

EVALUACIÓN DE DESCRIPTORES DE APARIENCIA GLOBAL EN TAREAS DE LOCALIZACIÓN BAJO CAMBIOS DE ILUMINACIÓN

Vicente Román, Luis Payá, Adrián Peidró, David Valiente, Luis Miguel Jimenez, Óscar Reinoso
Departamento de Ingeniería de Sistemas y Automática
Universidad Miguel Hernández, Avenida de la Universidad s/n 03202 Elche (Alicante), España
{vicente.roman, lpaya, adrian.peidro, dvaliente, luis.jimenez, o.reinoso}@umh.es

Resumen

La localización y creación de mapas son dos tareas que los robots móviles autónomos deben desarrollar. Por ello, se han llevado a cabo numerosas investigaciones sobre estos temas, proponiendo diferentes métodos para abordar estos problemas. Este trabajo describe brevemente algunas posibilidades para mejorar la tarea de navegación utilizando descriptores de apariencia global. Con este propósito, el método captura información visual del entorno y obtiene, mediante transformaciones, un descriptor de apariencia global de cada imagen completa. Utilizados estos descriptores de apariencia global, el robot es capaz de estimar su posición en un mapa previamente creado, donde el reto es verificar el funcionamiento los descriptores cuando el entorno sufre variaciones en su apariencia visual similares a las condiciones reales a las que un robot móvil se puede enfrentar en un entorno real. Para llevar a cabo el objetivo se ha realizado una evaluación comparativa entre algunos descriptores de apariencia global.

Palabras clave: Robótica móvil, descriptores de apariencia global, imágenes omnidireccionales, mapping, localización.

1. INTRODUCCIÓN

En las últimas décadas, la presencia de robots móviles en ambientes industriales y domésticos se ha incrementado notoriamente. Este aumento ha sido facilitado por un desarrollo de las capacidades de percepción y computación, el desarrollo de estas aptitudes ha permitido que los robots mejoren la habilidad de operar en áreas grandes y heterogéneas sin necesidad de cambiar la estructura del robot. Sin embargo, el despegue de los robots móviles no será definitivo hasta que el nivel de autonomía y adaptabilidad a diferentes condiciones no mejore sustancialmente. Un robot autónomo debe ser capaz de encontrar una buena solución en un tiempo aceptable a dos problemas críticos: la construcción de un modelo del entorno (mapping) y la estimación de la posición del robot en el modelo (localización); si es capaz de realizar ambas tareas simultáneamente, el robot realiza una tarea de SLAM (Simultaneous Loca-

lization And Mapping).

Para solucionar estos problemas el robot debe estar equipado con una serie de sensores que le permita extraer información del entorno. Con este objetivo muchos trabajos han usado cámaras como sensores de visión ya que pueden capturar información relevante para realizar las tareas de navegación. Existen dos formas de extraer información de una imagen: la extracción de puntos característicos o el tratamiento de la imagen en su conjunto sin extraer características locales.

La utilización de métodos de descripción de apariencia local, en los que se detectan puntos característicos conocidos comúnmente como landmarks, ha sido la manera clásica de obtener información relevante en sistemas de localización mediante visión. Esta es una alternativa madura usada por muchos investigadores en tareas de mapping y localización. Por ejemplo, Valgren y Lilienthal [16] han propuesto un método de localización topológico utilizando esta técnica, Muriello et al. [7] presentaron un estudio comparativo sobre diferentes descriptores locales en tareas de localización interior, por último, Gil et al. [3] presentaron una evaluación de este tipo de descriptores comparando su repetitividad, su invariancia y su singularidad bajo diferentes condiciones. Sin embargo, una nueva alternativa ha emergido recientemente con la que se trata la imagen de forma global sin extraer sus características locales. Esta técnica puede ofrecer resultados relativamente buenos, comparada con la técnica de localización mediante el uso de características locales. Además esta técnica ofrece la ventaja de simplificar la estructura de los mapas. Adicionalmente, el proceso de localización se reduce a una comparación entre descriptores globales. Como desventaja, se tiene que trabajar con una gran cantidad de datos así que es necesario el uso de una técnica de compresión que reduzca el coste computacional.

Algunos investigadores han propuesto diferentes métodos en los que se reduce la información de manera que el coste computacional no sea tan elevado, manteniendo información necesaria para realizar las tareas de mapping y localización. Estos métodos son denominados descriptores de apariencia global; por ejemplo, la transformada discreta de Fourier es uno de estos métodos, Menegatti [6] presenta el concepto de Firma de Fourier (FS) y lo utiliza para hacer localiza-

ción y mapping. Otros estudios exponen un método basado en el histograma de la orientación del gradiente (HOG). Por ejemplo Dalal y Triggs utilizaron este método para detectar peatones [2] y Paya et al. han estudiado su uso como descriptor en tareas de mapping y localización [12]. Además, cabe destacar la existencia de un método conocido como gist el cual trata de imitar la visión humana; Oliva y Torralba fueron los precursores de esta idea [8], otros autores como Siagian e Itti han estudiado este método testeándolo en ambientes de exterior [15]. Por último, también es posible utilizar la Transformada de Radon (RT) como descriptor de imagen [1].

La finalidad de este trabajo es investigar en la descripción de imágenes y utilizar el descriptor para realizar tareas de navegación. En los experimentos se utiliza imágenes omnidireccionales que ofrecen un rango de visión de 360° . Con este sistema, las características permanecen más tiempo dentro del rango de captura, esto permitirá obtener descriptores más estables. Además, las imágenes omnidireccionales tienen información suficiente para crear un modelo, aunque un objeto o persona haya ocluido parcialmente la imagen durante la ruta.

Algunos autores han trabajado antes en este área. En su trabajo, Payá et al. presentan una evaluación comparativa de algunos descriptores de apariencia global en la creación de mapas topológicos [12], para más tarde realizar un estudio de la localización utilizando los descriptores FS, HOG y gist y los mapas topológicos [11]. La finalidad de este trabajo es extender los estudios anteriores. Por un lado, añadir un nuevo descriptor, la transformada de Radon; por otro lado, comparar los métodos en un entorno real y variante.

En tareas realizadas en un entorno real un robot debe ser capaz de afrontar situaciones como escenas ocluidas parcialmente por gente u otros robots móviles, cambios en las condiciones de iluminación entre el modelo y una nueva localización y pequeños cambios en la escena como movimientos en la posición del mobiliario, entre otros eventos. Teniendo en cuenta estos hechos se ha estudiado la fiabilidad de los descriptores de apariencia global, verificando su rendimiento cuando el entorno visual cambia substancialmente.

2. DESCRIPTORES DE APARIENCIA GLOBAL

En esta sección se resumen las alternativas escogidas para la extracción de información global de la imagen. Inicialmente las imágenes son omnidireccionales; por ello, el primer paso es transformar las imágenes a panorámicas. Esta transformación es bastante lenta y afectará al coste computacional final. Para utilizar los descriptores, el punto inicial es una imagen panorámica $i(x,y) \in \mathbb{R}^{N_x \times N_y}$ y tras un proceso de descripción glo-

bal el resultado se simplifica a un vector $\vec{d} \in \mathbb{R}^{l \times 1}$.

El robot tiene un movimiento planar restringido al plano del suelo y captura las imágenes con un sistema de visión catadióptrico [13]. El sistema está formado por una cámara y un espejo convexo alineados y montados verticalmente en el robot.

2.1. Firma de Fourier

Este descriptor calcula la transformada discreta de Fourier para cada fila de la imagen panorámica $i(x,y)$ y se obtiene una matriz compleja $F(u,v)$ [6]. F se puede separar en dos matrices.: una matriz de módulos $A(u,v)$ y una matriz de argumentos $\Phi(u,v)$.

Dos imágenes que hayan sido capturadas desde la misma posición, pero con dos orientaciones diferentes tendrán la misma matriz de módulos y matrices de argumentos diferentes. La rotación relativa entre las imágenes se puede calcular [10]. De este modo, $A(u,v)$ puede utilizarse como un descriptor de posición y $\Phi(u,v)$ como descriptor de orientación.

Asimismo, en el dominio de Fourier la información más relevante se concentra en las frecuencias bajas; por ello, se puede eliminar las frecuencias altas, donde habitualmente se almacena ruido. Esto permite minimizar el coste computacional. En este trabajo se usa el parámetro k_1 el cual indica el número de frecuencias retenidas y N_x que indica el número de filas que tiene inicialmente la imagen panorámica. De esta manera la imagen queda reducida a un descriptor $\vec{d} \in \mathbb{R}^{N_x \cdot k_1 \times 1}$.

2.2. Histograma de Orientación del Gradiente

La técnica HOG trabaja con la orientación del gradiente en áreas específicas [2]. Autores como Hofmeister et al. han probado su uso en ambientes pequeños y controlados [5].

Esencialmente consiste en calcular el gradiente de una imagen y calcular orientación y módulo de cada píxel. A continuación, la imagen se divide en un conjunto de celdas horizontales por otro lado se crean b_1 recipientes denominados bins. Dependiendo de la orientación obtenida antes y por celdas horizontales, cada píxel, ponderado por su valor de magnitud de gradiente, se almacenará en un bin distinto. De este modo, dos imágenes panorámicas capturadas en el mismo punto tendrán la misma información en cada celda horizontal. Como resultado, el descriptor es invariante a la rotación del robot en el plano del suelo y puede ser utilizado como un descriptor de apariencia global.

Además, el vector obtenido tiene un tamaño reducido que dependerá del número de celdas y bins. Siendo b_1 en número de bins y k_2 el número de celdas horizontales, el método reduce la información a un descriptor $\vec{d} \in \mathbb{R}^{b_1 \cdot k_2 \times 1}$.

2.3. Gist

Este descriptor fue introducido por Oliva y Torralba [9] y el cual desarrollaron en otro estudio en el que se testeó en entornos de exterior [15]. El descriptor se construye con la información de intensidad conseguida tras convolucionar la imagen con un filtro de Gabor con diferentes orientaciones y niveles de resolución. Finalmente, los datos se guardan como la media del valor de intensidad, como al utilizar HOG, se divide la imagen en celdas horizontales y se calcula la intensidad media de cada celda.

Con este método, y usando m_1 como número de orientaciones del diferentes del filtro de Gabor y k_3 como el número celdas horizontales, la imagen puede reducirse a un vector que dependerá de m_1 , k_3 y el número de diferentes niveles de resolución de la imagen, en este trabajo este parámetro se ha mantenido constante a 2 niveles. Por tanto, el resultado es un descriptor $\vec{d} \in \mathbb{R}^{2 \cdot m_1 \cdot k_3 \times 1}$.

2.4. Transformada de Radon

La transformada de Radon (RT) también puede ser utilizada como descriptor de apariencia global. La transformada fue descrita en 2005 [14] y se ha utilizado en algunas tareas de visión por computador como puede ser la descripción de formas geométricas [4] o en tareas en el campo de la robótica donde RT fue utilizada para encontrar el vecino más cercano en un estimador de posición [1].

La principal ventaja que va a ofrecer este método es la posibilidad de trabajar con la imagen omnidireccional sin necesidad de pasarla a panorámica. Esto puede conllevar a una reducción del coste computacional. El método consiste en aplicar la transformada de Radon a lo largo de grupos de líneas rectas, con una variación de la orientación entre grupos consecutivos. La diferencia de orientación es un parámetro variable denominado p_1 y medido en grados. La transformada de Radon comprime la información de una imagen omnidireccional en una matriz de tamaño $\frac{360}{p_1} \times 0,5 \cdot N_x$ cuya información es invariante a cambios de orientación [1].

Una vez la transformada se ha calculado, el proceso de localización se puede afrontar con diversos métodos. En este trabajo se ha optado por aplicar la firma de Fourier para transformar la información que se tenía tras calcular RT y realizar la localización como se ha descrito en el apartado 2.1.

En este caso, el descriptor tiene un tamaño que depende de p_1 y del número de columnas que se retiene al aplicar FS el cual denominaremos k_4 . Finalmente la imagen omnidireccional se reduce a un vector $\vec{d} \in \mathbb{R}^{\frac{360}{p_1} \cdot 0,5 \cdot N_x \times 1}$.

3. Conjunto de imágenes

Para realizar los experimentos se ha decidido utilizar la base de datos COLD [13]. Esta base de datos proporciona diferentes secuencias de imágenes omnidireccionales tomadas bajo diferentes condiciones de iluminación: soleado, nublado y de noche. Las imágenes fueron adquiridas durante diversos días y bajo cambios en la iluminación así como cambios introducidos por la actividad humana como puede ser objetos o muebles recolocados o personas que aparecen en escena. Estas diferencias pueden observarse en la figuras 1.



(a) Imagen de la base nublada



(b) Imagen de la base soleada



(c) Imagen de la base noche

Figura 1: Imágenes panorámicas de la base de datos COLD.

Por ello, la base de datos COLD es ideal para realizar pruebas de cara a comprobar la eficacia de algoritmos de localización y mapping. La base de datos ofrece diferentes rutas capturadas en entornos de interior a través de distintas universidades. Entre todas ellas se ha elegido la ruta más larga de Friburgo, denominada Part A, Path 2, size 3 [13]. Realizar tareas de navegación autónoma en Friburgo supone un gran reto debido a la cantidad de paredes de cristal y a que hay muchas ventanas al exterior, de este modo las condiciones de luz afectarán sustancialmente a la localización.

La secuencia de imágenes ofrecida ha sido tomada por una cámara omnidireccional montada en un robot móvil. El sistema catadioóptico se construyó utilizando un espejo hiperbólico y una cámara montados en el mismo eje. Además, el robot se equipó con láseres y encoders en las ruedas para detectar la odometría, estos datos se usan para conocer las coordenadas reales en las que se han capturado las imágenes. Estas coordenadas serán utilizadas como ground truth realizándose la localización únicamente con las imágenes omnidireccionales.

Cuadro 1: Distancias

Medida	Distancia	Expresión matemática
d1	Cityblock	$d_1(\vec{a}, \vec{b}) = \sum_{i=1}^l a_i - b_i $
d2	Euclidean	$d_2(\vec{a}, \vec{b}) = \sqrt{\sum_{i=1}^l (a_i - b_i)^2}$
d3	Correlation	$d_3(\vec{a}, \vec{b}) = 1 - \frac{\vec{a}_d^T \cdot \vec{b}_d}{ \vec{a}_d \cdot \vec{b}_d }$
d4	Cosine	$d_4(\vec{a}, \vec{b}) = 1 - \frac{\vec{a}^T \cdot \vec{b}}{ \vec{a} \cdot \vec{b} }$

Donde:

$$\vec{a}_d = [a_1 - \bar{a}, \dots, a_l - \bar{a}]; \bar{a} = \frac{1}{l} \cdot \sum_j \cdot a_j$$

$$\vec{b}_d = [b_1 - \bar{b}, \dots, b_l - \bar{b}]; \bar{b} = \frac{1}{l} \cdot \sum_j \cdot b_j$$

4. EXPERIMENTOS

El propósito de esta sección es comprobar la efectividad de los descriptores de apariencia global. Primero se describe el modelo con el que se determina el error y las diferentes medidas con las que se calcula dicho error. A continuación, se representan los resultados de la localización considerando diferente iluminación y por último se muestra el coste computacional de los experimentos.

4.1. Métodos de creación del modelo

Como se ha explicado en la sección previa, COLD ofrece diferentes bases de datos capturadas bajo diversas condiciones de iluminación. Con la base de datos de Friburgo con tiempo nublado se construye el modelo o mapa en el cual se localizará posteriormente el robot. COLD ofrece una imagen cada 0.06m aproximadamente. Para construir el modelo se reduce la cantidad de imágenes cogiendo una de cada cinco; de esta manera el modelo sigue la misma ruta, pero con una distancia entre imágenes de 0.30m. Esta técnica permite tener un modelo con 556 imágenes en vez de 2778 como la base de datos de nublado (Cloudy Part A, Path 2, size 3) tiene.

Una vez se ha construido el modelo con las imágenes conocidas, el método para dar solución a la tarea de localización es comparar cada nueva imagen con las imágenes del modelo. El programa calcula el descriptor más cercano, para obtener este se puede utilizar diferentes distancias. En este estudio se han utilizado: *cityblock*, *Euclidean*, *correlation* y *cosine*. El cuadro 1 muestra la definición de cada una de estas distancias; donde $\vec{a} \in \mathbb{R}^{l \times 1}$ y $\vec{b} \in \mathbb{R}^{l \times 1}$ son dos vectores en los que: $a_i, b_i, i = 1, \dots, l$.

Cuando se ha obtenido el descriptor más cercano se calcula la distancia geométrica real entre la nueva imagen y la seleccionada como vecino, el resultado será el error de localización. Esta distancia se puede calcular ya que la base de datos COLD ofrece las coordenadas en las que cada imagen había sido capturada. Estas coordenadas solo se han utilizado como ground truth

Cuadro 2: Parámetros con impacto en la localización

Descriptor	Parámetros
FS	$k_1 \Rightarrow$ número de columnas retenidas.
HOG	$b_1 \Rightarrow$ número de bins por histograma. $k_2 F \Rightarrow$ número de celdas horizontales.
gist	$m_1 \Rightarrow$ número de filtros de Gabor. $k_3 \Rightarrow$ número de bloques horizontales.
Radon	$p_1 \Rightarrow$ grados entre pares de líneas. $k_4 \Rightarrow$ número de columnas retenidas.

para controlar el error. La tarea de localización se resuelve íntegramente con información visual.

Los experimentos se realizan utilizando diferentes descriptores y variando los parámetros que influyen en la descripción. De esta manera se puede evaluar la influencia de cada parámetro en la tarea. Estos parámetros se resumen en el cuadro 2. Como la sección 2 muestra, estos parámetros afectan al tamaño del descriptor. Cuanto mayor sea el descriptor, más lento será el proceso.

4.2. Estimación de la posición

Con las 556 imágenes del entorno nublado se ha creado el modelo de referencia. A continuación, para estudiar la eficiencia del descriptor de apariencia global bajo cambios de iluminación, las imágenes usadas para solventar los problemas de localización han sido las bases de datos de soleado y noche, las cuales están compuestas por 2231 y 2896 imágenes respectivamente. El proceso de localización evalúa que imagen del modelo de referencia es más similar a la nueva imagen de test. El error se calcula como la distancia geométrica entre las coordenadas en las que se capturaron las imágenes. Tras realizar el proceso considerando las imágenes soleado como imágenes de test, las figuras 2, 3, 4 y 5 muestran el error medio (m) al variar los parámetros de cada método.

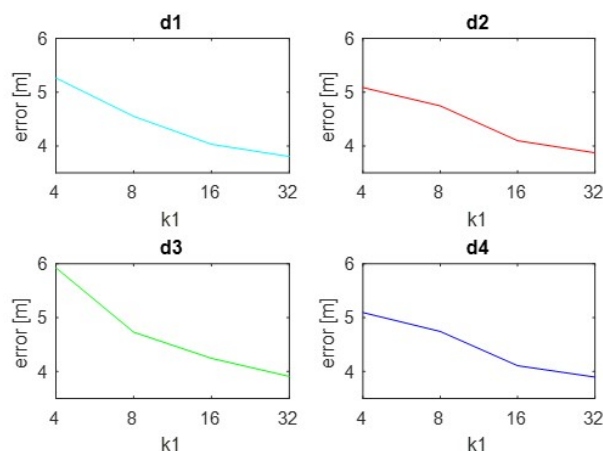


Figura 2: FS. Error de localización (m) utilizando dataset **soleado** como imágenes test al variar k_1 .

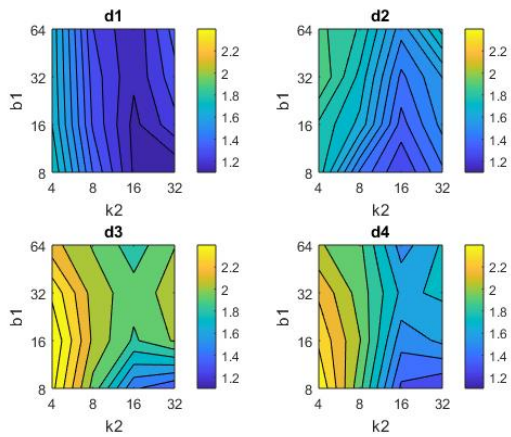


Figura 3: HOG. Error de localización (m) utilizando dataset **soleado** como imágenes test al variar k_2 y b_1 .

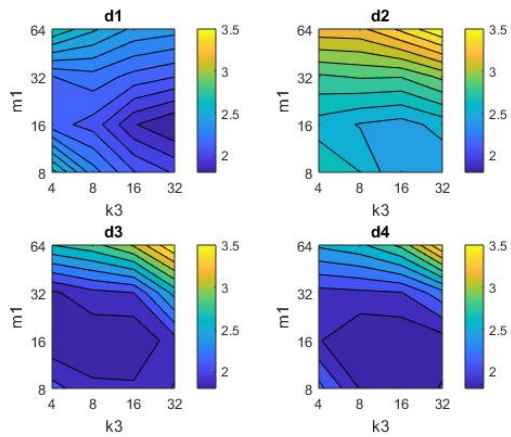


Figura 4: gist. Error de localización (m) utilizando dataset **soleado** como imágenes test al variar k_3 y m_1 .

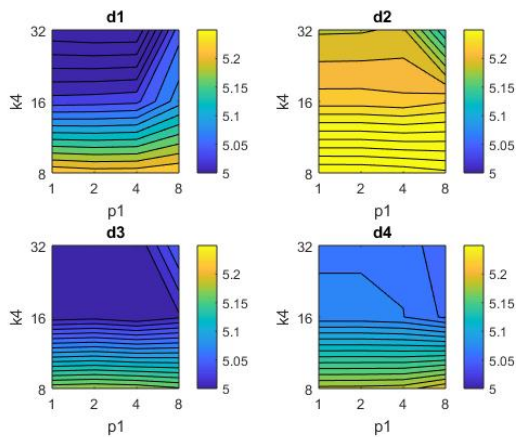


Figura 5: Radon. Error de localización (m) utilizando dataset **soleado** como imágenes test al variar p_1 y k_4 .

El comportamiento de FS varía dependiendo de la medida de distancia empleada. El error varía entre 3 y 7 metros. Cuanto mayor es k_1 mejor es el resultado, pero más lento será el proceso. Se obtienen los mejores resultados con distancia d_3 y $k_1=32$ donde el error de

localización es 3.91m.

Con HOG, valores medio-altos de k_2 e intermedios de b_1 ofrecen los mejores resultados. El error más bajo es 1.09m y se obtiene utilizando d_1 , $b_1=32$ y $k_2=8$.

En general, al utilizar el descriptor gist el error está entorno a 2m. El error baja con valores medios de k_3 y medio-bajos de m_1 . La mejor solución se detecta con d_3 y $k_3=m_1=16$ donde el error es 1.69m.

Por último, al utilizar RT, los resultados obtenidos son peores. El error es superior a 5m en la mayoría de casos. El error depende mucho de k_4 ; cuanto mayor es, menor es el resultado. El error más bajo se obtiene cuando $k_4=32$ y $p_1=1$. Con esos parámetros y d_1 el error es 4.93m.

Analizando globalmente las figuras, HOG y gist son los algoritmos de descripción con un error más bajo que varía entre 1 y 2 metros cuando se utiliza d_1 y d_3 .

Analizando globalmente las figuras, HOG y gist son los algoritmos de descripción con un error más bajo que varía entre 1 y 2 metros cuando se utiliza d_1 y d_3 . Tras estos experimentos, se repite el proceso de localización considerando las imágenes de noche como base de datos test. Como en el caso anterior, el modelo de referencia son las 556 imágenes nubladas. Las figuras 6, 7, 8 and 9 presentan los errores medios en metros.

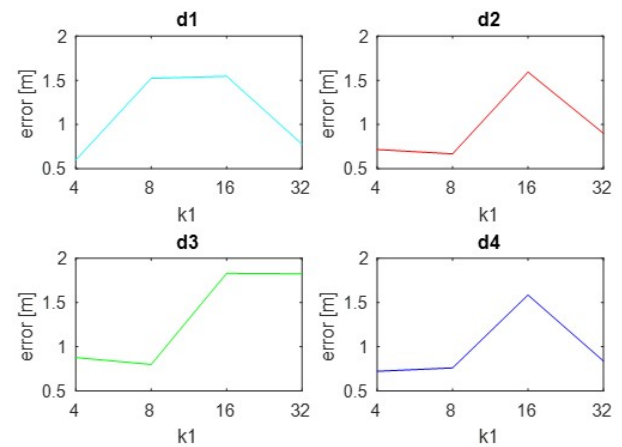


Figura 6: FS. Error de localización (m) utilizando dataset **noche** como imágenes test al variar k_1 .

Los resultados al utilizar FS tienen un error entre 0.6 y 2.5m. Valores bajos del parámetro k_1 ofrecen errores bajos, se obtiene el mejor con $k_1=4$ y la distancia d_1 donde el error es 0.597m.

El comportamiento con HOG presenta un error entre 0.2 y 0.6m. El mejor resultado se da con valores intermedios de k_2 and b_1 y la distancia d_1 . Con $k_2=32$, $b_1=16$ y d_1 se obtiene un error de 0.189m

Con gist el error medio está entre 0.2 y 1m. Si k_3 o m_1 son bajos, el error aumenta. Además, valores altos de m_1 hacen que el error crezca; así que valores medios

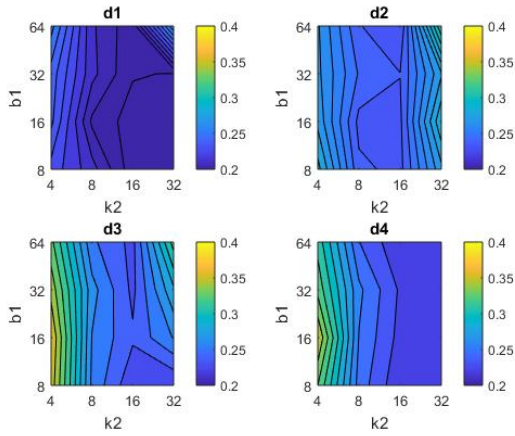


Figura 7: HOG. Error de localización (m) utilizando dataset **noche** como imágenes test al variar k_2 y b_1 .

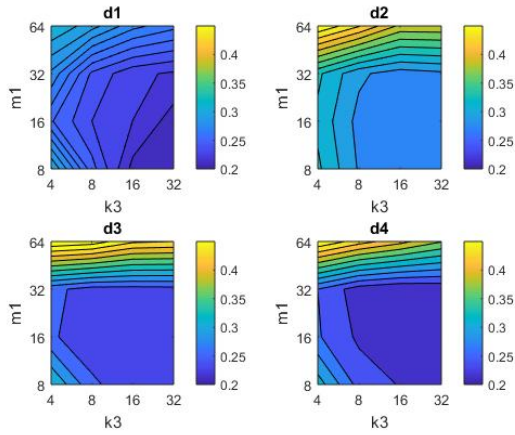


Figura 8: gist. Error de localización (m) utilizando dataset **noche** como imágenes test al variar k_3 y m_1 .

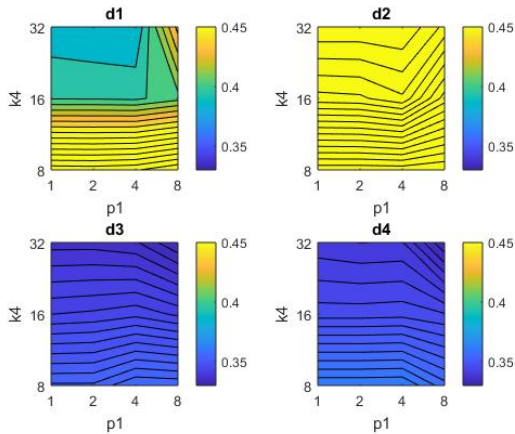


Figura 9: Radon. Error de localización (m) utilizando dataset **noche** como imágenes test al variar p_1 y k_4 .

de ambos parámetros es la mejor opción. El error más bajo se obtiene con distancia $d1$, $k_3=32$ y $b_1=8$, donde el error es 0.21m.

Finalmente, el error de localización con Radon está en-

tre 0.4 y 1.8m. Cuanto menor es el parámetro k_4 más bajo es el error; p_1 tiene menos influencia en el resultado. Cuando $k_4=32$, $p_1=1$ y $d3$ el error de localización es 0.328m.

El descriptor que mejor resultado ofrece es HOG y las distancias $d1$ y $d3$ las que mejor funcionan. Los errores oscilan alrededor del metro, resultados remarquables teniendo en cuenta los cambios de iluminación, los movimientos del mobiliario y las oclusiones que aparecen en las escenas.

4.3. Coste computacional

Además de tener un error bajo, también es importante tener un tiempo de computación razonable. Por ello, se ha estudiado el tiempo que tarda el proceso y de este modo saber si la localización se puede hacer en tiempo real. Las figuras 10, 11, 12 y 13 muestran el tiempo necesario para estimar la posición del robot. Los datos se dan en segundos.

FS ofrece los resultados más bajos. El tiempo más alto se obtiene con la distancia $d3$, teniendo un crecimiento exponencial proporcional a k_1 .

Los tiempos con HOG tienen un orden similar a los de FS. El resultado óptimo es 0.221s cuando $k_2=4$, $b_1=8$. El resultado es más lento al aumentar el valor de los parámetros, el parámetro b_1 tiene una influencia mayor.

Gist no ofrece un resultado rápido comparado con el resto de métodos. Al utilizar gist el tiempo de ejecución puede llegar a ser de 0.5 segundos. En este caso el parámetro m_1 es el que mayor influencia tiene. El menor tiempo se ha obtenido con $m_1=8$ y $k_3=4$, con un tiempo asociado de 0.2405s.

Finalmente se estudia el tiempo con RT. Se observa una alta dependencia de p_1 . Cuanto menor es, mayor es el coste computacional. El mayor tiempo obtenido es 0.447s con $p_1=1$ y $k_4=16$ pero RT puede ofrecer tiempos muy bajos como 0.059s cuando $p_1=8$ y $k_4=4$.

El tiempo empleado para pasar la imagen a panorámica tiene mucha relevancia en el tiempo final. Como resultado se puede destacar unos tiempos más adecuados con FS y HOG y un tiempo de rutina más lento con gist. En el proceso RT no es necesario transformar la imagen a panorámica, de este modo el tiempo se reduce y por ello la oscilación en el resultado es tan grande.

Los experimentos han sido llevados a cabo en un PC con CPU Quad-Core Intel i7-700® a 3.60 GHz y programando en Matlab®. Estos resultados no son absolutos, si no que variarían con la máquina con la que se realiza el proceso. De todas formas, los resultados sí que son comparables ya que se han calculado con la misma máquina.

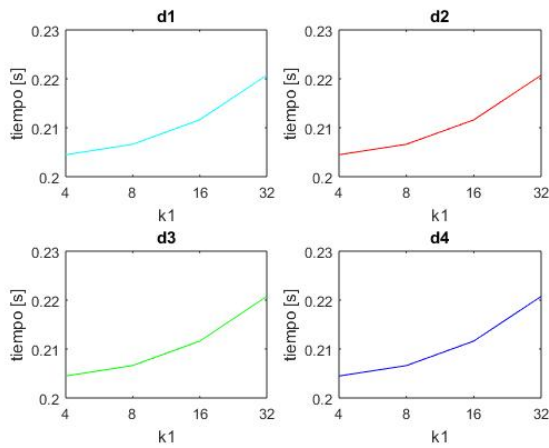


Figura 10: FS. Tiempo de localización (s) por imagen al variar k_1 .

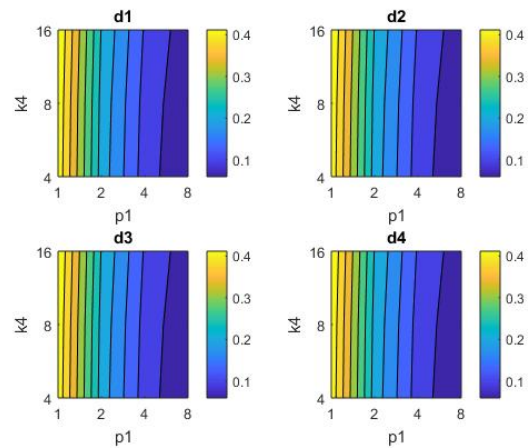


Figura 13: Radon. Tiempo de localización (s) por imagen al variar p_1 y k_4 .

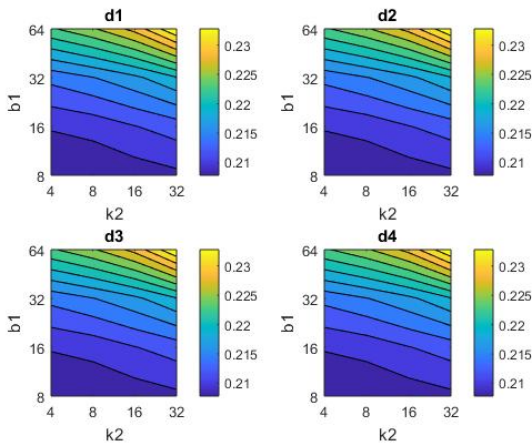


Figura 11: HOG. Tiempo de localización (s) por imagen al variar k_2 y b_1 .

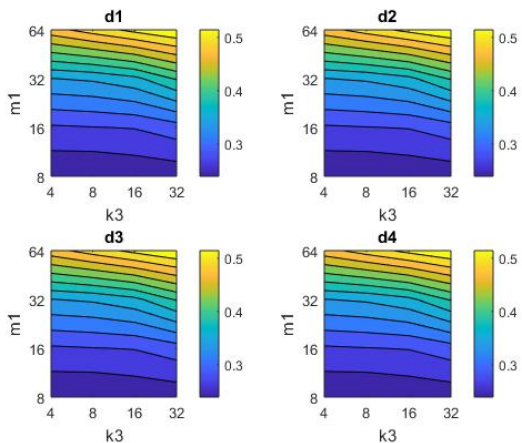


Figura 12: gist. Tiempo de localización (s) por imagen al variar k_3 y m_1 .

5. CONCLUSIONES

El trabajo se ha centrado en estudiar los problemas que los cambios de iluminación y la actividad humana pueden introducir en la tarea de navegación robótica, haciendo uso únicamente de información visual. Se ha comparado cuatro algoritmos de apariencia global. Una vez realizada la descripción de las imágenes y calculado el error de localización se realiza una evaluación del coste computacional.

Por un lado, la firma de Fourier y HOG ofrecen resultados más rápidos, por lo que pueden ser una opción más adecuada para tareas en tiempo real. Por otro lado, HOG y gist ofrecen unos resultados con mayor precisión en las tareas de localización.

Este trabajo abre la puerta a nuevas aplicaciones y estudios basados en métodos de apariencia global dentro del mundo de la robótica móvil. Una vez testada la eficacia de los métodos con variaciones de iluminación y cambios introducidos por la actividad humana, el siguiente paso es crear un método que continuamente renueve el modelo adaptándolo a nuevas condiciones de iluminación. Esto puede llevar también a una aproximación a un sistema en el que se creen mapas más sofisticados que faciliten un sistema de navegación autónoma.

Agradecimientos

Este trabajo ha sido apoyado por el gobierno español a través del proyecto DPI 2016-78361-R (AEI/FEDER, UE): “Creación de mapas mediante métodos de apariencia visual para la navegación de robots”.

Evaluating the robustness of global appearance descriptors in a visual localization task, under changing lighting conditions.

Abstract

Map building and localization are two important abilities that autonomous mobile robots must develop. This way, much research has been carried out on these topics, and researchers have proposed many approaches to address these problems. This work presents some possibilities to solve these problems robustly using global appearance descriptors. In this way, robots capture visual information from the environment and obtain, usually by means of a transformation, a global appearance description for each whole image. Using these descriptors, the robot is able to estimate its position in a previously built map. In this work, the global appearance descriptors' performance at the localization task inside a known environment has been studied. In the task we consider that the map is already built and the goal is to evaluate the descriptors' robustness when the environment visual appearance changes substantially. To achieve this objective, a comparative evaluation of several global appearance descriptors is carried out.

Keywords: Mobile robots, global appearance descriptors, omnidirectional images, mapping, localization.

Referencias

- [1] BERENGUER, Y., PAYÁ, L., PEIDRÓ, A., GIL, A., AND REINOSO, O. Nearest position estimation using omnidirectional images and global appearance descriptors. In *Robot 2015: Second Iberian Robotics Conference* (2016), Springer, pp. 517–529.
- [2] DALAL, N., AND TRIGGS, B. Histograms of oriented gradients for human detection. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)* (June 2005), vol. 1, pp. 886–893 vol. 1.
- [3] GIL, A., MOZOS, O. M., BALLESTA, M., AND REINOSO, O. A comparative evaluation of interest point detectors and local descriptors for visual slam. *Machine Vision and Applications* 21, 6 (2010), 905–920.
- [4] HOANG, T. V., AND TABBONE, S. A geometric invariant shape descriptor based on the radon, fourier, and mellin transforms. In *Pattern Recognition (ICPR), 2010 20th International Conference on* (2010), IEEE, pp. 2085–2088.
- [5] HOFMEISTER, M., LIEBSCH, M., AND ZELL, A. Visual self-localization for small mobile robots with weighted gradient orientation histograms. In *40th International Symposium on Robotics (ISR)* (2009), Barcelona, pp. 87–91.
- [6] MENEGATTI, E., MAEDA, T., AND ISHIGURO, H. Image-based memory for robot navigation using properties of omnidirectional images. *Robotics and Autonomous Systems* 47, 4 (2004), 251 – 267.
- [7] MURILLO, A. C., GUERRERO, J. J., AND SAGUES, C. Surf features for efficient robot localization with omnidirectional images. In *Robotics and Automation, 2007 IEEE International Conference on* (2007), IEEE, pp. 3901–3907.
- [8] OLIVA, A., AND TORRALBA, A. Modeling the shape of the scene: A holistic representation of the spatial envelope. *International journal of computer vision* 42, 3 (2001), 145–175.
- [9] OLIVA, A., AND TORRALBA, A. Building the gist of a scene: The role of global image features in recognition. *Progress in brain research* 155 (2006), 23–36.
- [10] PAYÁ, L., FERNÁNDEZ, L., REINOSO, Ó., GIL, A., AND ÚBEDA, D. Appearance-based dense maps creation-comparison of compression techniques with panoramic images. In *ICINCO-RA* (2009), pp. 250–255.
- [11] PAYÁ, L., MAYOL, W., CEBOLLADA, S., AND REINOSO, O. Compression of topological models and localization using the global appearance of visual information. In *Robotics and Automation (ICRA), 2017 IEEE International Conference on* (2017), IEEE, pp. 5630–5637.
- [12] PAYÁ, L., REINOSO, O., BERENGUER, Y., AND ÚBEDA, D. Using omnidirectional vision to create a model of the environment: A comparative evaluation of global-appearance descriptors. *Journal of Sensors* 2016 (2016).
- [13] PRONOBIS, A., AND CAPUTO, B. COLD: COsy Localization Database. *The International Journal of Robotics Research (IJRR)* 28, 5 (May 2009), 588–594.
- [14] RADON, J. 1.1 über die bestimmung von funktionen durch ihre integralwerte längs gewisser mannigfaltigkeiten. *Classic papers in modern diagnostic radiology* 5 (2005), 21.
- [15] SIAGIAN, C., AND ITTI, L. Biologically inspired mobile robot vision localization. *IEEE Transactions on Robotics* 25, 4 (2009), 861–873.
- [16] VALGREN, C., AND LILIENTHAL, A. J. Sift, surf & seasons: Appearance-based long-term localization in outdoor environments. *Robotics and Autonomous Systems* 58, 2 (2010), 149–156.