

USO DE TÉCNICAS DE MACHINE LEARNING PARA REALIZAR MAPPING EN ROBÓTICA MÓVIL

Sergio Cebollada, Vicente Román, Luis Payá, María Flores, Luis M. Jiménez, Oscar Reinoso
 Departamento de Ingeniería de Sistemas y Automática
 Universidad Miguel Hernández, Avda. de la Universidad s/n, 03202 Elche (Alicante), España
 {sergio.cebollada, vicente.roman, lpaya, m.flores, luis.jimenez, o.reinoso}@umh.es

Resumen

El trabajo presentado consiste en un estudio de diversos clasificadores basados en machine learning como herramienta para llevar a cabo la tarea de mapeo y localización en robótica móvil. En concreto, estos clasificadores son utilizados para solventar la tarea de localización "gruesa", la cual forma parte de los procesos a realizar para resolver la localización jerárquica. El proceso de localización llevado a cabo por el robot consistiría en (1) capturar una imagen desde una posición desconocida, (2) calcular su correspondiente descriptor de apariencia global, (3) introducir dicha información al clasificador y obtener la estancia en la cual se encuentra el robot en ese instante. Tras esto, (5) el robot realizará el problema de image retrieval con toda la información visual de entrenamiento contenida en la estancia seleccionada (localización fina). Este trabajo evalúa el uso de tres clasificadores (Naïve Bayes, SVM y clasificador basado en red neuronal) los cuales se entrenan con tres posibles descriptores de apariencia global (HOG, gist y un descriptor obtenido a partir de una CNN). Los experimentos se llevan a cabo mediante el uso de un dataset que contiene imágenes omnidireccionales capturadas en entornos de interior y que presenta cambios dinámicos (personas andando, cambios de mobiliario, etc.). Los resultados obtenidos demuestran que el método propuesto es una alternativa eficiente para realizar la tarea de localización jerárquica en cuanto error de localización y tiempo de cómputo.

Palabras clave: Mapping, Localización Jerárquica, Imágenes Omnidireccionales, Descriptores de Apariencia Global, Machine Learning

1. INTRODUCCIÓN

En la actualidad, la información visual desempeña un papel fundamental para solventar tareas en robótica móvil. En concreto, el robot debe crear un mapa (mapping) y posteriormente estimar su posición dentro de dicho mapa (localización). Si bien en el estado del arte se puede encontrar una amplia variedad de sensores para llevar a cabo di-

chas tareas, el uso de información visual resulta interesante puesto que presenta una alternativa económica y proporciona resultados eficientes para las tareas desempeñadas. Entre los sensores visuales, las cámaras omnidireccionales presentan una solución robusta puesto que dichas cámaras aportan información en 360 grados alrededor del robot.

Generalmente, para utilizar la información visual aportada por la cámara, es necesario extraer información característica de la imagen para así poder solventar de manera más efectiva las tareas de mapeo y localización. Una de estas técnicas es conocida como descripción de apariencia global. Mediante este método se extrae información relevante en un único vector que contiene información global sobre la imagen. Podemos encontrar varios ejemplos de sus uso en la bibliografía, como por ejemplo, Korrapati y Mezouar [5], quienes utilizan los descriptores de apariencia global para crear mapas topológicos y llevar a cabo posteriormente la tarea de detección de lazo cerrado. Más recientemente, Zhou *et al.* [17] propone el uso del descriptor *gist* para solventar la tarea de localización a través de un emparejamiento entre el dataset obtenido durante la fase de mapeo y la imagen proporcionada por el robot en el instante actual. Por otro lado, Román *et al.* [12] llevó a cabo una evaluación de la robustez de diferentes descriptores de apariencia global ante cambios de iluminación. Estos trabajos se basan en el uso de descriptores que se calculan de manera analítica, sin embargo, en los últimos años ha emergido el uso de descriptores de apariencia global basados en técnicas de deep learning. Por ejemplo, Xu *et al.* [14] propuso de auto-encoders para obtener descriptores de apariencia global que le permitiesen detectar tumores de pecho; Payá *et al.* [10] utiliza los vectores obtenidos de capas intermedias de una red neuronal convolucional (CNN) como descriptores de apariencia global y a partir de esta información lleva a cabo la creación de mapas jerárquicos.

Dada la potencia de las técnicas de inteligencia artificial desarrolladas durante las últimas décadas, este trabajo pretende estudiar algunas de las herramientas más destacadas dentro del área conocida como *Machine Learning* (aprendizaje au-

tomático) para llevar a cabo de manera mas eficiente la tarea de mapeo en robótica móvil. En concreto, este trabajo analiza el uso de clasificadores como herramienta para llevar a cabo el mapeo jerárquico. Dichos clasificadores tendrán como información de entrada el descriptor de apariencia global de una imagen capturada y a la salida indicarán en qué estancia se capturó la imagen con mayor probabilidad de acierto. Por tanto, a través de este trabajo se lleva a cabo una evaluación de diferentes tipos de clasificadores.

El resto de artículo se estructura de la siguiente manera: la sección 2 explica las herramientas de *machine learning* utilizadas para llevar a cabo los experimentos. Después, en la sección 3, se explica el uso de éstas herramientas para llevar a cabo el proceso de mapeo. La sección 4 presenta la información utilizada así como los experimentos llevados a cabo. Finalmente, en la sección 5 se discuten las conclusiones alcanzadas a través de los resultados obtenidos.

2. HERRAMIENTAS DE MACHINE LEARNING

El aprendizaje profundo o *machine learning* es un método de análisis de datos que automatiza la construcción de modelos analíticos. Esta técnica está incluida dentro de la rama de inteligencia artificial y se sustenta en la idea de que el sistema puede aprender a identificar patrones partiendo exclusivamente de los datos aportados. Para su uso, es necesario seleccionar previamente la información que se utilizará para entrenar el modelo y posteriormente llevar a cabo el proceso de entrenamiento, el cual, dependiendo de los requerimientos establecidos, puede llevar un mayor o menor tiempo. En la bibliografía podemos encontrar trabajos donde se utilizan algunas de estas herramientas para llevar a cabo tareas de visión por computador como por ejemplo, Zhang y Wu [15], que utilizan kSVM (Kernel Support Vector Machine) para clasificar imágenes de frutas. Dentro de la visión por computador, también encontramos trabajos para solventar tareas de robótica móvil, por ejemplo, Tribel [13] propone el uso del clasificador IVM (Informative Vector Machine) para crear mapas semánticos en robots autónomos; por otro lado, Duguleana y Mogan [4] proponen un algoritmo de planificación de trayectorias basado en Q-learning y utilizan redes neuronales para evitar obstáculos durante el camino. En este trabajo se propone el uso de clasificadores para llevar a cabo el mapeo jerárquico en la capa de alto nivel.

2.1. CLASIFICADORES

Un clasificador es básicamente una herramienta que se utiliza para predecir la clase o categoría a la que pertenece un dato. Previamente a poder utilizar esta herramienta, se necesitará llevar a cabo un entrenamiento. Dicho entrenamiento consistirá en introducir datos (x_{train}) etiquetados (y_{train}). De esta manera, se llevará a cabo un mapeo de aproximación de funciones para alcanzar un correcto ajuste de los parámetros. Finalmente, el modelo estará preparado para recibir datos de entrada (x_{in}) y estimar su categoría (y_{est}).

Este trabajo estudia tres tipos de clasificadores para utilizarlos en el proceso de mapeo.

- **Clasificador Naïve Bayes (NB).** Basado en el teorema de Bayes con independencia de suposición entre datos de entrada, este clasificador fue introducido por Maron [8] como método para categorizar correos de entrada (como spam, deportes, política, etc.) utilizando frecuencia de palabras como información de descripción.
- **Clasificador basado en red neuronal.** Este tipo de red neuronal se basa en el reconocimiento de patrones. Este sistema “aprende” a realizar tareas basándose en ejemplos (datos de entrenamiento) sin necesidad de ser programado ni de aprender ninguna regla específica.
- **Clasificador Support Vector Machine (SVM).** Introducido por Cortes y Vapnik [3], estos clasificadores no solo han sido utilizados para solventar tareas de clasificación, si no también han sido utilizados para propósitos de regresión. El algoritmo marca cada dato en un espacio n-dimensional (donde n es el número de características) con el valor de cada característica como situado en una coordenada particular. Después, la clasificación se lleva a cabo encontrando el hiper-plano o hiper-planos que mejor diferencia las categorías.

2.2. CLUSTERING

Los algoritmos de clustering son métodos que consisten en agrupar la información aportada en vectores de acuerdo a un criterio dado. Dicho criterio suele ser la similitud entre los vectores de información. Esta técnica ha resultado ser una buena solución para agrupar información basado en descriptores de apariencia visual ([10] y [1]). En [1], se estudiaron diversos métodos de clustering con el objetivo de comprobar su validez para crear modelos jerárquicos. Para este propósi-

to, los descriptores capturados en el proceso de mapeo (o entrenamiento) componen el dataset visual ($D = \{\vec{d}_1, \vec{d}_2, \dots, \vec{d}_N\}$). Tras esto, el algoritmo de clustering agrupa la información aportada por el dataset en clusters de acuerdo a su similitud ($C = \{C_1, C_2, \dots, C_{n_c}\}$ donde n_c es el número de clusters). Mediante el uso de clustering, se propone crear un mapa de alto nivel basado en representativos. Considerando este trabajo previo, se elige el clustering espectral [6] como el método para llevar a cabo la tarea de clustering para este estudio.

3. MAPPING

Además de las herramientas de *machine learning* mencionadas, en este trabajo también se estudia el uso de descriptores de apariencia global, los cuales han sido utilizados durante los últimos años para resolver la tarea de mapping. Por ejemplo, Payá *et al.* [10] utilizaron descripción de apariencia global para construir modelos topológicos jerárquicos. Basándonos en este trabajo, el presente estudio analiza el uso de los descriptores HOG, *gist* y un descriptor basado en deep learning. En cuanto a HOG, el cálculo de este descriptor se basa en dividir la imagen en k_1 celdas horizontales y calcular un histograma de gradientes de orientación para cada celda con b barras por histograma. Por último, los histogramas calculados se reordenan en una única fila, obteniendo así un vector del tipo $\vec{d}_{HOG} \in \mathbb{R}^{b \cdot k_1 \times 1}$. Para obtener el descriptor *gist*, primero se crean m_2 imágenes con diferente resolución partiendo de la imagen original. Después, se aplica un filtro de Gabor sobre las m_2 imágenes aplicando m_1 diferentes orientaciones. Tras esto, los píxeles de cada imagen se agrupan en k_2 bloques horizontales y por último, la información de orientación obtenida se agrupa para formar un vector del tipo $\vec{d}_{gist} \in \mathbb{R}^{m_1 \cdot m_2 \cdot k_2 \times 1}$. Se puede obtener una explicación mas detallada en [9]. Por último, en relación al descriptor basado en deep learning, el descriptor se obtiene de una capa intermedia incluida en una red neuronal convolucional (CNN), estas redes reciben un exhaustivo entrenamiento para llevar a cabo una determinada tarea (comúnmente de clasificación). Las CNNs están compuestas por varias capas ocultas, cuyos parámetros y ponderaciones se ajustan durante el proceso de entrenamiento. La idea consiste en introducir a la CNN la imagen y extraer un vector de de una de las capas intermedias, de manera que ese vector sea el descriptor de apariencia global. Esta técnica ha sido utilizada previamente por otros autores como por ejemplo Mancini *et al.* [7] para llevar a cabo una categorización de escenarios mediante un clasificador Naïve Bayes. Por otro lado, en trabajos previos ([10] y [2]) ya se ha probado el uso de estos descriptores para lle-

var a cabo tareas de mapping visual. El descriptor utilizado para este estudio es extraído de la red *places* [16] y corresponde con la capa completamente convolucional 'fc7'. este descriptor es de la forma ($\vec{d}_{fc7} \in \mathbb{R}^{4096 \times 1}$).

Por tanto, la tarea de mapping propuesta se resume en: el robot se mueve a través del entorno y captura imágenes omnidireccionales (ver fig. 1). Para cada imagen capturada, se calcula el descriptor de apariencia global (basado en alguno de los métodos propuestos). Finalmente, tras recorrer el entorno y obtener las N_{Train} imágenes, se obtiene un dataset visual formado por todos los descriptores calculados; $D = \{\vec{d}_1, \vec{d}_2, \dots, \vec{d}_{N_{Train}}\}$ donde cada descriptor es $\vec{d}_i \in \mathbb{C}^{l \times 1}$ y corresponde a la imagen im_i .

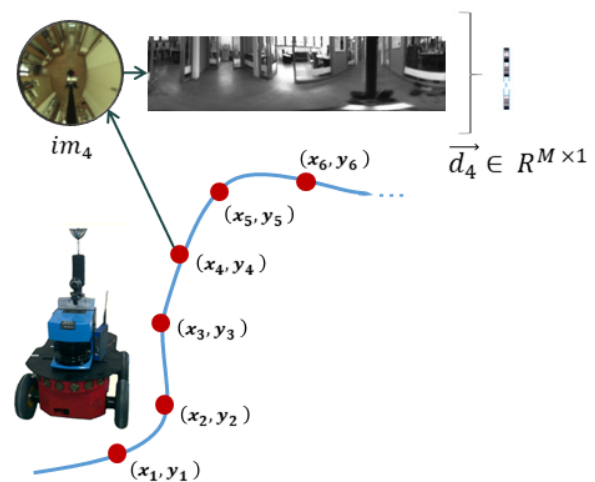


Figura 1: Proceso de mapping. El robot captura imágenes en distintas posiciones del escenario y calcula el descriptor de apariencia global.

En cuanto a la creación de modelos topológicos jerárquicos, los métodos propuestos en trabajos previos ([10] y [1]) consistieron en crear diferentes mapas con distinto nivel de detalle. De esta manera, con el mapa de alto nivel (que incorpora menos detalle) se llevaría a cabo una localización “gruesa” que determinaría el área en el cual se encuentra el robot. Tras esto, se llevaría a cabo una localización “fina” en el área seleccionada. La toma de decisión se basó en comparar la información proporcionada por el robot en la posición a estimar con respecto a un conjunto de descriptores representativos. Estos descriptores son calculados como la media de los descriptores que componen el área $R = \{r_1^-, r_2^-, \dots, r_{n_a}^-\}$ donde n_a es el número de áreas consideradas. De esta manera, se consigue una compactación del modelo que permite llevar a cabo una localización jerárquica más eficiente que mediante una comparación del dato obtenido por el robot con respecto a toda la información

recogida durante el proceso de entrenamiento.

3.1. CLASIFICADORES PARA CREACIÓN DE MAPAS DE ALTO NIVEL

Este trabajo pretende presentar una alternativa al uso de descriptores representativos para la creación de mapas de alto nivel en los modelos topológicos jerárquicos. De esta forma, los clasificadores estarían entrenados con la información visual (descriptores de apariencia global) y el etiquetado correspondiente a la habitación donde se capturó cada imagen. Una vez entrenado el clasificador, para solventar la localización gruesa bastaría con introducir en el sistema el descriptor calculado en la posición a estimar y éste devolvería la categoría (habitación) más probable.

En resumen, el proceso de localización consistiría en los siguientes pasos: (1) el robot captura una imagen desde una posición desconocida (im_{test}) y (2) calcula el descriptor de apariencia global (\vec{d}_{test}). (3) El descriptor obtenido es introducido en un sistema clasificador (previamente entrenado), el cual predecirá el área más probable donde se encuentra el robot (C_i). (4) Tras esta predicción, el descriptor \vec{d}_{test} es comparado con todos los descriptores que componen el área seleccionada ($D_i = \{\vec{d}_{i,1}, \vec{d}_{i,2}, \dots, \vec{d}_{i,N_i}\}$). Finalmente, se estima la posición del robot como la posición que corresponde al descriptor que presenta la mínima distancia entre \vec{d}_{test} y $\vec{d}_{i,j}$. La fig. 2 muestra el diagrama de localización propuesto.

4. EXPERIMENTOS

4.1. DATASET

El dataset empleado para llevar a cabo los experimentos es obtenido de la base de datos COLD (COsy Localization Database)[11], la cual aporta información obtenida por robots en entornos de interior ante cambios dinámicos y frente a tres posibles condiciones de iluminación. En particular se utiliza el dataset del entorno Friburgo, el cual contiene 9 habitaciones: una sala de impresoras, una cocina, cuatro despachos, un baño, una zona de escaleras y un pasillo que conecta todas las estancias. Para trabajar con la información visual aportada, en primer lugar se realiza un muestreo para retener únicamente una imagen cada 20-40 cm. Este muestreo permite por un lado establecer un dataset con una distancia entre posiciones de captura de imágenes similar a otros datasets empleados con anterioridad y por otro lado, del mismo dataset obtenemos un dataset de entrenamiento y otro dataset (dataset de test, que con-

tiene el número original de imágenes) el cual contendrá imágenes capturadas en mismas posiciones y también en distintas posiciones a las capturadas en el dataset de entrenamiento. El dataset de entrenamiento está compuesto por un total de 519 imágenes omnidireccionales y el dataset de testeo por 2595 imágenes que contiene una imagen del entorno cada 5 cm aproximadamente. Para este estudio no se tiene en cuenta los cambios de iluminación, por tanto, todas las imágenes son extraídas del dataset en condiciones de nublado, ya que es el que menos se ve afectado por la iluminación. Además, para poder utilizar los descriptores de apariencia global, se realiza una transformación de imágenes omnidireccionales a imágenes panorámicas.

4.2. EXPERIMENTO 1: EVALUACIÓN DE LOS CLASIFICADORES

Este experimento lleva a cabo la evaluación de los tres clasificadores propuestos para llevar a cabo la selección de la estancia a través del uso de descriptores de apariencia global. Previamente a testear el ratio de acierto de cada combinación “método de descripción + tipo de clasificador”. Para dicho entrenamiento, se utiliza los descriptores de las imágenes correspondientes al dataset de entrenamiento y las etiquetas correspondientes a cada imagen, las cuales se proporcionan por el *ground truth* de la base de datos y determina en qué estancia del entorno se capturó cada imagen. Una vez finalizado el entrenamiento, el clasificador ya estará preparado para recibir nueva información visual (datos de test) y obtener la categoría correspondiente.

Por tanto, el resultado que se obtiene en este experimento es el ratio de acierto de cada configuración clasificador + método de descripción. La tabla 1 muestra los resultados obtenidos. A partir de estos datos, se alcanza la conclusión de que tanto el clasificador basado en red neuronal como el clasificador SVM presentan una correcta opción para llevar a cabo la tarea de selección de estancia. Ambas opciones alcanzan valores de acierto de entorno al 98 % cuando se utilizan los descriptores *gist* o el descriptor basado en la capa *fc7* de la CNN (descriptor CNN-fc7). Por otro lado, se alcanza la conclusión de que el clasificador Naïve Bayes produce los peores resultados independientemente del tipo de descriptor utilizado. También se descarta el uso del descriptor HOG para llevar a cabo esta tarea. El clasificador SVM junto con el descriptor *gist* proporciona la configuración más óptima, la cual produce un ratio de acierto del 98.61 %. Para este resultado, se muestra la matriz de confusión obtenida a través de la fig. 3. A través de esta matriz se puede observar que la categori-

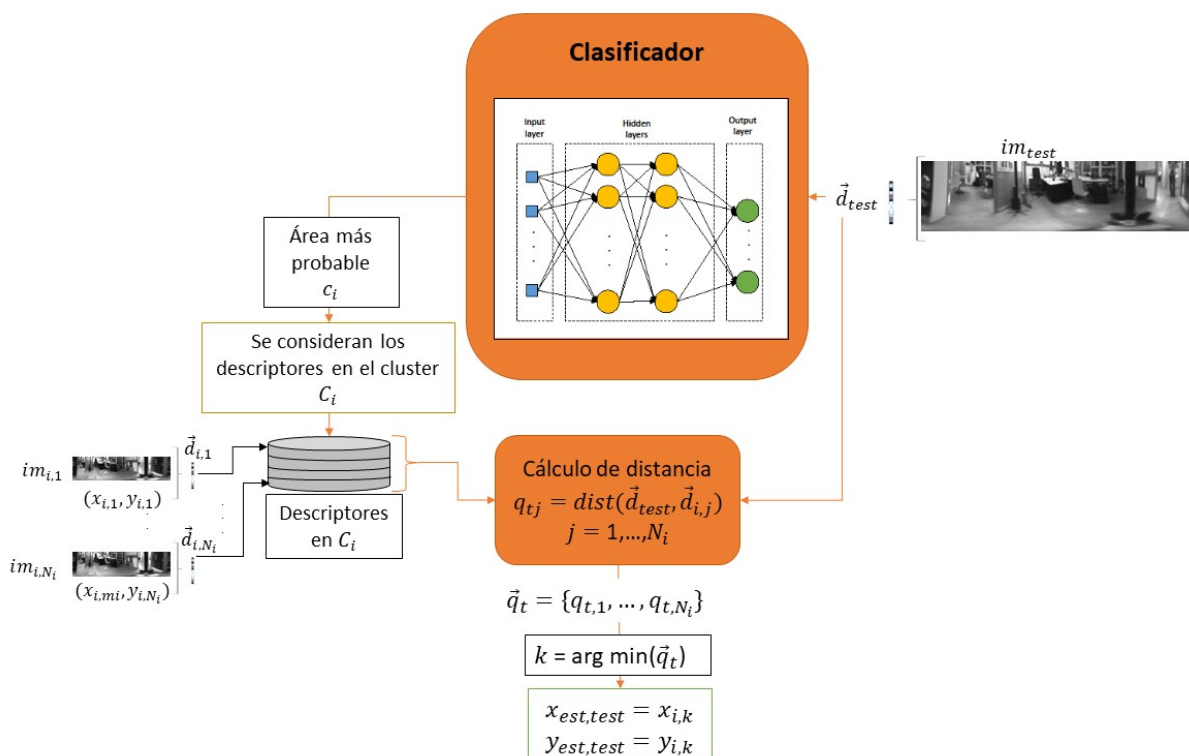


Figura 2: Diagrama de localización jerárquica. Se resuelve la localización gruesa mediante un clasificador para predecir el área más probable en el mapa. La localización fina se resuelve comparando los descriptores contenidos en el área seleccionada a través de un proceso de *image retrieval*.

Cuadro 1: Resultados experimento 1. Ratio de acierto de las configuraciones clasificador + Descriptor estudiadas respecto a datos del dataset de testeo.

Tipo de clasificador	Descriptor	Ratio de acierto(%)
Naïve Bayes	gist	0,8674373796
Naïve Bayes	CNN-fc7	0,8612716763
Naïve Bayes	HOG	0,04238921002
Neural Network	gist	0,9857418112
Neural Network	CNN-fc7	0,9741811175
Neural Network	HOG	0,04971098266
SVM	gist	0,9861271676
SVM	CNN-fc7	0,9849710983
SVM	HOG	0,07244701349

zación es correcta en la mayoría de casos y que los fallos se producen en su mayoría por el pasillo, lo cual se debe a las dimensiones y a la diferencia visual entre las imágenes que componen esta estancia.

4.3. EXPERIMENTO 2: LOCALIZACIÓN JERÁRQUICA

El experimento anterior ha demostrado que los clasificadores propuestos pueden utilizarse para

	Estancia predicha									
	Sala de impresoras	Pasillo	Cocina	Despacho grande	Despacho 2-P #1	Despacho 2-P #2	Despacho 1-P	Baño	Zona de escaleras	
Sala de impresoras	219	4	0	0	0	0	0	0	0	98,21%
Pasillo	0	1039	0	0	2	1	1	0	1	99,52%
Cocina	0	1	254	0	0	0	0	0	0	99,61%
Despacho grande	0	4	0	171	0	0	0	0	0	97,71%
Despacho 2-P #1	0	5	0	0	227	0	0	0	0	97,84%
Despacho 2-P #2	0	4	0	0	0	126	1	0	0	96,18%
Despacho 1-P	0	0	0	0	0	2	152	0	0	98,70%
Baño	0	4	0	0	0	0	0	242	1	97,98%
Zona de escaleras	0	2	0	0	0	0	0	3	129	96,27%
	100,00%	97,74%	100,00%	100,00%	99,13%	97,67%	98,70%	98,78%	98,47%	

Figura 3: Matriz de confusión del clasificador SVM mediante el uso del descriptor *gist* para estimar la estancia correspondiente para las imágenes del dataset de testeo.

llevar a cabo la localización gruesa dentro del proceso de localización jerárquica propuesto (sec. 3). En este experimento se pretende implementar el proceso completo de localización jerárquica, es decir, añadir el paso de localización fina. Este proceso ya ha sido testeado en trabajos previos ([2]) y ha demostrado ser una alternativa eficiente para resolver la tarea de localización en robótica móvil. En resumen, los pasos que se llevan a cabo en este proceso de localización jerárquica son los siguientes: (1) la localización gruesa se resuelve mediante el uso de clasificadores (tal como se ha estudiado en 4.2) y se estima la estancia más proba-

ble. Tras esto, (2) la localización fina se resuelve comparando la información de la imagen a testear únicamente con la información correspondiente a las imágenes de entrenamiento contenidas en la estancia estimada (image retrieval).

Para evaluar el proceso de localización jerárquica propuesto, se utilizan los clasificadores SVM y la red neuronal y los descriptores *gist* y CNN-fc7, dado que estas configuraciones generaron los mejores resultados en el experimento anterior. Por otro lado, para comparar con métodos llevados a cabo en trabajos previos, se propone también la localización jerárquica llevada a cabo mediante el uso de descriptores representativos para la localización gruesa (tal y como se hizo en [2]). Para este método, en primer lugar se lleva a cabo un algoritmo de clustering espectral, el cual agrupa la información de descripción con respecto a la similitud entre las imágenes. Una vez agrupada la información, se obtiene un representativo por grupo; dicho representativo se calcula como la media de de los descriptores contenidos en el grupo. De esta manera, la localización gruesa también esta basada en un problema de image retrieval pero a un nivel menor de precisión (a cambio de un menor tiempo de computo). Para este caso, se decide que el algoritmo de clustering agrupe la información en 9 clusters ($n_c = 9$) y así poder equiparar los resultados obtenidos. Además, se propone utilizar una de las mejores soluciones obtenidas mediante este método: clustering espectral mediante el uso del descriptor *gist* y localización mediante el uso del descriptor HOG. La evaluación de este experimento se lleva a cabo con el dataset de imágenes de test y se medirá la eficiencia de localización mediante el error medio de localización (distancia media entre la posición estimada y la posición real de la imagen test) y el tiempo medio de computo para estimar las posición del robot (calcular el descriptor, estimar la estancia/el área y estimar la posición dentro de esa estancia/área). Los experimentos se han realizado con un PC con CPU Intel Core i7-7700 ® @ 3,6 GHz y mediante el software Matlab ®.

La fig. 4 muestra los valores obtenidos para los métodos de localización jerárquica. Se muestra el error medio y tiempo medio de localización. Los métodos propuestos mediante el uso de clasificadores son (1) red neuronal con descriptor *gist* (NN+*gist*), (2) red neuronal con descriptor CNN-fc7 (NN+CNN-fc7), (3) SVM con descriptor *gist* (SVM+*gist*) y (4) SVM con descriptor CNN-fc7 (SVM+CNN-fc7). Por último, se muestra el método basado en representativos obtenidos a través de clustering espectral (Clustering Esp.+HOG) (5) y representa los resultados obtenidos cuando el número de clusters es 9. De esta gráfica se concluye

que la localización basada en clasificadores presenta soluciones más eficientes que la localización basada en descriptores representativos. Por un lado, los resultados obtenidos mediante el clasificador de red neuronal son los que proporcionan un menor tiempo de computo; por otro lado, pese a que el tiempo de computo es mucho mayor para el clasificador SVM (más que incluso los obtenidos con el método de clustering espectral), el error de localización obtenido es mucho menor, alcanzándose el mínimo para la configuración SVM+*gist* (error medio de 5,09 cm).

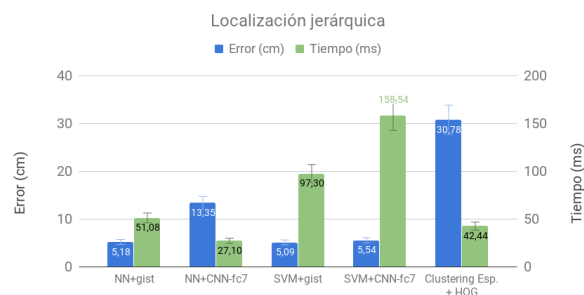


Figura 4: Comparación entre métodos de localización jerárquica mediante el uso de clasificadores o representativos para resolver la localización gruesa.

5. CONCLUSIONES

Este trabajo presenta un estudio llevado a cabo para resolver la tarea de localización jerárquica en robótica móvil mediante el uso de clasificadores de categorías basados en machine learning. Los clasificadores son utilizados para llevar a cabo el mapping correspondiente al mapa de alto nivel que compone el modelo topológico jerárquico. Se han evaluado tres clasificadores: Naïve Bayes, red neuronal y SVM. Además, los clasificadores reciben como entrada información visual mediante descriptores de apariencia global. Los descriptores propuestos para evaluar los clasificadores han sido: HOG, *gist* y un descriptor obtenido a partir de la capa 'fc7' de la CNN *places* (CNN-fc7). Los experimentos han sido realizados mediante el uso de un dataset que contiene imágenes panorámicas en entornos de interior con cambios dinámicos. Para medir la efectividad de acierto de los clasificadores propuestos, se calcula el ratio de acierto y para medir la eficiencia que presenta estos métodos para llevar a cabo la localización jerárquica, se utiliza el error medio de localización y el tiempo medio para estimar la pose. Los resultados obtenidos muestran que tanto el clasificador SVM como la red neuronal presentan un ratio de acierto lo suficientemente elevado para poder realizar la tarea de localización. Por otro lado, se ha probado que

el uso de los clasificadores mejora la eficiencia para realizar la localización jerárquica en comparación con otros métodos probados en anterioridad basados en el uso de descriptores representativos. Ya que, tanto el error de localización como el tiempo de cómputo se ven reducidos en comparación con los resultados recogidos mediante el método de clustering espectral (para un mismo nivel de precisión en la capa alta del mapa jerárquico).

En futuros trabajos, se extenderá este estudio comprobando la validez de estos métodos ante cambios de iluminación y efectos nocivos sobre las imágenes capturadas. Por otro lado, se estudiará el uso de otras herramientas de *machine learning* para conseguir soluciones más robustas.

Agradecimientos

Este trabajo está cofinanciado por la Unión Europea a través del Programa Operativo del Fondo Social Europeo (FSE) de la Comunitat Valenciana 2014-2020 a través de su beca predoctoral ACIF/2017/146 y ACIF/2018/224 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.”

English summary

USE OF MACHINE LEARNING TECHNIQUES TO TACKLE MAPPING IN MOBILE ROBOTICS

Abstract

This work introduces a study regarding the use of several classifiers based on machine learning tools to carry out the mapping and localization task in mobile robotics. These classifiers are used to solve the rough localization, which is part of the hierarchical localization process. Therefore, the localization tackled by the robot consists in (1) obtaining an image from an unknown position, (2) calculating its related global appearance descriptor, (3) putting this information into the classifier to estimate the current room. Afterwards, (5) the robot carries out the image retrieval problem with all the visual information provided by the training dataset contained in the selected room (fine localization step). This work evaluates the use of three types of classifiers (Naïve Bayes, SVM and a classifier based on neural networks) which are trained with three possible global appearance descriptors (HOG, gist and a descriptor calculated from a CNN). The experiments are carried out through the use of a dataset which contains omnidirectional images captured indoor under dynamic changes (people walking, furniture changes, etc.). The results obtained show that this method proposed is an efficient alternative to tackle the hierarchical localization regarding the localization error and the computing time.

Keywords: Mapping, Hierarchical Localization, Omnidirectional Images, Global Appearance Descriptors, Machine Learning.

Referencias

- [1] S. Cebollada, L. Payá, W. Mayol, and O. Reinoso. Evaluation of clustering methods

- in compression of topological models and visual place recognition using global appearance descriptors. *Applied Sciences*, 9(3):377, 2019.
- [2] S. Cebollada, L. Payá, V. Román, and O. Reinoso. Hierarchical localization in topological models under varying illumination using holistic visual descriptors. *IEEE Access*, 7:49580–49595, 2019.
- [3] C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, Sep 1995.
- [4] M. Duguleana and G. Mogan. Neural networks based reinforcement learning for mobile robots obstacle avoidance. *Expert Systems with Applications*, 62:104 – 115, 2016.
- [5] H. Korrapati and Y. Mezouar. Multi-resolution map building and loop closure with omnidirectional images. *Autonomous Robots*, 41(4):967–987, 2017.
- [6] U. Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17:395–416, 2007.
- [7] M. Mancini, S. R. Bulò, E. Ricci, and B. Caputo. Learning deep nbnn representations for robust place categorization. *IEEE Robotics and Automation Letters*, 2(3):1794–1801, 2017.
- [8] M. E. Maron. Automatic indexing: An experimental inquiry. *J. ACM*, 8(3):404–417, July 1961.
- [9] L. Payá, F. Amorós, L. Fernández, and O. Reinoso. Performance of global-appearance descriptors in map building and localization using omnidirectional vision. *Sensors*, 14(2):3033–3064, 2014.
- [10] L. Payá, A. Peidró, F. Amorós, D. Valiente, and O. Reinoso. Modeling environments hierarchically with omnidirectional imaging and global-appearance descriptors. *Remote Sensing*, 10(4):522, 2018.
- [11] A. Pronobis and B. Caputo. COLD: COsy Localization Database. *The International Journal of Robotics Research (IJRR)*, 28(5):588–594, May 2009.
- [12] V. Román, L. Payá, and O. Reinoso. Evaluating the robustness of global appearance descriptors in a visual localization task, under changing lighting conditions. In *ICINCO 2018. 15th International Conference on Informatics in Control, Automation and Robotics, Porto (Portugal), 29-31 July 2018*, pages 258–265, 2018.
- [13] R. Triebel, Hugo Grimmett, Rohan Paul, and Ingmar Posner. *Driven Learning for Driving: How Introspection Improves Semantic Mapping*, pages 449–465. Springer International Publishing, Cham, 2016.
- [14] J. Xu, L. Xiang, Q. Liu, H. Gilmore, J. Wu, J. Tang, and A. Madabhushi. Stacked sparse autoencoder (ssae) for nuclei detection on breast cancer histopathology images. *IEEE Transactions on Medical Imaging*, 35(1):119–130, 2016.
- [15] Y. Zhang and L. Wu. Classification of fruits using computer vision and a multiclass support vector machine. *Sensors*, 12(9):12489–12505, 2012.
- [16] B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva. Learning deep features for scene recognition using places database. In *Advances in Neural Information Processing Systems*, pages 487–495, 2014.
- [17] X. Zhou, Z. Su, D. Huang, H. Zhang, T. Cheng, and J. Wu. Robust global localization by using global visual features and range finders data. In *2018 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, pages 218–223, Dec 2018.



© 2019 by the authors.
Submitted for possible
open access publication

under the terms and conditions of the Creative Commons Attribution CC BY-NC-SA 4.0 license (<https://creativecommons.org/licenses/by-nc-sa/4.0/deed.es>).