Detección de Lugares con Cámaras RGB-D. Aplicación a Cierre de Bucles en SLAM

David Zúñiga-Noël, José-Raúl Ruiz-Sarmiento, Javier Gonzalez-Jimenez
Machine Perception and Intelligent Robotics Group, Departamento de Ingeniería de Sistemas y Automática, Universidad de Málaga, Campus de Teatinos, 29071, Málaga
{dzuniga,jotaraul,javiergonzalez}@uma.es

Resumen

En este trabajo se propone un método que combina descriptores de imágenes de intensidad y de profundidad para detectar de manera robusta el problema de cierre de bucle en SLAM. La robustez del método, proporcionada por el empleo conjunto de información de diversa naturaleza, permite detectar lugares revisitados en situaciones donde métodos basados solo en intensidad o en profundidad presentan dificultades (e.g. condiciones de iluminación deficientes, o falta de geometría). Además, se ha diseñado el método teniendo en cuenta su eficiencia, recurriendo para ello al detector FAST para extraer las características de las observaciones y al descriptor binario BRIEF. La detección de bucle se completa con una Bolsa de Palabras binarias. El rendimiento del método propuesto se ha evaluado en condiciones reales, obteniéndose resultados muy satisfactorios.

Palabras clave: Detección de lugares, Cámaras RGB-D, SLAM, Cierre de bucle.

1. INTRODUCCIÓN

La construcción de mapas métricos es una tarea clave para agentes que deseen localizarse o navegar en su entorno, como es el caso de los robots móviles [23, 8, 9]. Durante la construcción de dichos mapas se realizan estimaciones en las que influyen el ruido, *aliasing*, y otras distorsiones producidas por el sensor utilizado para percibir el entorno, dando lugar a imprecisiones que se acumulan conforme el proceso de reconstrucción avanza.

Un modo de mitigar estos errores es detectar cuándo el agente ha regresado a una región del entorno previamente visitada, para lo que es necesario que este cuente con la habilidad de reconocer dichos lugares. Cuando el reconocimiento es aplicado al problema de *SLAM* (Simultaneous Localization and Mapping) se le denomina *detección de cierre de bucle* [3, 23], y es una cuestión ampliamente estudiada por su relevancia para la construcción de mapas métricos consistentes.

En la actualidad, el uso de sensores que proveen información visual de intensidad (cámaras) es el más extendido para implementar esta capacidad, dado su bajo coste y la riqueza de la información que proporcionan. Para reconocer una zona previamente visitada, la información obtenida se describe y compara en base a una serie de características distintivas. Típicamente, la técnica de *Bolsa de Palabras* [20] es la elegida para abordar problemas de reconocimiento de lugares u objetos, dado que permite representar las características de cada imagen con un vector numérico que puede compararse eficientemente con otros vectores [5]. En dicho vector se codifica información sobre una serie de palabras extraídas de un *vocabulario* previamente definido. Si bien este enfoque es aplicable a una multitud de escenarios y escalable a grandes cantidades de datos [13], su rendimiento y robustez decaen en entornos con información visual insuficiente (como por ejemplo: falta de texturas o condiciones de iluminación muy cambiantes o deficientes).

En un intento por paliar esta limitación se pueden sustituir las cámaras convencionales por cámaras RGB-D que, además de la tradicional información de intensidad, también proporcionan información de profundidad [16, 18, 17, 11]. Esto permitiría aprovechar, además de información visual, información geométrica, que no depende de la iluminación externa [24]. No obstante, disponer de sendas fuentes de información independientes conlleva dos problemas fundamentales. De un lado, un mayor coste computacional, lo que puede comprometer su utilización en aplicaciones donde el mapeado deba realizarse en tiempo real. Por otro, se hace necesario adaptar los métodos de reconocimiento a la combinación de características de muy distinta naturaleza.

En este artículo se propone un método de reconocimiento robusto de lugares previamente visitados, que combina información tanto de intensidad como de profundidad, obtenida por medio de un sensor RGB-D. Para ello, y teniendo la eficiencia como requisito en el diseño del método, en una primera etapa se extrae de ambas fuentes de información una serie características empleando

do el conocido detector FAST [15]. Nuestra contribución se basa en la posterior utilización de BRIEF [4] para obtener descripciones binarias de dichas características y aprovecharlas para crear una representación unificada de una observación RGB-D mediante una Bolsa de Palabras binarias [7]. La utilización de estas descripciones binarias conlleva, a la hora de su generación y comparación, un bajo coste tanto de almacenamiento como computacional. Finalmente, se buscan similitudes entre la representación de la observación actual y las de observaciones anteriores almacenadas en una base de datos. Esto permite a un agente detectar si se encuentra en una región previamente visitada, y actuar en consecuencia.

En los experimentos de validación llevados a cabo se han utilizado secuencias de observaciones RGB-D públicamente disponibles. Por un lado, para la creación del vocabulario de la Bolsa de Palabras se han utilizado secuencias del conjunto de datos *RGB-D SLAM Dataset* [22], de la Universidad Técnica de Múnich. Por otro lado, para la evaluación del método se han grabado secuencias exigentes con el fin de mostrar las virtudes del método propuesto, obteniéndose resultados muy satisfactorios.

2. TRABAJOS RELACIONADOS

En la literatura se pueden encontrar trabajos empleando la combinación información de intensidad - Bolsa de Palabras para el reconocimiento de lugares. Un ejemplo es el sistema FAB-MAP, propuesto por Cummins y Newman [5], que emplea información de intensidad obtenida mediante cámaras omnidireccionales y descrita empleando SURF [2], y una Bolsa de Palabras para detectar bucles en recorridos de varios kilómetros de distancia en exteriores. Otro enfoque, que ha servido de inspiración para este artículo, es el uso de descriptores binarios en lugar de SURF, propuesto por Gálvez y Tardós [7]. Sin embargo, aunque estos métodos de detección consiguen un rendimiento destacable, su eficacia puede verse comprometida en entornos en los que la información visual capturada sea insuficiente, e.g. pocas texturas, condiciones de iluminación adversas, etc.

Una alternativa consiste en utilizar información de profundidad, puesto que, entre otras cosas, no depende de fuentes externas de iluminación. Por ejemplo, Steder *et. al.* [21] utilizan información obtenida mediante un escáner láser 3D para formar imágenes de rango de 360º, con el fin de detectar lugares previamente visitados en exteriores. Por otro lado, Scherer *et. al.* [18] evalúan el rendimiento de las Bolsas de Palabras para distintos descriptores, entre ellos BRIEF [4], sobre imágenes de profundidad en interiores. En general, estas alternativas suelen ser menos eficaces, puesto que es más difícil distinguir lugares utilizando únicamente información de esta naturaleza, pero son una opción válida si la información de apariencia en el entorno es pobre.

También existen descriptores capaces de combinar información de intensidad y profundidad, tratando de potenciar las virtudes de ambas y mitigar sus limitaciones. Concretamente, BRAND [12] es un descriptor binario capaz de combinar eficientemente estas dos fuentes de información, y ha sido utilizado en el reconocimiento de lugares por Zhang *et. al.* [24]. En dicho trabajo se emplea una técnica alternativa a la Bolsa de Palabras, *Locality Sensitive Hashing*, la cual ha obtenido una menor eficiencia en estudios como los realizados por Shahbazi y Zhang [19]. El método aquí propuesto combina información de intensidad y profundidad en una única Bolsa de Palabras binarias [7], buscando ser eficiente a la hora de crear representaciones de las imágenes empleando dicha bolsa, y que permitan detectar si se ha cerrado un bucle. Así mismo, el empleo del extractor de características FAST [15], y el descriptor binario BRIEF, permiten reducir el tiempo de ejecución y el espacio de almacenamiento requerido.

3. MÉTODO PROPUESTO

La Fig. 1 muestra el flujo de trabajo del método propuesto. Brevemente, con la llegada de nuevas observaciones de intensidad y profundidad, y tras un pre-procesamiento inicial (Sec. 3.1), se detectan y describen una serie de puntos de interés (Sec. 3.2, Sec. 3.3). Estas descripciones, junto con un vocabulario de palabras binarias previamente creado, se emplean para construir una representación conjunta de ambas imágenes mediante la Bolsa de Palabras (Sec. 3.4). Finalmente, dicha representación es cotejada con una base de datos (Sec. 3.5) que contiene representaciones anteriores para detectar si se ha completado un bucle (Sec. 3.6). Las siguientes secciones describen en más detalle los componentes principales del método.

3.1. PRE-PROCESAMIENTO

Para crear una única Bolsa de Palabras a partir de dos fuentes de información, se ha optado por trabajar con la imagen de rango de la información de profundidad, y aplicar sobre ella las técnicas tradicionales de Visión por Computador (detectores de puntos de interés y descriptores de los mismos) [18]. Esta imagen de rango puede considerarse una imagen en escala de grises, donde un

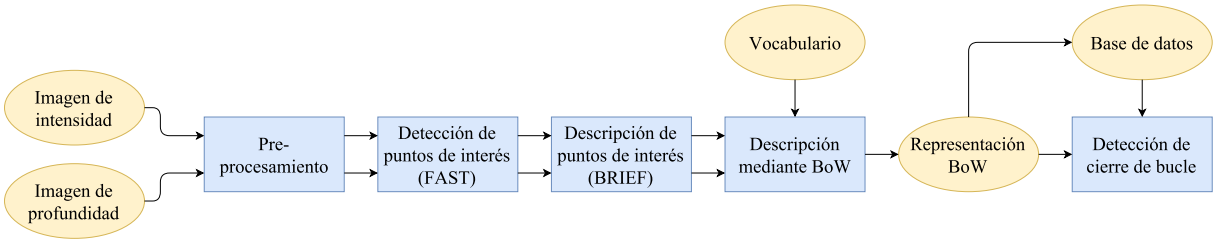

Figura 1: Flujo de información y procesos empleados por el método propuesto. Las formas rectangulares representan procesos, mientras que las ovaladas información consumida o producida por dichos procesos.

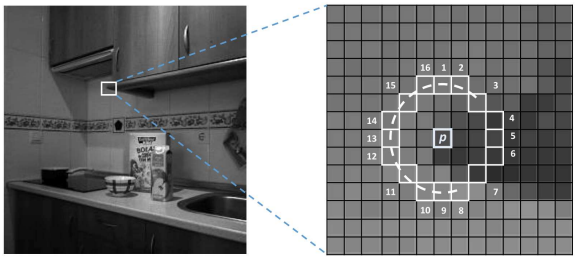

Figura 2: Ejemplo de comprobación de punto de interés por medio de FAST. Las celdas blancas representan los píxeles cuya intensidad se va a comparar con la del punto $\mathbf{p}$, extendiéndose el círculo punteado sobre los píxeles con menor intensidad que este.

píxel toma un valor bajo cuando representa una medición cercana, y va aumentando con la distancia. Es bien sabido que las mediciones de profundidad de un sensor RGB-D son más propensas a error conforme más lejanas son [16], por lo que en este trabajo se han considerado no válidas mediciones superiores a 5 metros.

Por otro lado, la imagen de intensidad también es transformada a escala de grises aplicando la fórmula:

$$I(\mathbf{p}) = 0{,}299R(\mathbf{p}) + 0{,}587G(\mathbf{p}) + 0{,}114B(\mathbf{p}) \quad (1)$$

para cada píxel $\mathbf{p}$, a partir de sus componentes R, G y B, de acuerdo con la *Recomendación 601* [10].

3.2. DETECCIÓN DE PUNTOS DE INTERÉS

Para seleccionar los puntos de interés se utiliza el detector de esquinas *Features from Accelerated Segment Test* (FAST) [15]. Este detector busca cambios de intensidad en un círculo de *Bresenham* de 16 píxeles de longitud, centrado en un punto $\mathbf{p}$ con intensidad $I(\mathbf{p})$ del cual se quiere determinar si es un punto de interés o no. Para que $\mathbf{p}$ sea considerado esquina, tendrá que haber en dicho círculo un número n de píxeles consecutivos (típicamente $n = 12$) tal que todos ellos tengan una intensidad bien mayor o bien menor que $I(\mathbf{p})$ (Fig. 2). Para realizar la detección de manera eficiente, primero se compara la intensidad de ciertos píxeles estratégicos, de tal manera que el punto $\mathbf{p}$ se pueda descartar rápidamente.

3.3. DESCRIPCIÓN DE PUNTOS DE INTERÉS

Con el objetivo de mantener al mínimo el tiempo de ejecución en la fase de descripción de características (y las posteriores comparaciones) se ha utilizado el descriptor BRIEF [4]. Este método, dado un punto de interés $\mathbf{p}$, genera una secuencia $\mathbf{B}(\mathbf{p})$ de L bits para describir la región cuadrada de lado S al rededor de dicho punto. En concreto, se tiene que:

$$B_i(\mathbf{p}) = \begin{cases} 1 & \text{si } I(\mathbf{a}_i) < I(\mathbf{b}_i) \\ 0 & \text{en otro caso} \end{cases} \quad (2)$$

para $i \in [1..L]$. Los pares de puntos $\mathbf{a}_i$ y $\mathbf{b}_i$ para los que se realizan las comparaciones se generan previamente de manera aleatoria. Para seleccionarlos, se ha seguido el proceso descrito en [6].

El tamaño del descriptor se ha fijado en $L = 256$, y el lado de la región cuadrada a $S = 48$, por haber mostrado una buena relación entre distinción y coste computacional [4].

3.4. DESCRIPCIÓN MEDIANTE BOLSAS DE PALABRAS

El proceso de descripción mediante Bolsas de Palabras requiere la creación de un vocabulario previo. Para definirlo, se establece una jerarquía o árbol a partir de las descripciones BRIEF extraídas de un conjunto de observaciones dado, discretizando de este modo el espacio de descripciones en $W = k^d$ palabras o nodos hoja.

Para ello, las descripciones se dividen en k grupos, mediante la técnica de las k-medias [1], formando así el primer nivel. Esta técnica, además de agrupar las descripciones, provee una descripción representativa de cada grupo, que es asociada a

un nodo. De forma recursiva, se repite el proceso para cada grupo, hasta un máximo de d veces.

De esta manera, para representar una imagen mediante una Bolsa de Palabras a partir de sus descripciones binarias, se recorre el árbol desde la raíz hasta las hojas para cada punto descrito, seleccionándose en cada nivel el nodo que minimice la *distancia de Hamming*. Como resultado de este proceso se obtiene un vector $\mathbf{v} \in \mathbb{R}^W$, al que llamaremos *representación BoW* (o simplemente *representación*), que se corresponde con un histograma pesado conforme a la frecuencia con la que aparecen las palabras tanto en la imagen como en el propio vocabulario [20].

En este trabajo, los descriptores extraídos de las imágenes de intensidad y de profundidad son considerados provenientes de una misma fuente de información, y como tal son utilizadas para generar el vocabulario y realizar las siguientes comparaciones a partir de una observación RGB-D.

3.5. BASE DE DATOS

La base de datos está compuesta por las representaciones BoW. Su fin es servir de repositorio para determinar si una nueva observación está ya incluida en esta base de datos. Para ello, es necesario obtener una medida de similitud entre dos representaciones.

Dadas las representaciones $\mathbf{v}_1$ y $\mathbf{v}_2$, su similitud se calcula utilizando la fórmula:

$$s(\mathbf{v}_1, \mathbf{v}_2) = 1 - \frac{1}{2}\left| \frac{\mathbf{v}_1}{|\mathbf{v}_1|_1} - \frac{\mathbf{v}_2}{|\mathbf{v}_2|_1} \right|_1 \qquad (3)$$

Para acelerar el proceso de búsqueda en la base de datos se mantiene un índice inverso, es decir, a partir de una palabra dada, se pueden recuperar las representaciones que la contienen. Esto permite comparar únicamente observaciones que tengan alguna palabra en común, dotando de eficiencia al proceso, tal y como se muestra en la sección de evaluación.

3.6. DETECCIÓN DE BUCLES

A la hora de comparar la representación BoW de una observación entrante con las ya existentes en la base de datos, se genera una serie de mediciones de similitud. En primer lugar, de entre ellas, se descartan las que se correspondan con observaciones demasiado próximas temporalmente a la observación consultada, puesto que, aún siendo muy similares, obviamente no constituyen un bucle real[1].

Dado que es difícil establecer un umbral fijo a partir del cual se considere que una medición de similitud es suficiente para corresponderse con la detección de un bucle, esta medida se normaliza dividiendo por una aproximación de la mejor medición o puntuación que se espera obtener. Dicha aproximación se calcula como la similitud entre la observación actual y la anterior. De esta manera, la puntuación final entre dos vectores $\mathbf{v}_t$ y $\mathbf{v}_s$ se obtiene como:

$$\eta(\mathbf{v}_t, \mathbf{v}_s) = \frac{s(\mathbf{v}_t, \mathbf{v}_s)}{s(\mathbf{v}_t, \mathbf{v}_{t-1})} \qquad (4)$$

donde $\mathbf{v}_t$ representa la descripción BoW de la observación actual, $\mathbf{v}_{t-1}$ la de la anterior, y $\mathbf{v}_s$ la de otra observación pasada. Si esta similitud normalizada es mayor que un cierto umbral α, se considera que se ha detectado un bucle.

Por último, con el fin de que sea objetivo de futuras detecciones de bucle, la representación $\mathbf{v}_t$ se añade a la base de datos, actualizándose también el índice inverso de la misma.

4. EVALUACIÓN

Con el objetivo de comprobar la eficacia y eficiencia del método propuesto se han conducido una serie de experimentos, presentándose aquí su metodología (Sec. 4.1) y resultados obtenidos (Sec. 4.2).

4.1. METODOLOGÍA

Dadas las particularidades de las fases de creación de vocabulario, entrenamiento del método, y evaluación del mismo, se han utilizado secuencias independiente en cada una de ellas. Además, todos los conjuntos de datos empleados están disponibles al público[2]. Concretamente:

Vocabulario: Para crear el vocabulario, se han utilizado observaciones de un conjunto de datos ofrecido públicamente por la Universidad Técnica de Múnich [22], concretamente las secuencias llamadas `fr1/room`, `fr2/desk` y `fr3/long_office_household`.

Entrenamiento: Con el objetivo de ajustar los distintos parámetros del método para el vocabulario creado, se han diseñado y recogido

[1] Típicamente los métodos de reconocimiento de lugares realizan una serie de comprobaciones de consistencia adicionales para descartar falsas detecciones, e.g. verifican si la localización de las características en las imagen de intensidad es similar. Aquí omitimos estas comprobaciones para poder obtener el rendimiento del método de reconocimiento por si mismo.

[2] `http://mapir.isa.uma.es/mapirwebsite/index.php/mapir-downloads/papers/235`

Tabla 1: Evaluación del rendimiento del método propuesto para detectar cierre de bucles (*Combinado*), junto con dos variantes (*Intensidad* y *Profundidad*). Resaltados los mejores resultados para cada secuencia.

Método	`lab`		`home`		`test`	
	Acierto	Precisión	Acierto	Precisión	Acierto	Precisión
Intensidad	77.08 %	**90.24** %	74.67 %	**81.3** %	32.10 %	25.81 %
Profundidad	85.42 %	77.36 %	68.09 %	73.36 %	36.39 %	33.96 %
Combinado	**87.5** %	80.76 %	**75.10** %	78.99 %	**43.55** %	**34.08** %

dos secuencias (llamadas `lab` y `home`, contando con 238 y 1292 observaciones respectivamente), que contienen una distribución equilibrada de información de intensidad y profundidad.

Evaluación: Se ha grabado una secuencia adicional (denominada `test`, compuesta por 1478 observaciones), que contiene bucles en regiones con pobre iluminación, con solo textura, y largos recorridos sin bucles, diseñada para evaluar el método propuesto.

El método presentado (*Combinado*) se compara con dos variantes (*Intensidad* y *Profundidad*), en las que se utiliza solamente una fuente de información de las dos disponibles durante todo el proceso (incluida la creación del vocabulario). Para poder evaluar la eficacia del método, se emplean los valores de acierto y precisión, tal y como se describen en [7]. El acierto se calcula como la razón entre el número de detecciones correctas y el número total de bucles existentes en la secuencia. La precisión se calcula como la razón entre el número detecciones correctas y el número de bucles detectados.

Para que el proceso de evaluación pueda llevarse a cabo de manera automática, y dado que no se cuenta con información sobre la localización del sensor a lo largo de la secuencia, se ha definido manualmente para cada observación un conjunto de observaciones anteriores a ella con las que cierra un bucle real, pudiéndose calcular de esta manera las medidas cuantitativas antes definidas.

4.2. RESULTADOS

Durante el ajuste de los parámetros del método se concluyó que el tamaño de vocabulario óptimo para los datos manejados es de $k = 10$ y $d = 5$ para la variante de *Intensidad*, $k = 5$ y $d = 5$ para la de *Profundidad*, y $k = 10$ y $d = 4$ para el método *Combinado*. Además, el valor del umbral α escogido para cada método es $\alpha = 0{,}5$; $\alpha = 0{,}8$; y $\alpha = 0{,}7$ respectivamente. Con estos parámetros, se muestran a continuación sus resultados en cuanto a eficacia y eficiencia.

4.2.1. Eficacia

La Tab. 1 muestra los resultados en cuanto a acierto y precisión obtenidos en tres secuencias: `lab`, `home` y `test`. Como se puede comprobar, no hay un método ganador en los tres casos, lo cual se debe a las peculiaridades de cada secuencia. Por un lado, `lab` es una secuencia grabada en un entorno de oficinas, rico en información tanto de apariencia como geométrica, por lo que el método propuesto no marca una diferencia: aunque se obtienen resultados ligeramente mejores que la variante de *Profundidad*, la de *Intensidad* consigue una precisión mayor en un $\sim$10 %, si bien a costa de un acierto menor en el mismo porcentaje.

La secuencia `home` se obtuvo en un entorno doméstico con condiciones de iluminación favorables, donde la información visual es más notoria y discriminativa. Esto se ve reflejado en los resultados de la variante *Intensidad*, que obtiene un acierto del $\sim$74 % y una precisión del $\sim$81 %. El método *Combinado* alcanza unos resultados similares, pero el basado en *Profundidad* ve mermado su rendimiento.

Por último, la secuencia `test` es la más desafiante y realista de todas, conteniendo un mayor número de observaciones y bucles de distinta naturaleza. Esta secuencia simula el recorrido de un agente en un entorno doméstico, desplazándose por regiones con distintas condiciones de iluminación y configuraciones geométricas. Es en estos casos donde el rendimiento de las variantes *Intensidad* y *Profundidad* puede verse comprometido, tal y como se refleja en los resultados obtenidos. Aquí, el método propuesto es más robusto frente a la detección de bucles en condiciones adversas, obteniendo mejores resultados que las dos variantes. A modo de ejemplo, la Fig. 3 muestra a la izquierda una región donde la información de intensidad es pobre, debido principalmente a las condiciones de iluminación, mientras que la información geométrica es distintiva. El método *Combinado* ha sido capaz de detectar un cierre de bucle en esa región, mientras que la variante de *Intensidad* falla. A la derecha en la Fig. 3 se muestra una región en la que se

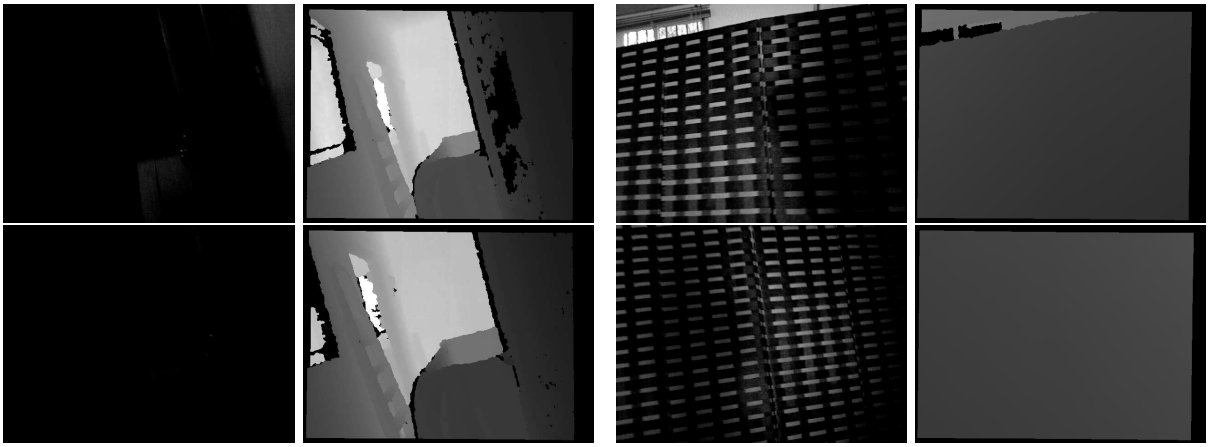

Figura 3: Ejemplos de detecciones de cierre de bucle obtenidas por el método propuesto. En la primera fila se presentan observaciones en las que se ha detectado un bucle con las mostradas en la segunda fila. A la izquierda se expone una región con pobre iluminación pero rica en geometría, mientras que a la derecha se puede apreciar una zona plana pero que presenta características visuales.

Tabla 2: Media y desviación típica del tiempo de ejecución de cada uno de los procesos del método.

Componente	Tiempo de ejecución	
	Media	Desviación
Puntos de interés	2.01 ms	1.02 ms
Descripción binaria	12.53 ms	6.96 ms
Descripción BoW	15.12 ms	10.18 ms
Detección de bucle	21.54 ms	21.11 ms

da el caso opuesto: geometría prácticamente nula pero información visual distintiva. En esta última región, la variante de *Profundidad* es incapaz de detectar un bucle, mientras que el método propuesto sí lo hace.

4.2.2. Eficiencia

Para la medición de los tiempos de ejecución del método se ha empleado un ordenador portátil con una configuración modesta: CPU Intel(R) Core(TM) i3-2328M a 2,20GHz, y una memoria RAM compartida de 4GB SO-DIMM DDR3 a 1.333MHz. Como se puede ver en la Tab. 2, se han obtenido mediciones de tiempo de ejecución medio y desviación estándar de cada uno de los procesos del método al trabajar con una observación (menos para el de pre-procesamiento, cuyo tiempo de ejecución es despreciable). De esta manera, el tiempo medio de ejecución del método completo es de 51,2ms, lo que permitiría ejecutar lo con una frecuencia de 19,53Hz.

Los tiempos de ejecución y frecuencia obtenidos son prometedores ya que, si bien no alcanzan la frecuencia de funcionamiento del sensor ($\sim 30Hz$), los sistemas de detección de cierre de bucle suelen lanzarse en un hilo de ejecución a parte del resto de procesos dentro de un sistema de SLAM, y se activan cuando este sistema detecta un *keyframe*, lo cual suele ocurrir con una frecuencia inferior a la conseguida por el método propuesto.

El único proceso cuyo tiempo de ejecución se incrementa conforme se van procesando nuevas observaciones es el de la detección de bucle. Para estudiar su escalabilidad, se ha analizado cómo varía su tiempo de ejecución medio, obteniéndose los resultados de la Fig. 4. Aquí vemos como el índice inverso consigue que el método escale bien con respecto al número de representaciones almacenadas, estando el incremento de tiempo medio por debajo de un aumento lineal.

En cuanto a los tiempos de ejecución de las dos variantes, *Intensidad* y *Profundidad*, y como era de esperar, son ligeramente inferiores a los del método *Combinado*. Esto se debe a que el número de características con el que trabajan es inferior, aproximadamente la mitad, por lo que los procesos de extracción de las mismas, descripción, y generación de la representación BoW son más livianos.

5. CONCLUSIONES

En este trabajo se han descrito los primeros pasos hacia un método de reconocimiento de lugares que permita detectar de forma robusta y eficiente cuándo se ha cerrado un bucle en SLAM. Para conseguir robustez, el método combina información de distinta naturaleza, i.e. apariencia y geometría, proveniente de cámaras RGB-D. Es-

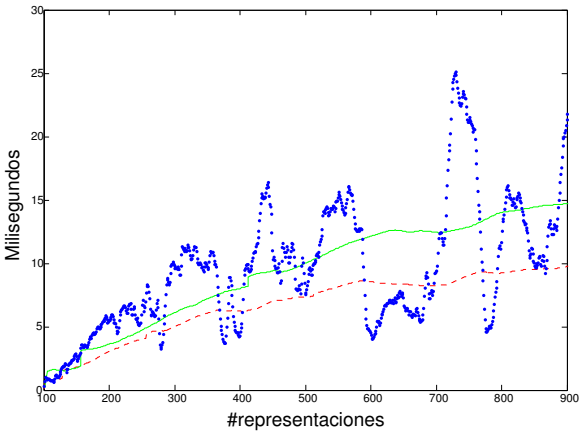

Figura 4: Evolución del tiempo empleado en detectar si se ha cerrado el bucle con respecto al número de representaciones almacenadas en la base de datos. Puntos en azul: tiempo empleado en la detección para una cierta representación, raya punteada en rojo: tiempo medio, raya verde: tiempo esperado si el incremento fuera lineal.

to permite detectar bucles en lugares con condiciones de iluminación pobres, falta de texturas, etc., donde un método basado en intensidad tendría dificultades, así como en lugares con insuficiente información geométrica, donde aquellos empleando información de profundidad no operarían correctamente. Por otro lado, para que la detección sea eficiente, las características de las observaciones son extraídas empleando FAST y descritas mediante BRIEF. Finalmente, la detección de cierre de bucle se realiza empleando una representación mediante Bolsas de Palabras binarias. Estos procesos son eficientes tanto en lo referente a tiempo de ejecución como a espacio de almacenamiento.

Para evaluar el rendimiento del método propuesto se han empleado secuencias de observaciones RGB-D públicamente disponibles: el vocabulario binario se ha generado utilizando secuencias del conjunto de datos *RGB-D SLAM Dataset* de la Universidad Técnica de Múnich; el ajuste de parámetros se ha realizado con secuencias grabadas en interiores; el método se evaluó con otra secuencia independiente grabada en un entorno doméstico, conteniendo múltiples observaciones en condiciones adversas que dificultan la detección de lugares. Los resultados de eficacia son muy satisfactorios, superándose en $\sim 10\,\%$ al rendimiento alcanzado por un método que emplea solo información de intensidad. En cuanto a la eficiencia, se ha conseguido combinar ambas fuentes de información y detectar bucles a una frecuencia de $\sim 20Hz$, suficiente para ser empleado en un sistema de SLAM.

En un futuro, se pretende evaluar el rendimiento del método al considerar dos representaciones mediante Bolsas de Palabras distintas: una para la información de intensidad y otra para la de profundidad. Esto permitiría, por ejemplo, decidir con mayor criterio cuándo aceptar una detección de bucle. Además se podrían aplicar técnicas de paralelización para procesar ambos tipos de información en dos hilos distintos, empleando por ejemplo OpenMP [14], disminuyéndose así el tiempo de ejecución del método. También se evaluará la utilización de descriptores como BRAND, que permiten una descripción conjunta de estas dos fuentes de información. Por último, se planea que el método desarrollado forme parte de un sistema de SLAM completo.

Agradecimientos

Este trabajo se ha desarrollado en el marco de los proyectos TEP2012-530 y DPI2014-55826-R, financiados por la Junta de Andalucía y el Ministerio de Ciencia e Innovación respectivamente, ambos contando con fondos del Fondo Europeo de Desarrollo Regional (FEDER).

Referencias

[1] D. Arthur and S. Vassilvitskii. `k-means++`: The advantages of careful seeding. In *Proceedings of the 18th Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '07, pages 1027–1035, Philadelphia, USA, 2007. Society for Industrial and Applied Mathematics.

[2] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool. Speeded-up robust features (surf). *Comput. Vis. Image Underst.*, 110(3):346–359, June 2008.

[3] J.-L. Blanco, J.-A. Fernández-Madrigal, and J. González-Jiménez. Towards a unified bayesian approach to hybrid metric-topological slam. *IEEE Transactions on Robotics*, 24(2):259–270, 2008.

[4] M. Calonder, V. Lepetit, C. Strecha, and P. Fua. Brief: Binary robust independent elementary features. In *Proceedings of the 11th European Conference on Computer Vision: Part IV*, ECCV'10, pages 778–792, Berlin, Heidelberg, 2010. Springer-Verlag.

[5] M. Cummins and P. Newman. Appearance-only slam at large scale with fab-map 2.0. *Int. J. Rob. Res.*, 30(9):1100–1123, Aug. 2011.

[6] D. Galvez-Lopez and J. D. Tardos. Real-time loop detection with bags of binary words. In *Intelligent Robots and Systems*

(IROS), 2011 IEEE/RSJ International Conference on, pages 51 –58, sept. 2011.

[7] D. Galvez-López and J. D. Tardos. Bags of binary words for fast place recognition in image sequences. *IEEE Transactions on Robotics*, 28(5):1188–1197, Oct 2012.

[8] J. González-Jiménez, C. Galindo, F. Melendez-Fernandez, and J. R. Ruiz-Sarmiento. Building and exploiting maps in a telepresence robotic application. In *10th International Conference on Informatics in Control, Automation and Robotics (ICINCO)*, 2013.

[9] J. González-Jiménez, A. Muñoz, C. Galindo, J.-A. Fernández-Madrigal, and J.-L. Blanco. A description of the sena robotic wheelchair. In *13th IEEE Mediterranean Electrotechnical Conference (MELECON)*, May 2006.

[10] International Telecomunication Union. Studio encoding parameters of digital television for standard 4:3 and wide screen 16:9 aspect ratios. `https://www.itu.int/rec/R-REC-BT.601/`. [Online; accessed 07-July-2016].

[11] M. Jaimez and J. González-Jiménez. Fast visual odometry for 3-d range sensors. *IEEE Transactions on Robotics*, 31(4):809–822, 2015.

[12] E. R. Nascimento, G. L. Oliveira, M. F. M. Campos, A. W. Vieira, and W. R. Schwartz. Brand: A robust appearance and depth descriptor for rgb-d images. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1720–1726, Oct 2012.

[13] D. Nister and H. Stewenius. Scalable recognition with a vocabulary tree. In *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2*, CVPR '06, pages 2161–2168, Washington, DC, USA, 2006. IEEE Computer Society.

[14] OpenMP Architecture Review Board. OpenMP API Specification for Parallel Programming. `http://openmp.org/wp/`. [Online; accessed 07-July-2016].

[15] E. Rosten and T. Drummond. Machine learning for high-speed corner detection. In *Proceedings of the 9th European Conference on Computer Vision - Volume Part I*, ECCV'06, pages 430–443, Berlin, Heidelberg, 2006. Springer-Verlag.

[16] J. R. Ruiz-Sarmiento, C. Galindo, and J. González-Jiménez. Experimental study of the performance of the kinect range camera for mobile robotics. Technical report, University of Malaga, 2013.

[17] J. R. Ruiz-Sarmiento, C. Galindo, and J. González-Jiménez. OLT: A Toolkit for Object Labeling Applied to Robotic RGB-D Datasets. In *European Conference on Mobile Robots*, 2015.

[18] S. A. Scherer, A. Kloss, and A. Zell. Loop closure detection using depth images. In *Mobile Robots (ECMR), 2013 European Conference on*, pages 100–106, 2013.

[19] H. Shahbazi and H. Zhang. Application of locality sensitive hashing to realtime loop closure detection. In *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1228–1233, Sept 2011.

[20] J. Sivic and A. Zisserman. Video google: A text retrieval approach to object matching in videos. In *Proceedings of the Ninth IEEE International Conference on Computer Vision - Volume 2*, ICCV '03, pages 1470–1477, Washington, DC, USA, 2003. IEEE Computer Society.

[21] B. Steder, G. Grisetti, and W. Burgard. Robust place recognition for 3d range data based on point features. In *Robotics and Automation (ICRA), 2010 IEEE International Conference on*, pages 1400–1405, May 2010.

[22] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers. A benchmark for the evaluation of rgb-d slam systems. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 573–580, Oct 2012.

[23] S. Thrun, W. Burgard, and D. Fox. *Probabilistic Robotics (Intelligent Robotics and Autonomous Agents)*. The MIT Press, 2005.

[24] H. Zhang, Y. Liu, and J. Tan. Loop closing detection in rgb-d slam combining appearance and geometric constraints. *Sensors*, 15(6):14639–14660, 2015.