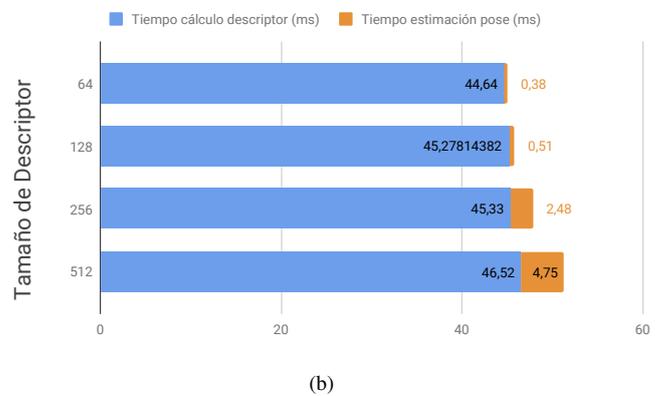


Figura 4: Gráfica de los resultados de localización obtenidos mediante descripción HOG. (a) Errores medios de localización calculados. (b) Tiempos de cómputo que han sido medidos (azul) para calcular el descriptor holístico y (naranja) para estimar la pose del robot en el mapa.

en redes CNN (ver fig. 8), todas las capas que se han evaluado han presentado resultados interesantes. Entre todas las capas estudiadas, las primeras (*'conv4'* y *'conv5'*) alcanzan la máxima precisión (error entorno 5 cm). Este comportamiento resulta lógico si se tiene en cuenta que las primeras capas de las redes CNN están entrenadas para obtener información característica de la imagen, mientras que las últimas capas (capas completamente conectadas) están entrenadas para realizar de la manera más exitosa posible la tarea de clasificación. Los resultados obtenidos de las capas *'conv4'* y *'conv5'* presentan un gran interés, dado que su uso para calcular descriptores holísticos había sido poco considerado y a pesar de ello han presentado resultados muy satisfactorios para resolver la tarea de localización mediante información visual. En cuanto a los tiempos de cómputo, ninguna de las capas estudiadas han presentado grandes valores; además, como era de esperar, el tiempo para calcular el descriptor es directamente proporcional a la posición que ocupa la capa dentro de la red. Por último, cabe destacar que aunque los tiempos de estimación de pose son proporcionales al tamaño del descriptor, se aprecia la particularidad de que el descriptor asociado a la capa *'conv5'* presenta menos tiempo de estimación de pose que el descriptor de la capa *'conv4'* pese a que ambos descriptores presentan mismo tamaño. Por tanto, el descriptor

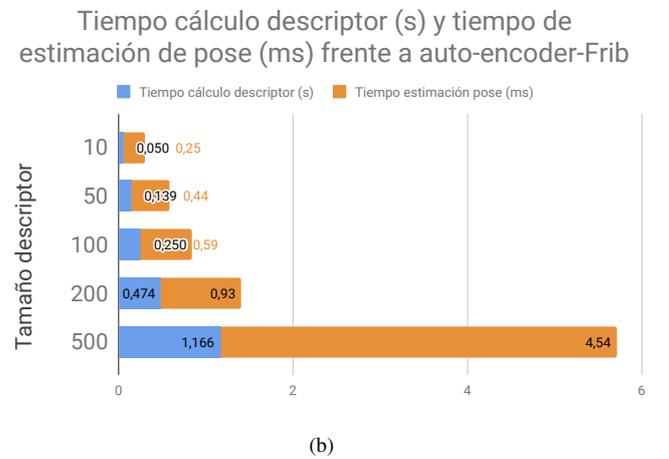
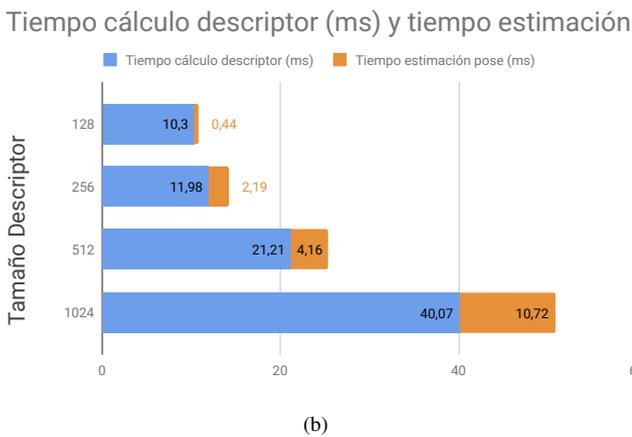
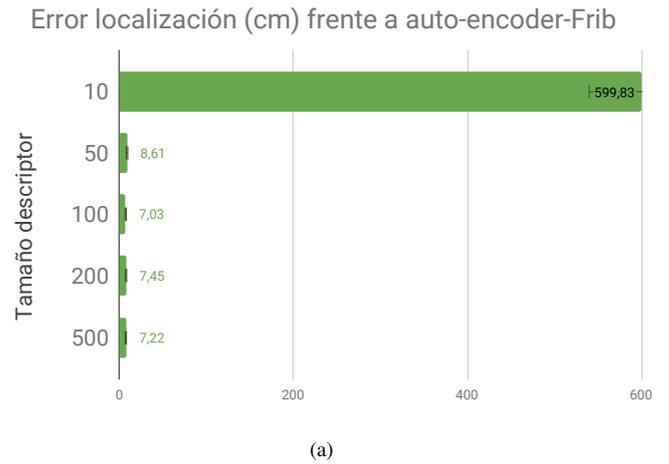
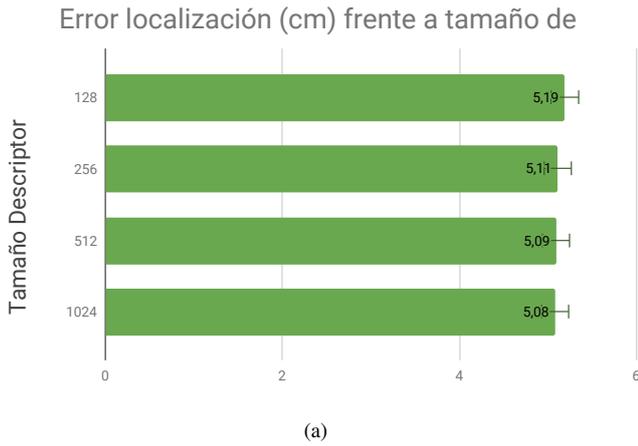


Figura 5: Gráfica de los resultados de localización obtenidos mediante descripción *gist*. (a) Errores medios de localización calculados. (b) Tiempos de cómputo que han sido medidos (azul) para calcular el descriptor holístico y (naranja) para estimar la pose del robot en el mapa.

Figura 6: Gráfica de los resultados de localización obtenidos mediante descripción basada en auto-encoders entrenados con el dataset de Friburgo. (a) Errores medios de localización calculados. (b) Tiempos de cómputo que han sido medidos (azul) para calcular el descriptor holístico y (naranja) para estimar la pose del robot en el mapa.

basado en la capa 'conv5' es seleccionado como el descriptor óptimo de entre todos los referidos a la CNN.

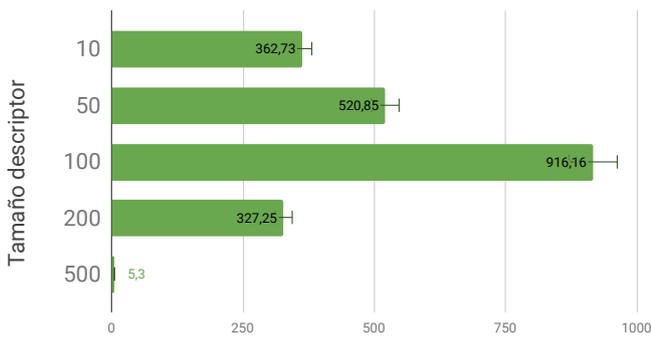
5. Conclusión

Este trabajo propone el estudio de descriptores holísticos para llevar a cabo la tarea de localización. Para resolver esta tarea, se considera el problema conocido como *image retrieval*. Los experimentos realizados se han llevado a cabo a través de un dataset que contiene imágenes panorámicas en entornos de interior con cambios dinámicos. Se han evaluado cinco métodos de descripción holística: dos basados en métodos analíticos (HOG y *gist*), dos basados en el uso de auto-encoders y uno basado en las capas de una red tipo CNN. El tamaño de cada descriptor varía en función de los parámetros de calibración (como por ejemplo el número de barras o *bins* en HOG o el tamaño de neuronas en la capa intermedia para el caso de los auto-encoders) o simplemente al seleccionar la capa de la CNN. Para medir la eficiencia de los descriptores se han utilizado tres parámetros: el error medio de localización, el tiempo medio de cálculo del descriptor y el tiempo medio necesario para estimar la posición del robot. La tabla. 2 muestra los resultados de error de localización, tiempos de cálculo de descriptor y los tiempos de esti-

mación de pose obtenidos para la mejor configuración de cada uno de los descriptores estudiados. El mínimo error de localización se obtiene con el descriptor obtenido de la CNN, aunque los descriptores *gist* y el descriptor basado en auto-enc-SUN presentan resultados muy próximos. De nuevo, el descriptor basado en CNN presenta la mejor solución en cuanto tiempo de cálculo de descriptor, sin embargo, el descriptor HOG, seguido de *gist* y del descriptor basado en auto-enc-Frib, es el más rápido para estimar la posición del robot.

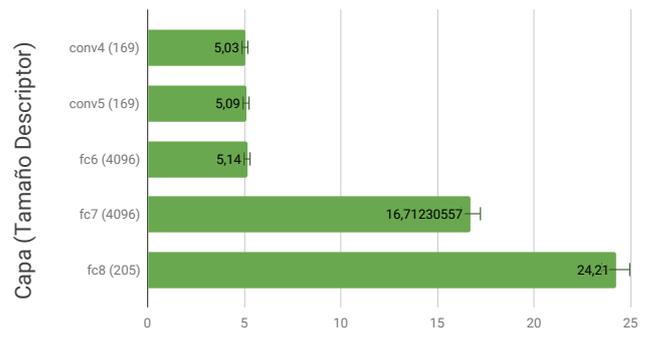
En cuanto al uso de los auto-encoders, se han presentado buenos resultados de precisión para el caso del auto-encoder entrenado con imágenes que pertenecen al escenario donde debe localizarse el robot (auto-enc-Frib). El caso del auto-encoder que se entrena con imágenes genéricas de entornos de interior (auto-enc-SUN), funciona de manera aceptable únicamente cuando la capa intermedia presenta un elevado número de neuronas, por consiguiente, los tiempos de cómputo obtenidos son muy elevados en comparación con el resto de opciones. Sin embargo, el uso de este auto-encoder para calcular descriptores holísticos podría ser una herramienta muy útil para llevar a cabo la descripción de imágenes con ciertas características (en este

Error localización (cm) frente a auto-encoder-SUN



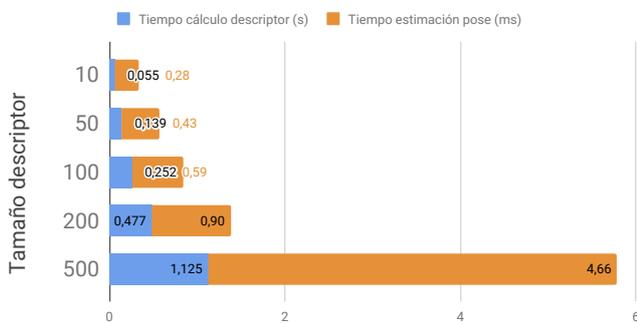
(a)

Error localización (cm) frente a capa de CNN



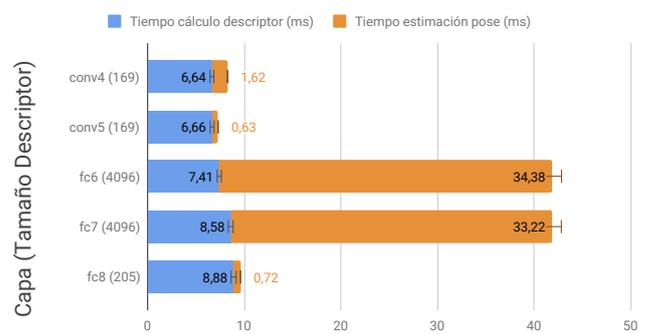
(a)

Tiempo cálculo descriptor (s) y tiempo de estimación de pose (ms) frente a auto-encoder-SUN



(b)

Tiempo cálculo descriptor (ms) y tiempo de



(b)

Figura 7: Gráfica de los resultados de localización obtenidos mediante descripción basada en auto-encoders entrenados con el dataset de SUN Xiao et al. (2012). (a) Errores medios de localización calculados.(b) Tiempos de cómputo que han sido medidos (azul) para calcular el descriptor holístico y (naranja) para estimar la pose del robot en el mapa.

Figura 8: Gráfica de los resultados de localización obtenidos mediante descripción basada en CNN. (a) Errores medios de localización calculados. (b) Tiempos de cómputo que han sido medidos (azul) para calcular el descriptor holístico y (naranja) para estimar la pose del robot en el mapa.

caso, imágenes panorámicas capturadas en entornos de interior) lo cual presentaría la gran ventaja de que este auto-encoder solo es necesario entrenarlo una vez y después podría ser utilizado para resolver la localización en cualquier entorno de interior. Por otro lado, para el uso de descriptores basados en redes CNN, hemos propuesto el uso de capas que no están situadas al final de la arquitectura ('conv4' y 'conv5'). A pesar de que estas capas no son completamente convolucionales (como las capas que típicamente se han empleado para obtener descriptores), los resultados obtenidos para estos descriptores han sido muy interesantes. Además, estos descriptores holísticos han producido las soluciones más óptimas de entre todas las opciones consideradas para resolver la localización: tamaño de descriptor relativamente pequeño (lo que lleva a tiempos de estimación de pose bajos), rapidez para calcular los descriptores y una localización muy precisa (error medio entorno a 5 cm para un dataset de test y dataset de entrenamiento que presentan una distancia media entre poses de alrededor de 4 cm y 20 cm respectivamente).

Por tanto, en trabajos futuros, estudiaremos en detalle el uso de técnicas de *deep learning* para obtener descriptores holísticos. Intentaremos optimizar los auto-encoders desde un punto de vista de tiempo de cómputo. Además, ampliaremos la eva-

luación de la localización cuando los imágenes son afectadas por cambios de iluminación y/o oclusiones.

English Summary

Study of localization for mobile robots through the use of holistic descriptors based on analytical methods and deep learning.

Abstract

Through this work, several holistic descriptors are evaluated to solve the localization task, which is essential in mobile robotics. For this purpose, the robot only uses information obtained from an omnidirectional camera and then calculates the related holistic descriptor. Afterwards, the position of the robot is estimated through comparing the calculated descriptor with all the descriptors contained in the visual model. Two types of descriptors are evaluated throughout this work: (1) descriptors based on analytical methods (HOG and *gist*) and (2) descriptors based on deep learning techniques (auto-encoders and Convolutional Neural Networks). In order to tackle the evaluation, a dataset with panoramic images is used. It provides images captured in indoor environments and also under real conditions. The results obtained from the experiments show that

Tabla 2: Tabla resumen. Se muestran las mejores configuraciones para cada tipo de descriptor estudiado.

Descriptor	Tamaño	Error localización (cm)	Tiempo cómputo descriptor (ms)	Tiempo estimación (ms)
HOG	64	16,33	44,64	0,39
<i>gist</i>	128	5,19	10,3	0,45
auto-enc-Frib	100	7,03	249,55	0,59
auto-enc-SUN	500	5,30	1125,06	4,66
CNN-conv5	169	5,09	6,63	0,62

the deep learning based descriptors can also provide interesting solutions to solve the localization task.

Keywords:

Mobile Robotics, Omnidirectional Images, Holistic Description, Localization, Deep Learning.

Agradecimientos

Este trabajo ha sido realizado gracias al apoyo de la Generalitat Valenciana y el Fondo Social Europeo a través de su beca predoctoral ACIF/2017/146 y gracias al Ministerio de Ciencia e Innovación a través del proyecto DPI 2016-78361-R (AEI/FEDER, UE): “Creación de mapas mediante métodos de apariencia visual para la navegación de robots.”

Referencias

Berenguer, Y., Payá, L., Ballesta, M., Reinoso, O., 2015. Position estimation and local mapping using omnidirectional images and global appearance descriptors. *Sensors* 15 (10), 26368–26395. DOI: 10.3390/s151026368

Cebollada, S., Payá, L., Mayol, W., Reinoso, O., 2019. Evaluation of clustering methods in compression of topological models and visual place recognition using global appearance descriptors. *Applied Sciences* 9 (3), 377.

Chang, C., Siagian, C., Itti, L., 2010. Mobile robot vision navigation and localization using gist and saliency. En: *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*. pp. 4147–4154. DOI: 10.1109/IROS.2010.5649136

Dalal, N., Triggs, B., 2005. Histograms of oriented gradients for human detection. En: *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, San Diego, USA*. Vol. II, pp. 886–893.

Faessler, M., Fontana, F., Forster, C., Mueggler, E., Pizzoli, M., Scaramuzza, D., 2016. Autonomous, vision-based flight and live dense 3d mapping with a quadrotor micro aerial vehicle. *Journal of Field Robotics* 33 (4), 431–450.

Gao, X., Zhang, T., 2017. Unsupervised learning to detect loops using deep neural networks for visual slam system. *Autonomous robots* 41 (1), 1–18.

Hubens, N., 2018. Deep inside: Autoencoders. Accessed February 11, 2019. URL: <https://towardsdatascience.com/deep-inside-autoencoders-7e41f319999f>

Korrapati, H., Mezouar, Y., 2017. Multi-resolution map building and loop closure with omnidirectional images. *Autonomous Robots* 41 (4), 967–987.

Leonardis, A., Bischof, H., 2000. Robust recognition using eigenimages. *Computer Vision and Image Understanding* 78 (1), 99–118.

Mancini, M., Bulò, S. R., Ricci, E., Caputo, B., 2017. Learning deep nbnn representations for robust place categorization. *IEEE Robotics and Automation Letters* 2 (3), 1794–1801.

Murillo, A. C., Singh, G., Kosecka, J., Guerrero, J. J., 2013. Localization in urban environments using a panoramic gist descriptor. *IEEE Transactions on Robotics* 29 (1), 146–160.

Oliva, A., Torralba, A., May 2001. Modeling the shape of the scene: A holistic representation of the spatial envelope. *Int. J. Comput. Vision* 42 (3), 145–175. DOI: 10.1023/A:1011139631724

Oliva, A., Torralba, A., 2006. Building the gist of a scene: the role of global image features in recognition. En: *Progress in Brain Research: Special Issue on Visual Perception*. Vol. 155.

Payá, L., Fernández, L., Gil, A., Reinoso, O., 2010. Map building and monte carlo localization using global appearance of omnidirectional images. *Sensors* 10 (12), 11468–11497. DOI: 10.3390/s101211468

Payá, L., Peidró, A., Amorós, F., Valiente, D., Reinoso, O., 2018. Modeling environments hierarchically with omnidirectional imaging and global appearance descriptors. *Remote Sensing* 10 (4), 522.

Pronobis, A., Caputo, B., May 2009. COLD: COsy Localization Database. *The International Journal of Robotics Research (IJRR)* 28 (5), 588–594. DOI: 10.1177/0278364909103912

Radon, J., 1917. Über die bestimmung von funktionen durch ihre integralwerte laengs gewisser mannigfaltigkeiten. *Berichte Saechsishe Acad. Wissenschaft. Math. Phys., Klass* 69, 262.

Rituerto, A., Murillo, A. C., Guerrero, J., 2014. Semantic labeling for indoor topological mapping using a wearable catadioptric system. *Robotics and Autonomous Systems* 62 (5), 685–695.

Román, V., Payá, L., Reinoso, O., 2018. Evaluating the robustness of global appearance descriptors in a visual localization task, under changing lighting conditions. En: *ICINCO 2018. 15th International Conference on Informatics in Control, Automation and Robotics, Porto (Portugal)*, 29–31 July 2018. pp. 258–265.

Xiao, J., Ehinger, K. A., Oliva, A., Torralba, A., 2012. Recognizing scene viewpoint using panoramic place representation. En: *Computer Vision and Pattern Recognition (CVPR), IEEE Conference on*. pp. 2695–2702.

Xu, J., Xiang, L., Liu, Q., Gilmore, H., Wu, J., Tang, J., Madabhushi, A., 2016. Stacked sparse autoencoder (ssae) for nuclei detection on breast cancer histopathology images. *IEEE Transactions on Medical Imaging* 35 (1), 119–130.

Xu, S., Chou, W., Dong, H., 2019. A robust indoor localization system integrating visual localization aided by cnn-based image retrieval with monte carlo localization. *Sensors* 19 (2), 249.

Zhou, B., Lapedriza, A., Xiao, J., Torralba, A., Oliva, A., 2014. Learning deep features for scene recognition using places database. En: *Advances in Neural Information Processing Systems*. pp. 487–495.