Artif Intell Rev (2015) 43: 55–81 DOI
10.1007 / s10462-012-9365-8

Localización y mapeo visual simultáneo: una encuesta

Jorge Fuentes-Pacheco · José Ruiz-Ascencio · Juan Manuel Rendón-Mancha

Publicado en línea: 13 de noviembre de 2012

Resumen Visual SLAM (localización y mapeo simultáneos) se refiere al problema de utilizar imágenes, como única fuente de información externa, para establecer la posición de un robot, un vehículo o una cámara en movimiento en un entorno, y al mismo tiempo , construir una representación de la zona explorada. SLAM es una tarea fundamental para la autonomía de un robot. Hoy en día, el problema de SLAM se considera resuelto cuando se utilizan sensores de rango como láseres o sonar para construir mapas 2D de pequeños entornos estáticos. Sin embargo, SLAM para entornos dinámicos, complejos y de gran escala, utilizando la visión como único sensor externo, es un área activa de investigación. Las técnicas de visión por computadora empleadas en SLAM visual, como la detección, descripción y emparejamiento de características destacadas, reconocimiento y recuperación de imágenes, entre otras, aún son susceptibles de mejora.

Palabras clave SLAM visual · Selección de características destacadas · Coincidencia de imágenes · Asociación de datos · Mapas topológicos y métricos

1. Introducción

El problema de la navegación autónoma de los robots móviles se divide en tres áreas principales: localización, mapeo y planificación de rutas (Cyrill 2009). La localización consiste en determinar de forma exacta la pose actual del robot en un entorno. El mapeo integra el

J. Fuentes-Pacheco · J. Ruiz-Ascencio (B)
Centro Nacional de Investigación y Desarrollo Tecnológico, Cuernavaca, Morelos, México
e-mail: josera@cenidet.edu.mx

J. Fuentes-Pacheco
correo electrónico: jorge_fuentes@cenidet.edu.mx

JM Rendón-Mancha
Universidad Autónoma del Estado de Morelos, Cuernavaca, Morelos, México
e-mail: rendon@uaem.mx

observaciones parciales del entorno en un único modelo coherente y la planificación de la ruta determina la mejor ruta en el mapa para navegar por el entorno.

Inicialmente, el mapeo y la localización se estudiaron de forma independiente, luego se reconoció que son dependientes. Esto significa que, para estar localizado con precisión en un entorno, es necesario un mapa correcto, pero para construir un buen mapa es necesario estar correctamente localizado cuando se agregan elementos al mapa. Actualmente, este problema se conoce como *Localización y mapeo simultáneos (* GOLPE). Cuando las cámaras se emplean como único sensor exteroceptivo, se denomina *SLAM visual.* Los términos *SLAM basado en la visión (* Se et al. 2005 ; Lemaire y col. 2007) o *vSLAM (* Solà 2007) también se utilizan. En este artículo se utiliza el término SLAM visual porque es el más conocido. Los sistemas Visual SLAM se pueden complementar con información de sensores propioceptivos, con el objetivo de incrementar la precisión y robustez. Este enfoque se conoce como *SLAM visual-inercial (* Jones y Soatto 2011). Sin embargo, cuando se utiliza la visión como único sistema de percepción (sin hacer uso de la información extraída de la odometría del robot o de los sensores inerciales) se puede llamar *SLAM solo para visión (* Paz y col. 2008 ; Davison y col. 2007) o *SLAM solo con cámara (* Milford y Wyeth 2008).

Muchos sistemas SLAM visuales fallan cuando funcionan en las siguientes condiciones: en entornos externos, en entornos dinámicos, en entornos con demasiadas o muy pocas características destacadas, en entornos a gran escala, durante movimientos erráticos de la cámara y cuando hay oclusiones parciales o totales de la sensor ocurra. Una clave para un sistema SLAM visual exitoso es la capacidad de operar correctamente a pesar de estas dificultades.

Las aplicaciones importantes de SLAM están orientadas al pilotaje automático de automóviles en terrenos todoterreno sin ensayar (Thrun y col. 2005a); tareas de rescate para entornos de alto riesgo o de navegación difícil (Thrun 2003 ; Piniés et al. 2006); exploración planetaria, aérea, terrestre y oceánica (Olson y col. 2007 ; Artieda y col. 2009 ; Steder y col. 2008 ; Johnson y col. 2010); Aplicaciones de realidad aumentada donde se incluyen objetos virtuales en escenas del mundo real (Chekhlov y col. 2007 ; Klein y Murray 2007); sistemas de vigilancia visual Mei y col. 2011); medicamento (Auat y col. 2010 ; Grasa y col. 2011), Etcétera.

En este artículo se presenta un estudio detallado del SLAM visual, así como las aportaciones más recientes y diversos problemas actuales. Previamente, Durrant y Bailey presentaron un tutorial dividido en dos partes que resume el problema SLAM (Durrant y Bailey 2006 ; Bailey y Durrant 2006). El último tutorial describe trabajos que se centran en el uso de sensores láser telémetro, construyendo mapas 2D bajo un enfoque probabilístico. Similar, Thrun y Leonard (2008) presentó una introducción al problema SLAM, analizó tres paradigmas de solución (el primero se basa en el Filtro Kalman Extendido y los otros dos utilizan técnicas de optimización basadas en gráficos y filtros de partículas) y propusieron una taxonomía del problema. Sin embargo, los artículos mencionados anteriormente no se centran en métodos que utilizan la visión como único sensor externo. Por otro lado, Kragic y Vincze (2009) presentan una revisión de la visión por computadora para robótica en un contexto general, considerando el problema visual SLAM pero no en detalle como se pretende en este artículo.

Este artículo está estructurado de la siguiente manera: Art. 2 describe el problema SLAM en general. En la secc. 3 , se discute el uso de cámaras como único sensor externo y se mencionan los puntos débiles de dichos sistemas. Sección 4 describe el tipo de características sobresalientes que se pueden extraer y los descriptores utilizados para lograr la invariancia a las diversas transformaciones que pueden sufrir las imágenes. Sección 5 se ocupa de la coincidencia de imágenes y el problema de asociación de datos. Sección 6 da una revisión detallada de los diferentes métodos para resolver el problema visual SLAM y se discuten las debilidades y fortalezas de cada uno. Las diferentes formas de representar el mundo observado se describen en la Secta. 7 . Sección 8 proporciona conclusiones y problemas potenciales para futuras investigaciones. La sección final presenta referencias bibliográficas.

2 localización y mapeo simultáneos

Durante el período 1985-1990, Chatila y Laumond (1985) y Smith y col. (1990) propuso realizar el mapeo y la localización de forma concurrente. Algún tiempo después, este problema recibió el nombre de SLAM (localización y mapeo simultáneos). El lector puede consultar el tutorial de Durrant y Bailey (2006), Bailey y Durrant (2006) para obtener una descripción detallada de la historia del problema SLAM. En algunas publicaciones de Newman y col. (2002) y Andrade y Sanfeliu (2002) también se conoce como CML (mapeo y localización concurrentes). SLAM o CML es el proceso mediante el cual una entidad (robot, vehículo o incluso una unidad central de procesamiento con dispositivos sensores llevados por una persona) tiene la capacidad de construir un mapa global del entorno visitado y, al mismo tiempo, utilizar este mapa para deducir su propia ubicación en cualquier momento.

Para construir un mapa a partir del entorno, la entidad debe poseer sensores que le permitan percibir y obtener medidas de los elementos del mundo circundante. Estos sensores se clasi fi can en *exteroceptivo* y *propioceptivo.* Entre los sensores exteroceptivos es posible encontrar: sonar (Tardós et al. 2002 ; Ribas y col. 2008), láseres de alcance (Nüchter y col. 2007 ; Thrun y col. 2006), cámaras (Se et al. 2005 ; Lemaire y col. 2007 ; Davison 2003 ; Bogdan y col. 2009) y sistemas de posicionamiento global (GPS) (Thrun y col. 2005a). Todos estos sensores son ruidosos y tienen capacidades de alcance limitado. Además, solo se pueden obtener vistas locales del entorno utilizando los tres primeros sensores mencionados anteriormente. Los sensores láser y la sonda permiten obtener información precisa y muy densa de la estructura del entorno. Sin embargo, tienen los siguientes problemas: no son útiles en entornos muy abarrotados o para reconocer objetos; ambos son caros, pesados y constan de grandes piezas de equipo, lo que dificulta su uso para robots aéreos o humanoides. Por otro lado, un sensor GPS no funciona bien en calles estrechas (cañones urbanos), bajo el agua, en otros planetas y, en ocasiones, no está disponible en interiores.

Los sensores propioceptivos permiten a la entidad obtener medidas como velocidad, cambio de posición y aceleración. Algunos ejemplos son: codificadores, acelerómetros y giroscopios. Permiten obtener una estimación incremental de los movimientos de la entidad mediante una *Deadreckoning* método de navegación (también conocido como *cálculo deducido),* pero debido a su ruido inherente, no son suficientes para tener una estimación precisa de la posición de la entidad en todo momento, ya que los errores son acumulativos.

Como se ha demostrado en algunas investigaciones (Castellanos y col. 2001 ; Majumder y col. 2005 ; Nützi y col. 2010), para mantener una estimación precisa y sólida de la posición del robot, es necesario utilizar el *fusión de información* de múltiples sensores de percepción. Sin embargo, la adición de sensores aumenta el costo, el peso y los requisitos de potencia de un sistema; por lo tanto, es importante investigar cómo una entidad puede ubicarse a sí misma y crear un mapa con solo cámaras.

3 cámaras como únicos sensores exteroceptivos

En los últimos 10 años, los artículos publicados reflejan una clara tendencia a utilizar la visión como único sistema externo de percepción sensorial para resolver el problema de SLAM (Paz y col. 2008 ; Davison y col. 2007 ; Klein y Murray 2007 ; Sáez y Escolano 2006 ; Piniés y Tardós 2008). El motivo principal de esta tendencia se atribuye a la capacidad de un sistema basado en cámaras para obtener información de rango, y también recuperar la apariencia, el color y la textura del entorno, dando al robot la posibilidad de integrar otras tareas de alto nivel como la detección y el reconocimiento de personas y lugares. Además, las cámaras son menos costosas, más livianas y tienen

menor consumo de energía. Desafortunadamente, puede haber errores en los datos por las siguientes razones: resolución de cámara insuficiente, cambios de iluminación, superficies con falta de textura, imágenes borrosas por movimientos rápidos, entre otros factores.

Los primeros trabajos sobre navegación visual se basaron en una configuración estéreo binocular (Se et al. 2002 ; Olson y col. 2003). Sin embargo, en muchos casos es difícil tener un dispositivo con cámaras estéreo binoculares o trinoculares debido a sus altos costos. Una alternativa es utilizar un par de cámaras monoculares (por ejemplo webcams), lo que lleva a considerar diferentes aspectos como: (a) la sincronización de la cámara mediante el uso de hardware o software, (b) las diferentes respuestas de cada sensor CCD a color y luminancia, y (c) la alineación mecánica según el esquema de geometría elegido (ejes paralelos o convergentes).

También existen obras que hacen uso de equipos multicámara con o sin superposición entre las vistas (Kaess y Dellaert 2010 ; Carrera y col. 2011) y cámaras con lentes especiales como gran angular (Davison y col. 2004) u omnidireccional (Scaramuzza y Siegwart 2008) con el objetivo de aumentar el alcance visual y así disminuir, hasta cierto punto, el error acumulativo de estimación de pose. Recientemente, se han utilizado sensores RGB-D (imágenes en color y mapas de profundidad) para mapear entornos interiores (Huang y col. 2011), demostrando ser una alternativa prometedora para las aplicaciones SLAM.

Independientemente de la configuración utilizada, las cámaras deben calibrarse (manualmente fuera de línea o automáticamente en línea). *Calibración* estimados *intrínseco* y *parámetros extrínsecos,* el primero depende de la geometría de la cámara (distancia focal y punto principal), mientras que el segundo depende de la posición de la cámara en el espacio (rotación y traslación con respecto a algún sistema de coordenadas). Los parámetros necesarios generalmente se estiman a partir de un conjunto de imágenes que contienen múltiples vistas de un patrón de calibración de tablero de ajedrez, para relacionar las coordenadas de la imagen con las coordenadas del mundo real (Hartley y Zisserman 2003). Existen muchas herramientas para ejecutar el proceso de calibración, algunas de ellas son: las funciones de calibración de OpenCV (2009) (basado en el algoritmo de Zhang (Zhang 2000)), Caja de herramientas de calibración de cámara para Matlab (Bouguet 2010), El software de calibración de la cámara Tsai (Willson 1995), OCamCalib Toolbox para cámaras omnidireccionales (Scaramuzza 2011), y Autocalibración de múltiples cámaras para calibrar varias cámaras (al menos 3) (Svoboda 2011).

Si la calibración de la cámara se realiza fuera de línea, se supone que las propiedades intrínsecas de la cámara no cambiarán durante todo el período de funcionamiento del sistema SLAM. Esta es la opción más popular, ya que reduce la cantidad de parámetros calculados en línea. No obstante, la información intrínseca de la cámara puede cambiar debido a algunos factores ambientales del entorno, como la humedad o la temperatura. Además, un robot que trabaja en condiciones del mundo real puede ser golpeado o dañado, lo que podría invalidar la calibración previamente adquirida (Koch y col. 2010).

Las configuraciones estéreo (binoculares, trinoculares o múltiples cámaras con sus campos de visión parcialmente superpuestos) ofrecen la ventaja de poder calcular de manera fácil y precisa las posiciones reales en 3D de los puntos de referencia contenidos en la escena, por medio de *triangulación* Hartley y Sturm 1997), que es información de gran utilidad en el problema visual SLAM. Las obras de Konolige y Agrawal (2008), Konolige y col. (2009), Mei y col. (2009) representan los sistemas SLAM estéreo binoculares más actuales y eficaces. Cuando la localización y el mapeo se realizan con una sola cámara, el mapa sufrirá un problema de ambigüedad de escala (Nister 2004 ; Strasdat y col. 2010a). Para obtener información 3D de una sola cámara, existen dos casos en función del conocimiento a priori de la cámara. Estos son: (a) con el conocimiento de los parámetros intrínsecos únicamente; con esta alternativa la estructura ambiental y los parámetros extrínsecos se recuperan con un factor de escala indeterminado. La escala se determina si se conoce la distancia real entre dos puntos en el espacio; y (b) donde solo se conocen correspondencias; en este caso, la reconstrucción se compone de una transformación proyectiva.

La idea de utilizar una cámara se ha vuelto popular desde la aparición de *SLAM de una sola cámara* o *MonoSLAM (* Davison 2003). Probablemente esto también se deba a que ahora es más fácil acceder a una sola cámara que a un par estéreo, a través de teléfonos celulares, asistentes digitales personales o computadoras personales. Este enfoque monocular ofrece una solución muy simple, flexible y económica en términos de hardware y tiempos de procesamiento.

Monocular SLAM es un caso particular de *SLAM solo de rodamientos.* Este último es un problema parcialmente observable, donde los sensores no proporcionan suficiente información a partir de una simple observación para determinar la profundidad de un punto de referencia. Esto provoca un problema de inicialización histórico, donde las soluciones se pueden dividir en dos categorías: *demorado* y *no retrasado* Lemaire y col. 2007 ; Vidal y col. 2007). Se debe realizar un seguimiento de características destacadas a través de múltiples observaciones para obtener información tridimensional de una sola cámara.

Aunque se han hecho muchas contribuciones al SLAM visual, todavía existen muchos problemas. Las soluciones propuestas para el problema visual SLAM se revisan en la Sección. 6 Muchos sistemas SLAM visuales sufren grandes errores acumulados mientras se explora el entorno (o fallan por completo en entornos visualmente complejos), lo que conduce a estimaciones inconsistentes de la posición del robot y mapas totalmente incongruentes. Existen tres razones principales:

(1) Primero, generalmente se asume que el movimiento de la cámara es suave y que habrá consistencia en la apariencia de las características sobresalientes (Davison 2003 ; Nistér y col. 2004), pero en general esto no es cierto. Los supuestos anteriores están muy relacionados con la selección del detector de características destacadas y de la técnica de coincidencia utilizada. Esto origina una inexactitud en la posición de la cámara al capturar imágenes con poca textura o que se ven borrosas debido a movimientos rápidos del sensor (por ejemplo, debido a vibraciones o cambios rápidos de dirección) (Pupilli y Calway 2006). Estos fenómenos son típicos cuando la cámara la lleva una persona, robots humanoides y helicópteros de cuatro rotores, entre otros. Una forma de aliviar este problema hasta cierto punto es mediante el uso de *fotogramas clave* consulte el "Apéndice I") (Mouragnon y col. 2006 ; Klein y Murray 2008). Alternativamente, Pretto y col. (2007) y Mei y Reid (2008) analizan el problema del seguimiento visual en tiempo real sobre secuencias de imágenes borrosas debido a una cámara desenfocada.

(2) En segundo lugar, la mayoría de los investigadores asume que los entornos a explorar son estáticos y que solo contienen elementos estacionarios y rígidos; la mayoría de los entornos contienen personas y objetos en movimiento. Si esto no se tiene en cuenta, los elementos móviles originarán coincidencias falsas y, en consecuencia, generarán errores impredecibles en todo el sistema. Los primeros enfoques a este problema son propuestos por Wang y col. (2007); Wangsiripitak y Murray (2009); Migliore y col. (2009), así como Lin y Wang (2010).

(3) En tercer lugar, el mundo es visualmente repetitivo. Hay muchas texturas similares, como los elementos arquitectónicos repetidos, el follaje y las paredes de ladrillo o piedra. Además, algunos objetos, como las señales de tráfico, aparecen repetidamente dentro de un entorno urbano al aire libre. Esto hace que sea difícil reconocer un área previamente explorada y también hacer SLAM en grandes extensiones de tierra.

4 Selección de características destacadas

Marcaremos la diferencia entre *características sobresalientes* y *puntos de referencia,* ya que en algunos artículos se tratan indistintamente. De acuerdo a Frintrop y Jensfelt (2008), un hito es una región del mundo real descrita por información de posición y apariencia en 3D. Por otro lado, una característica sobresaliente es una región de la imagen descrita por su posición 2D (en la imagen) y una

apariencia. En esta encuesta, el término característica sobresaliente se utiliza como una generalización que puede incluir puntos, regiones o incluso segmentos de borde que se extraen de las imágenes.

Las características más destacadas que son más fáciles de localizar son las producidas por hitos artificiales (Frintrop y Jensfelt 2008). Estos puntos de referencia se agregan intencionalmente al entorno con el propósito de servir como una ayuda para la navegación, por ejemplo, cuadrados o círculos situados en el piso o las paredes. Estos puntos de referencia tienen la ventaja de que su apariencia se conoce de antemano, lo que los hace fáciles de detectar en cualquier momento. Sin embargo, el entorno debe ser preparado por una persona antes de que se inicialice el sistema. Los hitos naturales son aquellos que existen habitualmente en el medio ambiente (Se et al. 2002). Para ambientes interiores es común utilizar como hitos naturales las esquinas de puertas o ventanas. En ambientes al aire libre, los troncos de los árboles (Asmar 2006), regiones (Matas y col. 2002), o puntos de interés (Lowe 2004) son usados. Un *punto de interés*
es un píxel de imagen con una vecindad tal que es fácil de distinguir de otros puntos usando un detector dado.

Una característica de buena calidad tiene las siguientes propiedades: debe ser notable (fácil de extraer), precisa (puede medirse con precisión) e invariable a los cambios de rotación, traslación, escala e iluminación (Lemaire y col. 2007). Por lo tanto, un hito de buena calidad tiene una apariencia similar desde diferentes puntos de vista en el espacio 3D. El proceso de extracción de características destacadas se compone de dos fases: *detección* y *descripción.* La detección consiste en procesar la imagen para obtener una serie de elementos destacados. La descripción consiste en construir un vector de características basado en la apariencia visual de la imagen. La invariancia del descriptor a los cambios de posición y orientación permitirá mejorar los procesos de correspondencia de imágenes y asociación de datos (descritos en la Sec. 5).

4.1 Detectores

En la mayoría de los sistemas SLAM basados en la visión, se han utilizado características naturales presentes en todas partes, como esquinas, puntos de interés, segmentos de borde y regiones. La selección del tipo de funciones a utilizar dependerá en gran medida del entorno en el que vaya a trabajar el robot.

Existe una gran cantidad de detectores de características destacadas. Algunos ejemplos son: detector de esquinas Harris (Harris y Stephens 1988), Detectores de puntos Harris-Laplace y Hessian-Laplace, así como sus respectivas versiones invariantes afines Harris-Af fi ne y HessianAf fi ne (Mikolajcczyk y Schmid 2002); Diferencia de gaussianos (DoG) utilizada en SIFT (Transformación de características invariantes de escala) (Lowe 2004); Regiones extremas máximamente estables (MSER) (Matas 2002), FAST (características de la prueba de segmento acelerado) (Rosten y Drummond 2006) y el Fast-Hessian utilizado en SURF (características robustas aceleradas) (Bay y col. 2006). Mikolajczyk y col. (2005) realizó una evaluación del rendimiento de estos algoritmos con respecto al punto de vista, zoom, rotación, desenfoque, compresión JPEG y cambios de iluminación. Los detectores Hessian-Affin y MSER tuvieron el mejor rendimiento, MSER fue el más robusto con respecto al punto de vista y los cambios de iluminación, y el Hessian-Affin fue el mejor en presencia de características desenfocadas y compresión JPEG. En (Tuytelaars y Mikolajczyk 2008) estos detectores y algunos otros se clasifican teniendo en cuenta su repetibilidad, precisión, robustez, e fi ciencia y características invariantes.

La mayoría de los sistemas SLAM visuales utilizan las esquinas como puntos de referencia debido a sus características invariables y su amplio estudio en el contexto de la visión por computadora. Sin emabargo, Eade y Drummond (2006a) proponen utilizar segmentos de borde llamados edgelets en un sistema MonoSLAM en tiempo real, lo que permite la construcción de mapas con altos niveles de información geométrica. Los autores demostraron que los bordes son buenas características para el seguimiento y SLAM, debido a su invariancia a

cambios de iluminación, orientación y escala. El uso de bordes como características parece prometedor, ya que los bordes se ven poco afectados por el desenfoque causado por los movimientos repentinos de la cámara (Klein y Murray 2008). No obstante, los bordes tienen la limitación de no ser fáciles de extraer y emparejar. Por otro lado, Gee y col. (2008) y Martinez y Calway (2010) investigan la fusión de características (es decir, puntos, líneas y estructuras planas) en un solo mapa, con el propósito de aumentar la precisión de los sistemas SLAM y crear una mejor representación del entorno.

4.2 Descriptores

Uno de los descriptores más utilizados para el reconocimiento de objetos es el descriptor SIFT de tipo histograma, propuesto por Lowe (2004), que se basa en la distribución espacial de características locales en la vecindad del punto saliente, obteniendo un vector de 128 componentes. Ke y Sukthankar (2004) proponen una modi fi cación de SIFT denominada PCA-SIFT, cuya idea principal es obtener un descriptor tan distintivo y robusto como SIFT pero con un vector de menos componentes. La reducción se logra mediante la técnica de Análisis de Componentes Principales. Los descriptores de tipo histograma tienen la propiedad de ser invariantes a la traslación, rotación y escala, y parcialmente invariantes a los cambios de iluminación y puntos de vista. Una evaluación exhaustiva de varios algoritmos de descripción y una propuesta para una extensión del descriptor SIFT (Gradient Location-Orientation HistogramGLOH) se puede encontrar en (Moreels y Perona 2005) y (Mikolajcczyk y Schmid 2005), respectivamente.

En (Gil y col. 2009) aparece un estudio comparativo de diferentes algoritmos de descripción local centrado en el problema visual de SLAM. La evaluación se basa en la cantidad de coincidencias correctas e incorrectas encontradas a través de secuencias de video con cambios significativos en la escala, el punto de vista y la iluminación. En este trabajo se demuestra que el descriptor SURF es superior al descriptor SIFT en términos de robustez y tiempo de cálculo. Más adelante, los autores manifiestan que SIFT no demuestra una gran estabilidad, lo que significa que muchos de los puntos de referencia detectados desde una determinada posición de la cámara desaparecen al moverla ligeramente. Actualmente existen muchas variantes que mejoran el rendimiento del algoritmo SIFT, por ejemplo: ASIFT, que incorpora invariancia a transformaciones afines (Morel y Yu 2009), BREVE (Funciones elementales independientes robustas binarias) (Calonder y col. 2010); ORB, un descriptor binario rápido basado en BRIEF pero invariante en rotación y resistente al ruido (Rublee y col. 2011); PIRF (característica robusta de posición invariable) (Kawewong y col. 2010) y GPU-SIFT, una implementación de SIFT en una GPU (Graphics Processing Unit) para realizar procesamientos en paralelo y en tiempo real (Sinha y col. 2006).

5 Los problemas de coincidencia de imágenes y asociación de datos

En correspondencia estéreo, el *coincidencia de imágenes* consiste en buscar cada elemento en una imagen, su correspondiente en la otra imagen. Las técnicas de emparejamiento se pueden dividir en dos categorías: *línea de base corta* y *línea de base larga.* Estas técnicas son necesarias durante la etapa de seguimiento de características destacadas y detección de cierre de bucle en SLAM visual. En el área de la navegación del robot, el *asociación de datos* consiste en relacionar las medidas del sensor con los elementos que ya están dentro del mapa del robot (Neira y Tardós 2001). Este problema también implica determinar si las mediciones son falsas o pertenecen a elementos no contenidos en el mapa. Una coincidencia de imágenes eficiente y una asociación de datos correcta son esenciales para una navegación exitosa. Los errores conducirán rápidamente a mapas incorrectos.

5.1 Coincidencia de línea de base corta

Base es la línea que separa los centros ópticos de dos cámaras que se utilizan para capturar un par de imágenes. Cuando la diferencia entre las imágenes tomadas desde diferentes puntos de vista es pequeña, el punto correspondiente tendrá casi la misma posición y apariencia en ambas imágenes, reduciendo la complejidad del problema. En este caso, el punto se caracteriza simplemente por los valores de intensidad de un conjunto de píxeles muestreados de una ventana rectangular (también conocida como *parche)*
que se centra sobre la característica más destacada. Los valores de intensidad de los píxeles se comparan mediante medidas de correlación como correlación cruzada, suma de diferencias cuadradas y suma de diferencias absolutas, entre otras. En (Ciganek y Siebert 2009) hay una lista de fórmulas para determinar la similitud entre dos parches. Artículos (Konolige y Agrawal 2008 ; Nistér y col. 2004) manifiestan que la medida de correlación cruzada normalizada (NCC) es la que presenta mejores resultados. La normalización hace que este método sea invariante a los cambios uniformes de brillo. En (Davison 2003) y (Molton y col. 2004) se calcula una homografía para deformar el parche y hacer que las correspondencias con NCC sean invariables con respecto a los puntos de vista, lo que permite una mayor libertad de movimiento de la cámara. Desafortunadamente, la correspondencia con NCC es susceptible de falsos positivos y falsos negativos. En una región de imagen con textura repetida, dos o más puntos dentro de una región de búsqueda pueden obtener una fuerte respuesta a NCC.

Para correspondencias breves de la línea de base, es importante tener en cuenta las dimensiones del parche, así como las dimensiones de la región de búsqueda, de lo contrario aparecerán errores (Nistér y col. 2004). Por ejemplo, los parches que son demasiado pequeños son buenos para la velocidad, pero tienden a generar correspondencias falsas y los parches demasiado grandes requieren más tiempo de procesamiento. Se recomienda utilizar parches de aproximadamente 9 × 9 u 11 × 11 píxeles y coloque el parche sobre una esquina, ya que en dicha región el degradado de la imagen tiene dos o más direcciones dominantes y, en consecuencia, facilita el proceso de correspondencia. El uso de descriptores es innecesario para la coincidencia de línea de base corta de fotograma a fotograma, pero si el seguimiento falla y la cámara se pierde, entonces se convierte en una gran ayuda.

Una desventaja de la línea base corta es que el cálculo de la profundidad es muy sensible al ruido, por ejemplo, las mediciones de error de las coordenadas de la imagen, debido a la distancia reducida entre diferentes puntos de vista. Sin embargo, es posible realizar un seguimiento preciso de las características correspondientes a través de secuencias de video. Yilmaz y col. (2006), Cañones (2008) y Lepetit y Fua (2005) presentan un estudio del estado del arte de las técnicas para realizar el seguimiento basado en características, contornos o regiones.

5.2 Emparejamiento de línea de base larga

Al trabajar con líneas de base largas, las imágenes presentan grandes cambios de escala o perspectiva, lo que origina que un punto de una imagen se mueva a cualquier lugar de la otra imagen. Esto crea un problema de correspondencia difícil, véase la sección. 3 de Brown y col. (2003). Los datos de la imagen en la vecindad de un punto están distorsionados por cambios en el punto de vista y la iluminación, y las medidas de correlación no darán buenos resultados.

La forma más fácil de encontrar correspondencias es comparar todas las características de una imagen con todas las características de alguna otra imagen (enfoque conocido como "fuerza bruta"). Desafortunadamente, este proceso crece de forma cuadrática por la cantidad de características extraídas, lo cual no es práctico para muchas aplicaciones que deben funcionar en tiempo real.

En los últimos años, ha habido un progreso considerable en el desarrollo de algoritmos de coincidencia para líneas de base largas que son invariantes a varias transformaciones de la imagen. Muchos de estos algoritmos obtienen un descriptor para cada característica detectada, calculan diferencias

mide la laridad entre descriptores y utiliza estructuras de datos para realizar la búsqueda de pares de forma rápida y eficiente.

Existen varias medidas de disimilitud, como la distancia euclidiana, la distancia de Manhattan, la distancia de Chi-Cuadrado, entre otras. Las estructuras de datos pueden ser árboles binarios balanceados llamados árboles kd (Beis y Lowe 1997 ; Silpa y Hartley 2008) o tablas hash (Grauman y Darrell 2007). También existen criterios para decidir cuándo se deben asociar dos características (Mikolajcczyk y Schmid 2005). Algunos ejemplos son: (a) umbral de distancia: dos características son relevantes si la distancia entre sus descriptores está por debajo de un umbral; (b) vecino más cercano: A y B están relacionados si el descriptor B es el vecino más cercano del descriptor A y si la distancia entre ellos está por debajo de un umbral, y (c) relación de distancia del vecino más cercano: este enfoque es similar al vecino más cercano excepto que el umbral se aplica a la relación de distancias del píxel actual al primer y segundo vecino más cercano. Al utilizar el primer criterio descrito anteriormente, una característica de la primera imagen se puede emparejar con varias características de la segunda imagen y viceversa. Existen diferentes técnicas para desambiguar estas coincidencias candidatas, por ejemplo: mediante técnicas de relajación (Zhang y col. 1994) o considerando colecciones de puntos (Dufournaud y col. 2004). A *restricción geométrica* se puede utilizar para acelerar el proceso de emparejamiento. La restricción epipolar establece que: una condición necesaria para X y X' ser puntos correspondientes, es que el punto X' tiene que estar en la línea epipolar de X (Hartley y Zisserman 2003). De esta manera, la búsqueda de coincidencias se restringe a una sola línea en lugar de a la imagen completa. Algunos detalles se pueden encontrar en Tuytelaars y Van-Gool (2004); Zhang y Kosecka (2006) y Matas y col.

(2002).

Otras investigaciones como (Lepetit y Fua 2006 ; Grauman 2010 ; Kulis y col. 2009 ; Özuysal y col. 2010) utilizan estrategias de aprendizaje para determinar la similitud entre características. Esto reformula el problema de la correspondencia como un problema de clasificación, lo que parece muy prometedor. En el caso concreto de las aplicaciones SLAM en tiempo real, este podría no parecer del todo adecuado ya que es necesario realizar un entrenamiento constante en línea. Sin embargo, (Hinterstoisser y col. 2009 ; Taylor y Drummond 2009) han propuesto métodos más rápidos para lograr el aprendizaje en línea, que podrían utilizarse en el futuro para aplicaciones SLAM.

Aguilar y col. (2009), Li y col. (2010) y Gu y col. (2010) proponen una imagen diferente enfoque de espondence, donde las relaciones de vecindad de puntos se representan mediante un gráfico. Los gráficos correspondientes son aquellos que son iguales o similares en ambas imágenes. Del mismo modo, Sanromá et al. (2010) proponen un algoritmo de emparejamiento iterativo basado en gráficos, que se utiliza para recuperar la pose de un robot móvil. Desafortunadamente, estas investigaciones aún son limitadas porque no funcionan en tiempo real y no pueden manejar oclusiones temporales.

El uso de descriptores de alta calidad o incluso diferentes tipos de medidas de similitud no garantiza evitar correspondencias falsas. Si estas correspondencias se utilizan dentro de un sistema SLAM, se generarán errores importantes para la posición de la cámara y la estimación del mapa. Por tanto, es necesario utilizar estimadores robustos como RANSAC (Consenso de Muestra Aleatoria), PROSAC (Consenso de Muestra Progresiva), entre otros, pueden manejar automáticamente correspondencias falsas. Un análisis comparativo de estos estimadores se puede encontrar en (Raguram y col. 2008). La principal diferencia entre ellos es la forma en que evalúan la calidad de un modelo. Los estimadores robustos se utilizan comúnmente para estimar los parámetros del modelo a partir de datos que contienen valores atípicos. RANSAC estima una relación global adaptando los datos y, al mismo tiempo, clasifica los datos en valores inliers (datos que son consistentes con la relación) y outliers (no consistentes con la relación). Debido a la capacidad de tolerar una gran cantidad de valores atípicos, este algoritmo es una opción popular para resolver una gran variedad de problemas de estimación.

Una alternativa a RANSAC es presentada por Chli y Davison (2008 , 2009), que proplantean una técnica bayesiana para la correspondencia marco a marco llamada coincidencia activa. La coincidencia activa realiza una búsqueda solo en las partes de la imagen donde es más probable encontrar verdaderos positivos, lo que reduce el número de valores atípicos y el número de operaciones de procesamiento de imágenes para procesar imágenes. Este algoritmo utiliza una búsqueda guiada basada en el principio de la teoría de la información de Shannon. El emparejamiento activo presenta buenos resultados frente a movimientos rápidos de cámara. La limitación de esta técnica es su escasa escalabilidad cuando aumenta el número de funciones. Para resolver este problema, Handa y col. (2010) proponen una extensión que permite gestionar cientos de funciones en tiempo real, sin perder precisión en la correspondencia.

Una forma de medir el rendimiento de los algoritmos de emparejamiento es mediante el *Curva característica de funcionamiento del receptor,* o curva ROC (Fawcett 2006). Esta es una representación gráfica que involucra el cálculo de verdaderos positivos, falsos positivos, falsos negativos y verdaderos negativos, además del valor predictivo positivo y la precisión. Donde los verdaderos positivos son el número de coincidencias correctas, los falsos negativos son coincidencias que no se detectaron correctamente, los falsos positivos son coincidencias que son incorrectas y los verdaderos negativos son no coincidencias que se rechazaron correctamente. En algunos artículos de la literatura sobre recuperación de información (Majumder y col. 2005), se utilizan las dos métricas siguientes: *precisión* número de coincidencias correctas dividido por el número total de correspondencias encontradas) y *recordar (* número de coincidencias correctas dividido por el número total de correspondencias esperadas).

5.3 Asociación de datos en SLAM visual

El problema de asociación de datos en SLAM visual se apoya mediante técnicas de Reconocimiento Visual de Lugares. La asociación de datos tiene casos particulares, como: *detección de cierre de bucle, robot secuestrado (* o cámara), y *multisesión* y *mapeo cooperativo;* que se describen en las siguientes líneas:

5.3.1 Detección de cierre de bucle

La detección de cierre de bucle consiste en reconocer un lugar que ya ha sido visitado en una excursión cíclica de longitud arbitraria (Ho y Newman 2007 ; Clemente y col. 2007 ; Mei y col. 2010). Este problema ha sido uno de los mayores impedimentos para realizar SLAM a gran escala y recuperarse de errores críticos. De este problema surge otro llamado *alias perceptual* Angeli y col. 2008 ; Cummins y Newman 2008); donde dos lugares diferentes del entorno se reconocen como lo mismo. Esto representa un problema incluso cuando se utilizan cámaras como sensores debido a las características repetitivas del entorno, por ejemplo, pasillos, elementos arquitectónicos similares o zonas con una gran cantidad de arbustos. Un buen método de detección de cierre de bucle no debe devolver ningún falso positivo y debe obtener un mínimo de falsos negativos.

De acuerdo a Williams y col. (2009) métodos de detección de cierres de bucle en SLAM visual se puede dividir en tres categorías: (1) *mapa a mapa; (* 2) *imagen a imagen;* y (3) *imagen al mapa.* Las categorías difieren principalmente acerca de dónde se toman los datos de asociación (espacio de mapa métrico o espacio de imagen). Sin embargo, lo ideal sería construir un sistema que combine las ventajas de las tres categorías. La detección de cierre de bucle es un problema importante para cualquier sistema SLAM, y teniendo en cuenta que las cámaras se han convertido en un sensor muy común para aplicaciones robóticas, muchos investigadores se centran en métodos de visión para solucionarlo.

Ho y Newman (2007) proponen utilizar una matriz de similitud para codificar las relaciones de semejanza entre todos los pares posibles en las imágenes capturadas. Demuestran mediante una descomposición de valor único que es posible detectar cierres de bucle, a pesar de la presión.

presencia de imágenes repetitivas y visualmente ambiguas. Eade y Drummond (2008) presentan un método unificado para recuperarse de los fallos de seguimiento y detectar cierres de bucle en el problema del SLAM visual monocular en tiempo real. También proponen un sistema llamado GraphSLAM donde cada nodo almacena puntos de referencia y mantiene estimaciones de las transformaciones relacionadas con los nodos. Para detectar fallas o cierres de bucles, modelan la apariencia como un *Bolsa de palabras visuales* (BoVW) para encontrar los nodos que tienen una apariencia similar en la imagen de video actual (consulte el "Apéndice II"). Angeli y col. (2008) presentan un método para detectar cierres de bucles bajo un esquema de filtrado bayessiano y un método de BoVW incremental, donde se calcula la probabilidad de pertenecer a una escena visitada para cada imagen adquirida. Cummins y Newman (2008) proponen un marco probabilístico para reconocer lugares, que utiliza solo datos de apariencia de imagen. Mediante el aprendizaje de un modelo generativo de apariencia, demuestran que no solo es posible calcular la semejanza de dos observaciones, sino también la probabilidad de que pertenezcan al mismo lugar; y, por tanto, calculan una función de distribución de probabilidad (*pdf)* de la posición observada. Finalmente, Mei y col. (2010) proponen una nueva representación topométrica del mundo, basada en la co-visibilidad, que permite simplificar la asociación de datos y mejorar el rendimiento del reconocimiento basado en la apariencia.

Todos los trabajos de cierre de bucle descritos anteriormente, tienen como objetivo lograr una precisión del 100%. Esto se debe a que un solo falso positivo puede causar fallas irremediables durante la creación del mapa. En el contexto de SLAM, los falsos positivos son más graves que los falsos negativos (Magnusson y col. 2009). Los falsos negativos reducen el porcentaje de recuerdo pero no tienen ningún impacto en el porcentaje de precisión. Por lo tanto, para determinar la e fi ciencia de un detector de cierre de bucle, la tasa de recuperación debe ser lo más alta posible, con una precisión del 100%.

5.3.2 Robot secuestrado

En el problema del robot secuestrado, la pose del robot en el mapa se determina sin información previa de su paradero. Este caso puede ocurrir si el robot se vuelve a colocar en una zona ya mapeada, sin el conocimiento de su desplazamiento mientras se transporta a ese lugar, o cuando el robot realiza movimientos a ciegas debido a oclusiones, mal funcionamiento temporal del sensor o movimientos rápidos de la cámara (Eade y Drummond 2008 ; Chekhlov y col. 2008 ; Williams y col. 2007).

Chekhlov y col. (2008) proponen un sistema capaz de tolerar la incertidumbre sobre la cam-Era posar y recuperarse de fallas de seguimiento menores generadas por movimientos erráticos continuos o por oclusiones. El trabajo consiste en generar un descriptor (basado en SIFT) a múltiples resoluciones para brindar robustez en la tarea de asociación de datos. Además, utiliza un índice basado en coeficientes de orden bajo de la ondícula de Haar. Williams y col. (2007) presentan un módulo de re-localización que monitorea el sistema SLAM, detecta fallas de rastreo, determina la posición de la cámara en el marco de los puntos de referencia del mapa y reanuda el rastreo tan pronto como las condiciones mejoran. La re-localización se realiza mediante un algoritmo de reconocimiento de puntos de referencia utilizando la técnica de clasi fi cador de árboles aleatorizados propuesta por Lepetit y Fua (2006) y capacitado en línea a través de una técnica de recolección de características. De esta forma se obtienen una alta tasa de recuperación y un rápido tiempo de reconocimiento. Para encontrar la pose de la cámara, las poses candidatas se generan a partir de las correspondencias entre el cuadro actual y los puntos de referencia en el mapa. Hay una selección de conjuntos de tres posibles coincidencias, luego, todas las poses consistentes con estos conjuntos se calculan mediante un algoritmo de tres puntos. Estas poses se evalúan buscando el consenso entre las otras correspondencias en la imagen encontrada por RANSAC. Si se encuentra una pose con un gran consenso, se asume que esa pose es correcta.

5.3.3 Mapeo cooperativo y multisesión

El mapeo multisesión y cooperativo consiste en alinear dos o más mapas parciales del entorno recogidos por un robot en diferentes periodos de funcionamiento o por varios robots al mismo tiempo (*SLAM cooperativo visual) (* Ho y Newman 2007 ; Gil y col. 2010 ; Vidal y col. 2011).

En el pasado, el problema de asociar mediciones con puntos de referencia en el mapa se resolvía mediante algoritmos como Vecino más cercano, Vecino más cercano de compatibilidad secuencial y Rama y límite de compatibilidad conjunta (Neira y Tardós 2001). Sin embargo, estas técnicas son similares porque funcionan solo si se dispone de una buena suposición inicial del robot en el mapa (Cummins y Newman 2008).

6 Soluciones al problema visual de SLAM

Las técnicas utilizadas para resolver el problema visual de SLAM se pueden dividir en tres grandes grupos: (a) clásicas, basadas en filtros probabilísticos, con los que el sistema mantiene una representación probabilística tanto de la pose del robot como de la ubicación de los hitos en el entorno, (b) las técnicas que emplean Structure from Motion (SfM) de manera incremental (causal) y, finalmente, (c) las técnicas inspiradas en la biología. En las siguientes secciones se describen algunos detalles de cada una de estas técnicas.

6.1 Filtros probabilísticos

La mayoría de las soluciones SLAM reportadas hasta la fecha se basan en técnicas probabilísticas. Algunos de estos son: el filtro de Kalman extendido (EKF), la solución factorizada para SLAM (FastSLAM), la probabilidad máxima (ML) y la maximización de la expectativa (EM) (Thrun y col. 2005b). Las dos primeras técnicas enumeradas anteriormente son las más utilizadas porque ofrecen los mejores resultados cuando minimizan conjuntamente las incertidumbres de la entidad y el mapa. Estos enfoques tienen éxito a pequeña escala, pero tienen una capacidad limitada para navegar en entornos grandes o para agregar información al cierre de bucles.

Una metodología para construir mapas de una manera incremental (causal), fue presentada por primera vez en el trabajo de Smith y col. (1990). Smith y col. (1990) introdujo el concepto de mapa estocástico y desarrolló una solución precisa al problema SLAM utilizando el Filtro Kalman Extendido. El enfoque de SLAM basado en EKF se caracteriza por un vector de estado compuesto por la ubicación de la entidad y algunos elementos del mapa, estimados de forma recursiva a partir de los modelos no lineales de observación y transición. La incertidumbre está representada por funciones de densidad de probabilidad (*pdfs).* Se supone que la propagación recursiva de la media y covarianza de estos *pdfs* están cerca de la solución óptima. El EKF tiene la desventaja de ser particularmente sensible a las malas asociaciones, una medición incorrecta puede conducir a la divergencia de todo el filtro. La complejidad de EKF es cuadrática con respecto al número de puntos de referencia en el mapa, siendo difícil mantener mapas grandes. En la literatura existen diferentes métodos para reducir esta complejidad a través de técnicas como: Atlas Framework (Bosse y col. 2003), Filtro de Kalman extendido comprimido (CEKF) (Guivant 2002), Filtro de información extendida dispersa (SEIF) (Thrun y col. 2002), Divide y vencerás en *En)* dada por Paz y col. (2008) o submapas condicionalmente independientes (CI-Submaps) desarrollados por Piniés y Tardós (2008). FastSLAM fue propuesto por Montemerlo y col. (2002) y posteriormente mejorado en (Montemerlo 2003). Este método mantiene una distribución de pose de entidad como un conjunto de partículas Rao-Blackwellized, donde cada partícula representa una trayectoria de la entidad, mantiene su propio mapa usando

el EKF, tiene una hipótesis sobre la asociación de datos (hipótesis múltiples) y sobrevive con una probabilidad. El algoritmo consta de un proceso de generación de partículas y un proceso de remuestreo, para evitar la degeneración de las partículas con el tiempo. El costo computacional de esta solución es logarítmico, *O (plogn),* dónde *pag* es el número de partículas utilizadas y *norte* es el número de puntos de referencia en el mapa. Su principal problema es que no hay forma de determinar el número de partículas necesarias para representar con precisión la posición de la entidad. Es decir, muchas partículas requieren mucha memoria y tiempo de cálculo, pero pocas partículas dan lugar a resultados inexactos.

Davison (2003) fue el primero en presentar un sistema probabilístico monocular en tiempo real, que llamó MonoSLAM. Esta técnica de SLAM, realiza simultáneamente un mapeo métrico 3D de puntos y ubicación a 30 fotogramas por segundo, utilizando solo una cámara digital con cable de fuego (IEEE-1394). Considera el movimiento completo de la cámara (6gdl): posición (x, y, z) y orientación (cabeceo, guiñada y balanceo). El trabajo de Davison tiene la limitación de trabajar solo en espacios cerrados e interiores, ya que emplea el EKF para estimar datos.

El sistema MonoSLAM utiliza un modelo de movimiento con velocidades lineales y angulares constantes. Esto representa un inconveniente debido a la incapacidad del modelo para lidiar adecuadamente con los movimientos bruscos, lo que limita la movilidad de la cámara. Por tanto, la distancia a la que se pueden mover las características destacadas entre los fotogramas es muy pequeña, con el fin de garantizar el seguimiento (de lo contrario, podría resultar muy costoso, ya que se propone una gran región para buscar características).

Para hacer frente a movimientos erráticos de la cámara con MonoSLAM, Gee y col. (2008) desarrollado una versión optimizada, capaz de operar a 200 Hz utilizando un modelo de movimiento extendido que tiene en cuenta la aceleración y las velocidades lineales y angulares; sin embargo, su rendimiento en tiempo real se limita a unos pocos segundos, debido a que el tamaño del mapa y el costo computacional crecen extremadamente rápido.

Para aumentar el número de puntos de referencia mantenidos en el mapa, Eade y Drummond (2006b) utilizó una técnica de filtro de partículas inspirada en el método propuesto por Montemerlo y col. (2002), FastSLAM. El método de Eade y Drummond es capaz de rastrear hasta 30 características por cuadro de video y mantener mapas densos de miles de puntos de referencia. Clemente y col. (2007) proponen una alternativa para utilizar el MonoSLAM en grandes entornos al aire libre. Este enfoque se basa en una técnica de mapeo jerárquico y un algoritmo de asociación de datos robusto basado en restricciones geométricas Branch and Bound (GCBB) capaz de realizar grandes cierres de bucles (250 m aprox.).

Como se mencionó anteriormente, un problema en el SLAM visual monocular es la inicialización de los puntos de referencia, porque su profundidad no se puede calcular a partir de una sola observación. Para esto, Davison (2003) utiliza una técnica de inicialización retardada, mientras que Montiel y col. (2006) proponen una técnica llamada parametrización de profundidad inversa, que realiza una inicialización de hito sin retardo en un sistema EKF-SLAM desde el primer momento en que se detectan.

6.2 Estructura por movimiento

Las técnicas de Structure from Motion (SfM) permiten calcular la estructura 3D de la escena y la posición de la cámara a partir de un conjunto de imágenes (Pollefeys y col. 2004). SfM tiene sus orígenes en la fotogrametría y la visión por computadora. El procedimiento estándar (realizado fuera de línea) consiste en extraer características destacadas de las imágenes entrantes, hacerlas coincidir y realizar una optimización no lineal denominada *Ajuste de paquete (* BA) para minimizar el error de reproyección (Triggs y col. 1999 ; Engels et al. 2006).

SfM permite una alta precisión en la ubicación de las cámaras pero no necesariamente tiene la intención de crear mapas consistentes. A pesar de ello, se han realizado varias propuestas utilizando SFM para localizar con precisión y al mismo tiempo crear una buena representación del entorno.

Un método para resolver el problema de SfM de forma incremental es el *odometría visual* publicado por Nistér y col. (2004). La odometría visual consiste en determinar simultáneamente la pose de la cámara para cada fotograma de video y la posición de las características en el mundo 3D, utilizando solo imágenes de forma causal y en tiempo real. Mouragnon y col. (2006 , 2009) utiliza una odometría visual similar a la propuesta de Nister, pero agrega una técnica llamada Ajuste de paquete local, que informa trayectorias de hasta 500 m. La odometría visual permite trabajar con miles de características por cuadro, mientras que las técnicas probabilísticas manejan solo unas pocas características.

Klein y Murray (2007) presentan un método monocular llamado seguimiento paralelo y Mapeo (PTaM). Utiliza un enfoque basado en fotogramas clave (consulte el "Apéndice I") con dos subprocesos de procesamiento en paralelo. El primer hilo de ejecución realiza la tarea de realizar un seguimiento robusto de muchas características, mientras que el otro produce un mapa de puntos en 3D con la ayuda de técnicas BA. El sistema PTaM presenta fallas de rastreo en presencia de texturas similares y objetos en movimiento.

En (Konolige y Agrawal 2008 ; Konolige y col. 2009) los autores utilizan una técnica llamados FrameSLAM y mapas basados en vistas, respectivamente. Los dos métodos se basan en hacer una representación del mapa como un "esqueleto" que consiste en un gráfico de restricción no lineal entre cuadros (en lugar de características 3D individuales). Los autores utilizan un dispositivo estéreo montado en un robot con ruedas. Sus resultados muestran un buen comportamiento en trayectorias largas (aproximadamente 10 km) en condiciones cambiantes como el paso por un entorno urbano.

Recientemente Strasdat y col. (2010b) han reconocido que para aumentar la precisión de la posición de un sistema SLAM monocular se recomienda aumentar el número de características (propiedad esencial de SfM) en lugar del número de fotogramas; además, que las técnicas de optimización de Bundle Adjustment son mejores que los filtros. Sin embargo, manifiestan que el filtro puede ser beneficioso en situaciones de alta incertidumbre. El sistema SLAM ideal aprovecharía los beneficios tanto de las técnicas SfM como de los filtros probabilísticos.

6.3 Modelos bioinspirados

Milford y col. (2004) utilizan modelos del hipocampo (responsable de la memoria espacial) de roedores para crear un sistema de localización y mapeo llamado RatSLAM. RatSLAM puede generar representaciones consistentes y estables de entornos complejos utilizando una sola cámara. Los experimentos llevados a cabo en (Milford y Wyeth 2008 ; Glover y col. 2010) muestra un buen desempeño en tareas en tiempo real tanto en ambientes interiores como exteriores. Además tiene la capacidad de cerrar más de 51 bucles de hasta 5 km de longitud y en diferentes horas del día. En (Milford 2008) se presenta un estudio más amplio de RatSLAM y otros sistemas biológicos y de navegación de abejas, hormigas, primates y humanos.

Collett (2010) examina el comportamiento de las hormigas en el desierto para analizar cómo son guiadas por puntos de referencia visuales y no rastros de feromonas. Aunque esta investigación se centra en comprender cómo navegan los usuarios utilizando información visual, el autor afirma que la solución propuesta sería viable y fácil de implementar en un robot.

7 Representación del mundo observado

Cartografía es hoy en día un área de investigación muy activa. Los espacios libres y ocupados del entorno (obstáculos) se representan en mapas mediante una representación geométrica. Hay diferentes tipos de mapas reportados en la literatura, ampliamente divididos en *métrico* y *topológico* mapas.

Los mapas métricos capturan las propiedades geométricas del entorno, mientras que los mapas topológicos describen la conectividad entre diferentes ubicaciones.

En la categoría de mapas métricos se pueden considerar los mapas de cuadrícula de ocupación (Gutmann y col. 2008) y mapas basados en puntos de referencia (Klein y Murray 2007 ; Se et al. 2002 ; Sáez y Escolano 2006 ; Mouragnon y col. 2006). Los mapas de cuadrícula modelan el espacio libre y ocupado mediante una discretización del entorno en forma de celdas, que pueden contener información 2D, 2.5D o 3D. Los mapas basados en puntos de referencia identifican y mantienen la ubicación en 3D de determinadas características destacadas del entorno. Thrun (2002) realiza un estudio detallado sobre el tema del mapeo robótico utilizando técnicas probabilísticas en ambientes interiores.

Con la representación a través de hitos, solo se capturan hitos aislados de la estructura del entorno, minimizando así, los recursos de memoria y los costos de computación. Por lo anterior, este tipo de mapas no son ideales para evitar obstáculos o planificar caminos, ya que la falta de un hito en un lugar no implica que el espacio sea libre. Sin embargo, cuando la determinación de la pose de la entidad es más importante que el mapa, estas representaciones son las más adecuadas.

Los mapas topológicos representan el medio ambiente como una lista de lugares significativos que están conectados por arcos (similar a un gráfico) (Fraundorfer y col. 2007 ; Eade y Drummond 2008 ; Konolige y col. 2009 ; Botterill y col. 2010). Una representación del mundo basada en gráficos simplifica el problema de mapear grandes extensiones. Sin embargo, es necesario realizar una optimización global del mapa para reducir el error local (Frese y col. 2005 ; Olson y col. 2006). Se puede consultar un tutorial para formular el problema SLAM mediante gráficas en (Grisetti y col. 2010). Otros esquemas relevantes basados en gráficos son los siguientes: Konolige y Agrawal (2008), Konolige y col. (2009) construyó una secuencia de poses relativas entre fotogramas, que pueden recuperarse de errores críticos. Muestran resultados en trayectorias de 10 km utilizando visión estéreo, aunque requiere posiciones generadas por un sensor IMU (Unidad de medida inercial) cuando se produce una oclusión de las cámaras. Los autores afirman que su esquema es aplicable a SLAM monocular, aunque no está demostrado. Otra alternativa es presentada por Mei y col. (2009), que logra mantener una complejidad constante en el tiempo para optimizar los submapas locales que consisten en los nodos más cercanos mediante una técnica llamada ajuste de paquete relativo. Generan una trayectoria de aproximadamente 2km, a través de cámaras estéreo.

Una limitación de la representación topológica es la falta de información métrica, por lo que es imposible utilizar el mapa con el propósito de guiar a un robot. Como consecuencia, Bazeille y Filliat (2010); Angeli y col. (2009) y Konolige y col. (2011) proponen estrategias para mezclar información métrica y topológica en un único modelo consistente.

Actualmente, las representaciones ambientales más prometedoras se basan en gráficos. Pero aún quedan una serie de retos por superar, como la capacidad de editar el gráfico al detectar estimaciones erróneas de la posición, o la generación de mapas globales de dimensiones muy grandes.
(*mapeo de por vida).*

En el “Apéndice III” se describen varios conjuntos de datos que contienen secuencias de imágenes reales para la evaluación de sistemas SLAM visuales.

Las características clave de algunos sistemas SLAM visuales revisados en este documento se resumen en la Tabla 1 . Específicamente, informamos: (1) el nombre del autor y su referencia respectiva, (2) el tipo de dispositivo de detección utilizado, (3) el núcleo de la solución SLAM visual, (4) el tipo de representación del entorno, (5) detalles del proceso de extracción de características, (6) la capacidad y robustez del sistema para operar en una variedad de condiciones: objetos en movimiento, movimientos abruptos y entornos grandes, y también para realizar cierres de bucles, y (7) el tipo de entorno utilizado para probar el rendimiento del sistema.

tabla 1 Resumen de algunos sistemas revisados

Autor	Tipo de dispositivo sensor	Núcleo de la solución	Tipo de mapa	Extracción de características	
				Detector	Descriptor
Davison (2003)	Cámara monocular	MonoSLAM (EKF)	Métrico	Esquinas de Harris del operador	Parches de imagen
Nistér y col. (2004)	Cámaras estéreo o monoculares	Odometría visual	Métrico	de Shi y Tomasi	Parches de imagen
Sáez y Escolano (2006)	Cámara estéreo	Entropía global Minimización Algoritmo	Métrico	Operador de Nitzberg	Parches de imagen
Mouragnon y col. (2006)	Cámara monocular	Vi + odometría sual Paquete local ajustamiento	Métrico	Esquinas de Harris	Parches de imagen
Klein y Murray (2007)	Cámara monocular	Seguimiento paralelo y mapeo (Visual odometria + Manojo ajustamiento)	Métrico	Rápido-10	Parches de imagen
Ho y Newman (2007)	Monóculo camara y laser	Estado retrasado formulación	Métrico	Regiones afines de Harris	Tamiz 128D
Clemente y col. (2007)	Cámara monocular	Mapa jerárquico + EKF	Métrico	Shi y Tomasi operador	Parches de imagen
Lemaire y col. (2007)	Estéreo o monocular cámaras	EKF	Métrico	Esquinas de Harris	Parches de imagen
Milford (2008)	Cámara monocular	RatSLAM (modelos del roedor hipocampo)	Topológico	Basado en apariencia pareo	
Scaramuzza y Siegwart (2008)	Cámara omnidireccional	Odometria visual	Métrico	SIFT (diferencia de gaussianos)	Parches de imagen
Eade y Drummond (2008)	Cámara monocular	GraphSLAM	Topológico	Extremos del espacio de escala detector	Tamiz 16D
Paz y col. (2008)	Cámara estéreo	Condicionalmente independiente dividir y conquistar (EKF)	Métrico	Shi y Tomasi operador	Parches de imagen

tabla 1 continuado

Autor	Tipo de sintiendo dispositivo	Centro de la solución	Tipo de mapa	Característica extracción	
				Detector	Descriptor
Angeli y col. (2008)	Cámara monocular	EKF	A + pológico Métrico	SIFT (diferencia de los gaussianos)	128D SIFT + histogramas de tonos locales
Cummins y Newman (2008)	Monóculo Cámara montado en un *Pan Tilt*	Apariencia rapida Mapeo basado (FAB-MAPA)	Topológico	Harris-Af fi ne	U-SURF 128D
Piniés y Tardós (2008)	Cámara monocular	Condicionalmente independiente mapas locales (EKF)	Métrico	Esquinas de Harris	Parches de imagen
Konolige y col. (2009)	Cámara estéreo + IMU	Vi + odometría sual Paquete escaso ajustamiento	Topológico	Rápido	Firmas de árboles al azar
Williams (2009)	Cámara monocular	Hola + mapa jerárquico EKF + Visual odometria	Métrico	Rápido	Parches de imagen + SIFT 16D
Kaess y Dellaert (2010)	Plataforma multicámara	Expectativa maximización + Paquete estándar ajustamiento	Métrico	Esquinas de Harris	Parches de imagen
Botterill y col. (2010)	Cámara monocular	Od + ometría visual Bolsa de palabras	Topológico	Rápido	Parches de imagen
Mei y col. (2010)	Cámara estéreo	Vi + odometría sual Relativo manojo ajuste + FAB-MAP	Topológico + Métrico	Rápido	Tamiz 128D

tabla 1 continuado

Autor	Tipo de detección dispositivo	Centro de la solución	Tipo de mapa	Poder con				Tipo de ambiente
				Moviente ¿objetos?	Círculo cierre ¿eventos?	El secuestro d problema del robot lem?	Gran escala ¿cartografía?	
Davison (2003)	Cámara monocular	MonoSLAM (EKF)	Métrico	No	No	No	No	Interior
Nistér y col. (2004)	Estéreo o monocular cámaras	Odometría visual	Métrico	No	No	No	No	Exterior
Sáez y Escolano (2006)	Cámara estéreo	Entropía global Minimización Algoritmo	Métrico	No	No	No	No	Exterior Interior
Mouragnon y col. (2006)	Cámara monocular	Vi + odometría sual Paquete local ajustamiento	Métrico	sí	No	No	sí	Exterior Interior
Klein y Murray (2007)	Cámara monocular	Seguimiento paralelo y mapeo (Visual odometria + Manojo ajustamiento)	Métrico	No	No	No	No	Interior
Ho y Newman (2007)	Monóculo camara y laser	Estado retrasado formulación	Métrico	No	sí	No	sí	Exterior Interior
Clemente y col. (2007)	Cámara monocular	Hola + mapa jerárquico EKF	Métrico	sí	sí	No	sí	Exterior
Lemaire y col. (2007)	Estéreo o monocular cámaras	EKF	Métrico	No	sí	No	No	Exterior
Milford (2008)	Cámara monocular	RatSLAM (modelos del roedor hipocampo)	Topológico	Sí	sí	sí	sí	Exterior
Scaramuzza y Siegwart (2008)	Omnidireccional cámara	Odometria visual	Métrico	No	No	No	sí	Exterior
Eade y Drummond (2008)	Cámara monocular	GraphSLAM	Topológico	Sí	sí	sí	sí	Exterior Interior

tabla 1 continuado

Autor	Tipo de sintiendo dispositivo	Centro de la solución	Tipo de mapa	Poder con				Tipo de ambiente
				Moviente ¿objetos?	Círculo cierre ¿eventos?	El secuestro d problema del robot lem?	Gran escala ¿cartografía?	
Paz y col. (2008)	Cámara estéreo	Condicionalmente independiente dividir y conquistar (EKF)	Métrico	No	No	No	sí	Exterior Interior
Angeli y col. (2008)	Cámara monocular	EKF	Topológico + Métrico	No	sí	No	No	Interior
Cummins y Hombre nuevo (2008)	Monóculo Cámara montado en un *Pan Tilt*	Apariencia rapida Mapeo basado (FAB-MAPA)	Topológico	sí	sí	No	sí	Exterior
Piniés y Tardós (2008)	Cámara monocular	Condicionalmente local independiente mapas (EKF)	Métrico	sí	sí	No	sí	Exterior
Konolige y col. (2009)	Cámara estéreo + IMU	Vi odometría sual Paquete escaso ajustamiento	Topológico	sí	sí	sí	sí	Exterior
Williams (2009)	Cámara monocular	Hola mapa jerárquico EKF + Visual odometria	Métrico	sí	sí	sí	sí	Exterior
Kaess y Dellaert (2010)	Plataforma multicámara	Expectativa maximización + Paquete estándar ajustamiento	Métrico	No	sí	No	No	Exterior
Botterill y col. (2010)	Cámara monocular	Od + ometría visual Bolsa de palabras	Topológico	sí	sí	sí	sí	Exterior Interior
Mei y col. (2010)	Cámara estéreo	Odometría visual + Paquete relativo ajuste + FAB-MAP	Topológico + Métrico	Sí	sí	sí	sí	Exterior

8 Conclusiones

Este trabajo comprueba que existe una gran preocupación por solucionar el problema SLAM utilizando la visión como único sensor exteroceptivo. Esto se debe principalmente a que una cámara es un sensor ideal, ya que es ligero, pasivo, tiene un bajo consumo de energía y captura información abundante y distintiva de una escena. Sin embargo, el uso de la visión requiere algoritmos confiables con buen desempeño y consistentes bajo condiciones de luz variables, oclusiones o cambios en la apariencia del entorno debido al movimiento de personas u objetos, aparición de regiones sin rasgos distintivos, transiciones entre el día y la noche o cualquier otra situación imprevista. . Por lo tanto, los sistemas SLAM que utilizan la visión como único sensor siguen siendo un área de investigación desafiante y prometedora.

La coincidencia de imágenes y la asociación de datos siguen siendo áreas de investigación abiertas en los campos de la visión por computadora y la visión robótica, respectivamente. El detector y el descriptor elegido afectan directamente el rendimiento del sistema para rastrear las características destacadas, reconocer áreas vistas anteriormente, construir un modelo consistente del entorno y trabajar en tiempo real. Particular a la asociación de datos es la necesidad de navegación a largo plazo, a pesar de una base de datos en crecimiento y entornos cambiantes y extremadamente curiosos. La aceptación de una mala asociación provocará graves errores en todo el sistema SLAM, lo que significa que tanto el cálculo de la ubicación como la construcción del mapa serán inconsistentes. Por tanto, es importante proponer nuevas estrategias para reducir la tasa de falsos positivos.

Los métodos basados en apariencia han sido muy populares para resolver problemas de asociación de datos en SLAM visual. La técnica más común en esta categoría es el BoVW, debido a su rapidez para encontrar imágenes similares. Sin embargo, el BoVW se ve afectado por el fenómeno del aliasing perceptual. Asimismo, esta técnica aún no ha sido probada a fondo para detectar imágenes con grandes variaciones de punto de vista o escala, que son transformaciones que suelen ocurrir durante la detección de cierre de bucle, el problema del robot secuestrado y el mapeo multisesión y cooperativo. Además, no tiene en cuenta la distribución espacial entre las características detectadas y la información geométrica 3D, lo que podría ser útil a la hora de establecer asociaciones.

Aunque ha habido varias propuestas para construir mapas para toda la vida, este tema sigue siendo un tema de interés, así como la capacidad de construir mapas a pesar de todos los problemas causados por trabajar en entornos del mundo real.

Hasta la fecha, no existen estándares para evaluar y comparar la eficiencia y efectividad general de un sistema completo de SLAM visual. No obstante, existen varios indicadores que pueden caracterizar su desempeño, como el grado de intervención humana, precisión de ubicación, consistencia del mapa, operación en tiempo real y el control del costo computacional que surge con el crecimiento del mapa, entre otros.

Expresiones de gratitud Este trabajo ha sido posible gracias al generoso apoyo de las siguientes instituciones a las que nos complace reconocer: CONACYT (Consejo Nacional de Ciencia y Tecnología) y CENIDET (Centro Nacional de Investigación y Desarrollo Tecnológico).

Apéndice I: Fotogramas clave

Un fotograma clave es un fotograma de video que es lo suficientemente diferente de su predecesor en la secuencia, para representar una nueva ubicación. Los fotogramas clave también se utilizan para estimar de manera eficiente la pose de la cámara y reducir la redundancia de información. La forma más sencilla de clasificar un fotograma de video como fotograma clave es comparar un fotograma de video con respecto a otro tomado anteriormente, seleccionando aquellos que maximizan tanto la distancia a la que fueron capturados como el número de características.

coincidencias que existen entre ellos. En (Zhang y col. 2010) Se presenta un estudio comparativo de diferentes técnicas para detectar keyframes orientados al problema visual SLAM.

Apéndice II: Bolsa de palabras visuales (BoVW)

Recientemente, la mayoría de las contribuciones para resolver la asociación de datos en SLAM visual utilizan BoVW (Sivic y Zisserman 2003) y su versión mejorada llamada Árbol de vocabulario (Nistér y Stewenius 2006). El BoVW ha tenido un gran éxito en el área de recuperación de información (Manning y col. 2008) y recuperación de imágenes basada en contenido desarrollada por la comunidad de visión por computadora, debido a su velocidad para encontrar imágenes similares. Sin embargo, esta técnica no es del todo precisa porque detecta varios falsos positivos. Para solucionar este problema hasta cierto punto, la información espacial se introduce normalmente en la última fase de recuperación, realizando una postverificación teniendo en cuenta la restricción epipolar (Angeli y col. 2008) o, recientemente, mediante campos aleatorios condicionales (Calonder y col. 2010). Esta verificación permite rechazar aquellas imágenes recuperadas que no son geométricamente consistentes con la imagen de referencia.

El modelo clásico de BoVW describe las imágenes como un conjunto de características locales llamadas palabras visuales y el conjunto completo de estas palabras se conoce como vocabulario visual. Muchos esquemas BoVW generan un vocabulario fuera de línea mediante un agrupamiento de K-medias (pero se puede usar cualquier otro) de descriptores de un gran corpus de imágenes de entrenamiento (Ho y Newman 2007 ; Cummins y Newman 2008). Un enfoque alternativo y más efectivo es construir dinámicamente el vocabulario a partir de las características que se encuentran a medida que se explora el entorno. Tal esquema es descrito por Angeli y col. (2008) y Botterill y col. (2010).

Algunas palabras visuales son más útiles que otras para identificar si dos imágenes muestran el mismo lugar. El esquema más común para asignar a cada palabra un peso específico es el TF-IDF. Combina la importancia de las palabras en la imagen (TF- Término Frecuencia) y la importancia de las palabras en la colección (IDF- Frecuencia Inversa del Documento). Además, existen otros esquemas, los cuales se dividen en local (TF al cuadrado, Logaritmo de frecuencia, Binario, BM25 TF, entre otros) y global (IDF probabilístico, IDF al cuadrado, etc.) (Tirilly y col. 2010). Se utiliza un índice invertido para acelerar las consultas, que organiza todo el conjunto de palabras visuales que representan imágenes. Un índice invertido está estructurado como un índice de libros. Tiene una entrada para cada palabra de la colección de imágenes, seguida de una lista de todas las imágenes en las que está presente la palabra.

Apéndice III: Conjuntos de datos para probar sistemas visuales SLAM

Algunos conjuntos de datos públicos disponibles para probar los sistemas SLAM visuales son: (a) Conjuntos de datos de New College y City Center (al aire libre) (Cummins 2008), usado por Cummins y Newman (2008); (b) El conjunto de datos láser y de visión de NewCollege (exterior) (Smith 2012), Capturado por Smith y col. (2009) (c) Bovisa (exterior) y Bicocca (interior) Conjunto de datos del proyecto Rawseeds (Semillas crudas 2012), Capturado por Ceriani y col. (2009); (d) El conjunto de datos de Cheddar Gorge (exterior), capturado por Simpson y col. (2012) y conjuntos de datos RGB-D (interior) (Sturm 2012) (Sturm y col. 2011).

Referencias

Aguilar W, Frauel Y, Escolano F et al (2009) Un gráfico robusto que coincide con el registro no rígido. Vis de imagen Computación 27 (7): 897–910

Andrade J, Sanfeliu A (2002) Construcción y localización de mapas concurrentes con validación de puntos de referencia. En: Procedidos de la 16ª conferencia internacional de la IAPR sobre reconocimiento de patrones, vol 2, págs. 693–696

Angeli A, Doncieux S, Filliat D (2008) Detección visual de cierre de bucle en tiempo real. En: Actas del IEEE conferencia internacional sobre robótica y automatización

Angeli A, Doncieux S, Meyer J (2009) SLAM topológico visual y localización global. En: Actas de la conferencia internacional IEEE sobre robótica y automatización, págs. 4300–4305

Artieda J, Sebastian J, Campoy P et al (2009) SLAM 3D visual de UAV. J Intell Robot Syst 55 (4): 299–321

AsmarD (2006) SLAM inercial de visión usando características naturales en ambientes al aire libre. Disertación, Universidad de Waterloo, Canadá

Auat C, Lopez N, Soria C, et al (2010) Algoritmo SLAM aplicado a la asistencia robótica para la navegación en entornos desconocidos. J Neuroeng Rehabil. doi: 10.1186 / 1743-0003-7-10

Bailey T, Durrant H (2006) Localización y cartografía simultánea (SLAM): Parte II. IEEERobot AutomMag 13 (3): 108-117

Bay H, Tuytelaars T, Van L (2006) SURF: características robustas aceleradas. En: Proceedings of the European conferencia sobre visión artificial

Bazeille S, Filliat D (2010) Combinando odometría y detección visual de cierre de bucle para topo-metri- mapeo cal. RAIRO Int J Oper Res 44 (4): 365–377

Beis J, LoweD (1997) Indización de formas utilizando la búsqueda del vecino más cercano aproximado en espacios de alta dimensión. En: Actas de la conferencia IEEE sobre visión por computadora y reconocimiento de patrones, págs. 1000–1006

Bogdan R, Sundaresan A, Morisset B et al (2009) Leaving fl atland: e fi cient real-time tridimensional percepción y planificación de movimiento. Robot de campo J. Número especial sobre cartografía tridimensional 26 (10): 841-862

Bosse M, Newman P, Leonard J, et al (2003) Un marco de atlas para mapeo escalable. En: Actas del Conferencia internacional IEEE sobre robótica y automatización, págs. 1899–1906

Botterill T, Mills S, Green R (2010) Localización y mapeo simultáneos de una sola cámara basada en una bolsa de palabras silbido. Robot de campo J 28 (2): 204–226

Bouguet (2010) Caja de herramientas de calibración de cámara para matlab. http://www.vision.caltech.edu/bouguetj/calib_doc/ . Consultado el 06 de marzo de 2012.

Brown MZ, Burschka D, Hager G (2003) Avances en estéreo computacional. IEEE Trans Pattern Anal Mach Intell 25 (8): 993–1008

Cadena C, Gálvez-López D, Ramos F, et al (2010) Reconocimiento robusto de lugares con cámaras estéreo. En: Proceedings de la conferencia internacional IEEE sobre robots y sistemas inteligentes, págs. 5182–5189

Calonder M, Lepetit V, et al (2010) BREVE: características elementales independientes robustas binarias. En: Actas de la conferencia europea sobre visión artificial

Cannons K (2008) Una revisión del seguimiento visual. Informe técnico CSE-2008-07, Departamento de la Universidad de York de Ciencias de la Computación e Ingeniería

Carrera G, Angeli A, Andrew D (2011) Calibración extrínseca automática basada en SLAM de un equipo multicámara. En: Actas de la conferencia internacional IEEE sobre robótica y automatización

Castellanos J, Tardós JD, Neira J (2001) Fusión multisensor para localización simultánea y construcción de mapas. IEEE Trans Robot Autom 17 (6): 908–914

Ceriani S, Fontana G, Giusti A et al (2009) Sistemas de recolección de la verdad del suelo de Rawseeds para la auto-localización en interiores zation y mapeo. J Auton Robots 27 (4): 353–371

Chatila R, Laumond J (1985) Referenciación de posición y modelado mundial coherente para robots móviles. En: Actas de la conferencia internacional IEEE sobre robótica y automatización, vol 2, págs. 138-145 Chekhlov D,

Mayol W, Calway A (2007) Ninja en un avión: descubrimiento automático de planos físicos para agosto realidad mentada utilizando visual SLAM. En: Actas del sexto simposio internacional de IEEE y ACM sobre realidad mixta y aumentada, págs. 1-4

Chekhlov D, Mayol W, Calway A (2008) Indexación basada en la apariencia para la relocalización en visual en tiempo real GOLPE. En: Actas de la conferencia británica sobre visión artificial, págs. 363–372

Chli M, Davison A (2008) Emparejamiento activo. En: Actas de la conferencia europea sobre visión artificial: parte I. doi: 10.1007 / 978-3-540-88682-2_7

Chli M, Davison A (2009) Coincidencia activa para seguimiento visual. Robot Autonom Syst 57 (12): 1173–1187 Ciganek B,

Siebert J (2009) Una introducción a las técnicas y algoritmos de visión por computadora en 3D. Wiley, nuevo York, págs. 194-195

Clemente L, Davison A, Reid I, et al (2007) Mapeo de bucles grandes con una sola cámara de mano. En: Procimientos de la robótica: conferencia de ciencia y sistemas

Collett M (2010) Cómo las hormigas del desierto utilizan un punto de referencia visual como guía a lo largo de una ruta habitual. En: Psychol Cogni Sci 107 (25): 11638-11643

Cummins (2008) Nuevo conjunto de datos del centro de la ciudad y la universidad. http://www.robots.ox.ac.uk/~mobile/IJRR_2008_ Conjunto de datos . Consultado el 06 de marzo de 2012
CumminsM, Newman P (2008) FAB-MAP: localización probabilística y mapeo en el espacio de aparición. Int J Robot Res 27 (6): 647–665
Cyrill S (2009) Mapeo y exploración robóticos. Springer Tracts in Advanced Robotics, vol 55, ISBN: 978-3-642-01096-5
Davison A (2003) Localización y mapeo simultáneos en tiempo real con una sola cámara. En: Actas de la conferencia internacional IEEE sobre visión artificial, vol2, págs. 1403–1410
Davison A, González Y, Kita N (2004) SLAM 3D en tiempo real con visión gran angular. En: 5o IFAC / EURON simposio sobre vehículos autónomos inteligentes
Davison A, Reid I, Molton N (2007) MonoSLAM: SLAM de cámara única en tiempo real. IEEE Trans Patrón Anal Mach Intell 29 (6): 1052–1067
Dufournaud Y, Schmid C, Horaud R (2004) Coincidencia de imágenes con ajuste de escala. Comput Vis Image Underst 93 (2): 175-194
Durrant H, Bailey T (2006) Localización y mapeo simultáneos (SLAM): parte I los algoritmos esenciales. IEEE Robot AutomMag 13 (2): 99–110
Eade E, Drummond T (2006a) Marcas de borde en SLAM monocular. En: Actas de la máquina británica conferencia de visión
Eade E, Drummond T (2006b) SLAM monocular escalable. En: Actas de la conferencia IEEE sobre visión por ordenador y reconocimiento de patrones, vol 1, págs. 469–476
Eade E, DrummondT (2008) Cierre y recuperación de bucle unificado para SLAM monocular en tiempo real. En proceso de la conferencia británica de visión artificial
Engels C, Stewénius H, Nistér D (2006) Reglas de ajuste de paquetes. En: Visión por computadora fotogramétrica
Fawcett T (2006) Una introducción al análisis ROC. Patrón de reconocimiento Lett 27 (8): 861–874
Fraundorfer F, Engels C, Nister C (2007) Mapeo topológico, localización y navegación usando colores de imagen lecturas. En: Actas de la conferencia internacional IEEE sobre robots y sistemas inteligentes, págs. 3872–3877
Frese U, Larsson P, Duckett T (2005) Un algoritmo de relajación multinivel para localización simultánea y cartografía. IEEE Trans Robot, págs. 196–207, ISSN 1552-3098
Frintrop S, Jensfelt P (2008) Puntos de referencia de atención y control activo de la mirada para SLAM visual. IEEE Trans Robot 24 (5): 1054–1065
Gee A, Chekhlov D, Calway A, Mayol W (2008) Descubriendo una estructura de nivel superior en SLAM visual. IEEE Trans Robot 24 (5): 980–990
Gemeiner P, Davison A, Vincze M (2008) Mejora de la robustez de la localización en SLAM monocular usando un cámara de alta velocidad. En: Actas de robótica: ciencia y sistemas IV
Gil A, Martínez O, Ballesta M, Reinoso O (2009) Una evaluación comparativa de detectores de puntos de interés y descriptores locales para SLAM visual. Mach Vis Appl 21 (6): 905–920
Gil A, Reinoso O, Ballesta M, Juliá M (2010) SLAM visual multi-robot usando una partícula rao-blackwellizada filtro. Robot Autonom Syst 58 (1): 68–80
Glover A, Maddern W, Milford M, et al (2010) FAB-MAP + RatSLAM: slam basado en la apariencia para múltiples Tiempos del Día. En: Actas de la conferencia internacional IEEE sobre robótica y automatización
Grasa O, Civera J, Montiel J (2011) SLAM monocular EKF con relocalización para secuencias laparoscópicas. En: Actas de la conferencia internacional IEEE sobre robótica y automatización, págs. 4816–4821
Grauman K (2010) Búsqueda eficiente de imágenes similares. Comun ACM 53 (6): 84–94
Grauman K, Darrell T (2007) Hash de coincidencia piramidal: indexación de tiempo sublineal sobre correspondencias parciales. En: Actas de la conferencia IEEE sobre visión por computadora y reconocimiento de patrones
Grisetti G, Kümmerle R, Stachniss C, Burgard W (2010) Un tutorial sobre SLAM basado en gráficos. IEEE Trans Intell Transp Syst Mag 2 (4): 31–43
Gu S, Zheng Y, Tomasi C (2010) Redes críticas y características beta-estables para la coincidencia de imágenes. En: Actas de la conferencia europea sobre visión artificial, págs. 663–676
Guivant J (2002) Localización y mapeo simultáneos eficientes en entornos grandes. Disertación, Universidad de Sydney, Australia
Gutmann J, Fukuchi M, Fujita M (2008) Percepción 3D y generación de mapas ambientales para robots humanoides. Int J Robot Res 27 (10): 1117–1134
Handa A, Chli M, Strasdat H, Davison A (2010) Coincidencia activa escalable. En: Actas del IEEE conferencia sobre visión por computadora y reconocimiento de patrones, págs. 1546–1533
Harris C, Stephens M (1988) Un detector combinado de esquinas y bordes. En: Actas de la cuarta visión de Alvey conferencia, págs. 147–151
Hartley R, Sturm P (1997) Triangulación. Comput Vis Image Underst 68 (2): 146-157

Hartley R, Zisserman A (2003) Geometría de vista múltiple en visión por computadora, 2ª ed. Cambridge, ISBN: 0521540518

Hinterstoisser S, Kutter O, Navab N, et al (2009) Aprendizaje en tiempo real de la rectificación precisa del parche. En: Procedidos de la conferencia IEEE sobre visión por computadora y reconocimiento de patrones

Ho K, Newman P (2007) Detectando el cierre de bucle con secuencias de escenas. Int J Comput Vis 74 (3): 261–286 Huang A,

Bachrach A, Henry P, et al (2011) Odometría visual y mapeo para vuelos autónomos usando rgb-d cámara. Simposio internacional de investigación en robótica

Johnson M, Pizarro O, Williams S, Mahon I (2010) Generación y visualización de tres dimensiones a gran escala reconstrucciones sionales a partir de estudios robóticos submarinos. Robot de campo J 27 (1): 21–51

Jones E, Soatto S (2011) Navegación, mapeo y localización visual-inercial: un factor causal escalable en tiempo real Acercarse. Int J Robot Res 30 (4): 407–430

KaessM, Dellaert F (2010) Comparación de estructuras probabilísticas para SLAM visual con un equipo multicámara. Computación Vis Image Underst 114: 286–296

Kawewong A, Tangruamsub S, Hasegawa O (2010) Características robustas de posición invariante para reconocimiento a largo plazo ción de escenas dinámicas al aire libre. IEICE Trans Inform Syst 9: 2587–2601

Ke Y, Sukthankar R (2004) PCA-SIFT: una representación más distintiva para los descriptores de imágenes locales. En: Actas de la conferencia IEEE sobre visión por computadora y reconocimiento de patrones, vol 2, págs. 506–513 Klein

G, Murray D (2007) Seguimiento y mapeo en paralelo para pequeños espacios de trabajo de RA. En: Actas del 6 Simposio internacional IEEE y ACM sobre realidad mixta y aumentada

Klein G, Murray D (2008) Mejora de la agilidad del SLAM basado en fotogramas clave. En: Proceedings of the European conferencia sobre visión artificial, págs. 802–815

Koch O, Walter M, Huang A, Teller S (2010) Navegación de robots terrestres con cámaras no calibradas. En: Procedidos de la conferencia internacional IEEE sobre robótica y automatización, págs. 2423–2430

Konolige K, Agrawal M (2008) FrameSLAM: del ajuste del paquete al mapeo visual en tiempo real. IEEE Trans Robot 24 (5): 1066–1077

Konolige K, Bowman J, Chen J (2009) Mapas basados en vistas, en: Proceedings of robótica: ciencia y sistemas

Konolige K, Marder-Eppstein E, Marthi B (2011) Navigation in Hybrid métrico-topological maps. En: Procedidos de la conferencia internacional IEEE sobre robótica y automatización

Kragic D, Vincze M (2009) Visión de la robótica. Found Trends Robot 1 (1): 1-78, ISBN: 978-1-60198-260-5 Kulis B, Jain P,

Grauman K (2009) Búsqueda rápida de similitudes para métricas aprendidas. IEEE Trans Pattern Anal Mach Intell 31 (12): 2143–2157

Lemaire T, Berger C, Jung I et al (2007) SLAM basado en la visión: enfoques estéreo y monoculares. Int J Comput Vis 74 (3): 343–364

Lepetit V, Fua P (2005) Seguimiento 3D de objetos rígidos basado en modelos monoculares. Gráfico de cálculo de tendencias encontradas Computación Vis 1 (1): 1–89

Lepetit V, Fua P (2006) Reconocimiento de puntos clave utilizando árboles aleatorios. IEEE Trans Pattern Anal Mach Intell 28 (9): 1465–1479

Li H, Kimi E, Huang X, He L (2010) Coincidencia de objetos con una restricción invariante localmente afín En: Proceedings de la conferencia internacional sobre reconocimiento de patrones, págs. 1641–1648

Lin K, Wang C (2010) Localización, mapeo y seguimiento de objetos en movimiento simultáneos basados en estéreo. En: Actas de la conferencia internacional IEEE sobre robots y sistemas inteligentes, págs. 3975–3980 Lowe D (2004)

Características de imagen distintivas de puntos clave invariantes de escala. Int J Comput Vis 60 (2): 91–110 Magnusson M,

Andreasson H, et al (2009) Detección automática de bucles basada en apariencia a partir de datos láser 3D utilizando la transformación de distribución normal. Robot de campo J. Mapeo tridimensional, parte 2, 26 (12): 892–914

Manning C, Schütze H, Raghavan P (2008) Introducción a la recuperación de información, Cambridge University Press, Cambridge, ISBN: 0521865719

Majumder S, Scheding S, Durrant H (2005) Fusión de sensores y creación de mapas para la navegación submarina. En: Actas de la conferencia australiana sobre robótica y automatización

Martinez J, Calway (2010) Un mapeo plano y puntual unificador en SLAM monocular. En: Actas del Conferencia británica sobre visión artificial, págs. 1-11

Matas J, Chum O, et al (2002) Estéreo de línea de base amplia y robusto de regiones extremas máximamente estables. En: Actas de la conferencia británica sobre visión artificial vol 22, no. 10, págs. 761–767

Mei C, Reid I (2008) Modelado y generación de desenfoque de movimiento complejo para seguimiento en tiempo real. En: Actas de la conferencia IEEE sobre visión por computadora y reconocimiento de patrones, págs. 1-8

Mei C, Sibley G, Cummins M, et al (2009) Un sistema SLAM estéreo eficiente en tiempo constante. En: Actas de la conferencia británica de visión artificial

Mei C, Sibley G, Cummins M et al (2010) RSLAM: un sistema para mapeo a gran escala en tiempo constante usando estéreo. Int J Comput Vision 94 (2): 1–17

Mei C, Sommerlade E, Sibley C, et al (2011) Síntesis de vista oculta usando SLAM visual en tiempo real para simplificar fying análisis de videovigilancia. En: Actas de la conferencia internacional IEEE sobre robótica y automatización, vol 8, págs. 4240–4245

Migliore D, Rigamonti R, Marzorati D, et al (2009) Use una sola cámara para la localización simultánea y mapeo con seguimiento de objetos móviles en entornos dinámicos. En: Taller ICRA sobre navegación segura en entornos abiertos y dinámicos: aplicación a vehículos autónomos

Mikolajcczyk K, Schmid C (2002) Un detector de punto de interés invariante afín. En: Proceedings of the European conferencia sobre visión artificial, págs. 128–142

Mikolajcczyk K, Schmid C (2005) Una evaluación del desempeño de descriptores locales. IEEE Trans Patrón Anal Mach Intell 27 (10): 1615–1630

Mikolajczyk K, Tuytelaars T, Schmid S et al (2005) Una comparación de detectores de regiones afines. Int J Comput Vis 65: 43–72

MilfordM (2008) Navegación robótica desde la naturaleza: simultánea, basada en localización, mapeo y planificación de rutas sobre modelos hipocampales, vol. 41. Springer Tracts in Advanced Robotics, ISBN: 3540775196

Milford M, Wyeth G (2008) Mapeo de un suburbio con una sola cámara usando un SLAM de inspiración biológica sistema. IEEE Trans Robot 24 (5): 1038–1053

Milford M, Wyeth G, Prasser D (2004) RatSLAM: un modelo de hipocampo para localización simultánea y cartografía. En: Actas de la conferencia internacional IEEE sobre robótica y automatización, vol 1, págs. 403–408

Molton N, Davison A, Reid I (2004) Características de parche localmente planas para estructura en tiempo real a partir del movimiento. En: Actas de la conferencia británica sobre visión artificial

Montemerlo M (2003) FastSLAM: una solución factorizada al problema de localización y mapeo simultáneos con asociación de datos desconocidos, disertación, Universidad Carnegie Mellon, EE. UU.

Montemerlo M, Thrun S, Koller D, et al (2002) FastSLAM: una solución factorizada para la localización simultánea problema de mapeo y En: Actas de la conferencia nacional AAAI sobre inteligencia artificial, págs. 593–598

Montiel J, Civera J, Davison A (2006) Parametrización de profundidad inversa unificada para SLAM monocular. En: Procimientos de la robótica: ciencia y sistemas

Morel J, Yu G (2009) ASIFT: un nuevo marco para la comparación de imágenes invariantes totalmente afines. Imágenes SIAM J Ciencia 2 (2): 438–469

Moreels P, Perona P (2005) Evaluación de detectores de características y descriptores basados en objetos 3D. En: Procederings de la conferencia internacional IEEE sobre visión artificial, págs. 800–807

Mouragnon E, Dhome M, Dekeyser F, et al (2006) SLAM basado en visión monocular para robots móviles. En: Actas de la conferencia internacional sobre reconocimiento de patrones, págs. 1027–1031

Mouragnon E, Lhuillier M, Dhome M, et al (2009) Estructura genérica y en tiempo real del movimiento usando local paquete de ajuste. Image Vis Comput, págs. 1178–1193, ISSN: 0262-8856

Neira J, Tardós JD (2001) Asociación de datos en cartografía estocástica mediante la prueba de compatibilidad conjunta. En: Procedidos de la conferencia internacional IEEE sobre robótica y automatización 17 (6): 890–897

Newman P, Leonard J, Neira J, Tardós J (2002) Explore and return: experimental validation of real timecon mapeo y localización actual. En: Actas de la conferencia internacional IEEE sobre robótica y automatización, vol 2, págs. 1802–1809

Nistér D (2004) Una solución eficiente al problema de la pose relativa de cinco puntos. IEEE Trans Pattern Anal Mach Intell 26 (6): 756–770

Nistér D, Stewenius H (2006) Reconocimiento escalable con árbol de vocabulario. En: Actas del IEEE conferencia sobre visión por computadora y reconocimiento de patrones, vol 2, págs. 2161–2168

Nistér D, NaroditskyO, Bergen J (2004) Odometría visual. En: Actas de la conferencia IEEE en computadora visión y reconocimiento de patrones vol 1, págs. 652–659

Nüchter A, Lingemann K, Hertzberg J et al (2007) 6D SLAM: mapeo 3D de entornos al aire libre. Campo J Robot 24 (8): 699–722

Nützi G, Weiss S, Scaramuzza D, Siegwart R (2010) Fusión de IMU y visión para la estimación de escala absoluta en SLAM monocular. J Intell Robot Syst. doi: 10.1007 / s10846-010-9490-z

Olson C, Matthies L, Schoppers M, Maimone M (2003) Navegación móvil con ego-movimiento estéreo. Robot Autonom Syst 43 (4): 215–229

Olson E, Leonard J, Teller S (2006) Optimización iterativa rápida de gráficos de pose con estimaciones iniciales deficientes. En: Actas de la conferencia internacional IEEE sobre robótica y automatización, págs. 2262-2269

Olson C, Matthies L, Wright J et al (2007) Mapeo visual del terreno para la exploración de Marte. Comput Vis Image Underst 105 (1): 73–85

OpenCV (2009) OpenCV: Calibración de cámara y reconstrucción 3D. http://opencv.willowgarage.com/documentation / camera_calibration_and_3d_reconstruction.html Consultado el 06 de marzo de 2012

Özuysal M, Calonder M, Lepetit V, Fua P (2010) Reconocimiento rápido de puntos clave utilizando helechos aleatorios. IEEE Trans Patrón Anal Mach Intell 32 (3): 448–461

Paz L, Piniés P, Tardós JD, Neira J (2008) SLAM 6DOF a gran escala con estéreo en mano. Robot Trans IEEE 24 (5): 946–957

Piniés P, Tardós JD, Neira J (2006) Localización de víctimas de avalanchas mediante SLAM robocéntrico. En: Actas de la conferencia internacional IEEE sobre robots y sistemas inteligentes. págs. 3074–3079

Piniés P, Tardós JD (2008) Construcción de SLAM a gran escala mapas locales condicionalmente independientes: aplicación a visión monocular. IEEE Trans Robot 24 (5): 1094-1106

Pollefeys M, Van L, Vergauwen M et al (2004) Modelado visual con una cámara de mano. Int J Comput Vis 59 (3): 207–232

Pretto A, Menegatti E, Pagello E (2007) Funciones confiables que coinciden con robots humanoides. En: IEEE-RAS conferencia internacional sobre robots humanoides, págs. 532–538

Pupilli M, Calway A (2006) SLAM visual en tiempo real con resistencia al movimiento errático. En: Actas del Conferencia IEEE sobre visión por computadora y reconocimiento de patrones, vol 1, págs. 1244–1249

Raguram R, Frahm J, Pollefeys M (2008) Un análisis comparativo de las técnicas RANSAC que conducen a consenso de muestras aleatorias en tiempo real. En: Actas de la conferencia europea sobre visión artificial, págs. 500–513

Rawseeds (2012) Conjuntos de datos de Bovisa y bicocca. http://www.rawseeds.org/rs/datasets . Consultado el 06 de marzo de 2012 Ribas D, Ridao P, Tardós JD et al (2008) SLAM subacuático en entornos estructurados artificiales. Campo J Robot 25 (11): 898–921

Rosten E, Drummond T (2006) Aprendizaje automático para la detección de esquinas de alta velocidad. En: Actas del Conferencia europea sobre visión artificial, págs. 430–443

Rublee E, Rabaud V, Konolige K, Bradski G (2011) ORB: una alternativa eficiente a SIFT o SURF. En: Actas de la conferencia internacional IEEE sobre visión artificial

Scaramuzza (2011) Caja de herramientas OcamCalib: cámara omnidireccional y caja de herramientas de calibración para matlab. https: // sites.google.com/site/scarabotix/ocamcalib-toolbox . Consultado el 06 de marzo de 2012

Scaramuzza D, Siegwart R (2008) Odometría visual omnidireccional monocular guiada por apariencia para exteriores vehículos terrestres. IEEE Trans Robot 24 (5): 1015–1026

Se S, Lowe D, Little J (2002) Localización y mapeo de robots móviles con incertidumbre utilizando invariantes de escala puntos de referencia visuales. Int J Robot Res 21 (8): 735–758

Se S, Lowe D, Little J (2005) Localización y mapeo global basados en la visión para robots móviles. IEEE Trans Robot 21 (3): 364–375

Sáez J, Escolano F (2006) SLAM de minimización de entropía 6DOF. En: Proceedings of the IEEE international conferencia sobre robótica y automatización, págs. 1548-1555

Sanromá G, Alquézar R, Serratosa F (2010) Emparejamiento gráfico usando descriptores SIFT — una aplicación para posar recuperación de un robot móvil. En: 13º taller internacional conjunto de la IAPR sobre reconocimiento de patrones estructurales, sintácticos y estadísticos, págs. 254–263

Silpa C, Hartley R (2008) Árboles KD optimizados para una rápida coincidencia de descriptores de imágenes. En: Actas del IEEE conferencia sobre visión artificial y reconocimiento de patrones

Sinha S, Frahm J, Pollefeys M, Genc Y (2006) Seguimiento y coincidencia de funciones de video basadas en GPU. En: Taller en la informática de borde utilizando nuevas arquitecturas de productos básicos

Sivic J, Zisserman A (2003) Video google: un enfoque de recuperación de texto para la coincidencia de objetos en videos. En: Proceder- ings de la conferencia internacional IEEE sobre visión por computadora Simpson R, Cullip J, Revell J (2012) The cheddar gorge data set. http://www.openslam.org/misc/BAE_ RSJCJR_2011.pdf . Consultado el 06 de marzo de 2012

Smith (2012) El nuevo conjunto de datos de láser y visión universitaria. http://www.robots.ox.ac.uk/NewCollegeData/ . Consultado el 06 de marzo de 2012

Smith R, Self M, Cheeseman P (1990) Estimación de relaciones espaciales inciertas en robótica. En: Autónomo vehículos robotizados. Springer, Nueva York, págs. 167–193, ISBN: 0-387-97240-4

Smith M, Baldwin I, Churchill W et al (2009) El nuevo conjunto de datos de láser y visión universitaria. Int J Robot Res 28 (5): 595–599

Solà J (2007) VSLAM multicámara: de las antiguas pérdidas de información a la autocalibración. En: Actas de la conferencia internacional IEEE sobre robots y sistemas inteligentes, taller sobre SLAM visual

Steder B, Grisetti G, Stachniss Cet al (2008) Visual SLAM para vehículos voladores. IEEETrans Robot 24 (5): 1088–1093

Sturm (2012) Conjunto de datos RGB-D y punto de referencia. http://cvpr.in.tum.de/data/datasets/rgbd-dataset . Consultado el 06 de marzo de 2012

Sturm J, Magnenat S, et al (2011) Hacia un punto de referencia para la evaluación RGB-D SLAM. En: Actas del Taller de RGB-D sobre razonamiento avanzado con cámaras de profundidad en robótica: conferencia de ciencia y sistemas

Strasdat H, Montiel J, Davison A (2010a) SLAM monocular a gran escala consciente de la deriva a escala. En el procedimiento de robótica: ciencia y sistemas

Strasdat H, Montiel J, Davison A (2010b) SLAM monocular en tiempo real: ¿por qué filtrar ?. En: Actas del Conferencia internacional IEEE sobre robótica y automatización Svoboda (2011) Autocalibración multicámara. http://cmp.felk.cvut.cz/~svoboda/SelfCal/index.html Accedido 06 de marzo de 2012

Tardós JD, Neira J, Newman P et al (2002) Mapeo y localización robustos en ambientes interiores usando datos de la sonda. Int J Robot Res 21: 311–330

Taylor S, Drummond T (2009) Localización de múltiples objetivos a más de 100 FPS. En: Proceedings of the British conferencia de visión artificial

Thrun S (2002) Mapeo robótico: una encuesta. Explorando la inteligencia artificial en el nuevo milenio, ISBN: 1-55860-811-7

Thrun S (2003) Un sistema para el mapeo robótico volumétrico de minas abandonadas. En: Actas del IEEE conferencia internacional sobre robótica y automatización, vol 3, págs. 4270–4275

Thrun S, Leonard J (2008) Localización y mapeo simultáneo. Manual de robótica de Springer; Siciliano, Editores de Khatib, ISBN: 978-3-540-23957-4, págs. 871–886

Thrun S, Koller D, Ghahramani Z, et al (2002) Mapeo y localización simultáneos con escasa extensión filtros de información: teoría y resultados iniciales. Informe técnico CMU-CS-02-112, Carnegie Mellon Thrun S, Montemerlo M, Dahlkamp H et al (2005a) Stanley: el robot que ganó el gran desafío de DARPA. J Field Robot 23 (9): 661–692

Thrun S, Burgard W, Fox D, (2005b) Robótica probabilística. The MIT Press, Nueva York, ISBN: 0262201623 Thrun S, Montemerlo M, Aron A (2006) Análisis probabilístico del terreno para la conducción en el desierto a alta velocidad. En: Actas de robótica: ciencia y sistemas

Tirilly P, Claveau V, Gros P (2010) Distancias y esquemas de ponderación para la recuperación de imágenes de la bolsa de palabras visuales. En: Actas de la conferencia internacional sobre recuperación de información multimedia, págs. 323–333

Triggs B, Mclauchlan P, Hartley R, Fitzgibbon A (1999) Bundle Adjustment — a modern Synthesis. En: Procedidos del taller internacional sobre algoritmos de visión: teoría y práctica, págs. 298–375 Tuytelaars T, Mikolajczyk K (2008) Detectores locales de características invariantes: una encuesta. Gráfico de cálculo de tendencias encontradas Vis

Tuytelaars T, Van-Gool L (2004) Emparejar visiones muy separadas basadas en regiones invariantes afines. Int J Computación Vis 59 (1): 61–85

Vidal T, Bryson M, Sukkarieh S, et al (2007) Sobre la observabilidad de SLAM de solo rodamientos. En: Actas de la conferencia internacional IEEE sobre robótica y automatización, págs. 4114–4119

Vidal T, Berger C, Sola J, Lacroix S (2011) Mapeo visual de robots múltiples a gran escala con heterogeneidad hitos en terreno semiestructurado. Robot Autonom Syst, págs. 654–674

Wang C, Thorpe Ch, Thrun S et al (2007) Localización, mapeo y seguimiento de objetos en movimiento simultáneos. Int J Robot Res 26 (9): 889–916

Wangsiripitak S, Murray D (2009) Evitar el movimiento de valores atípicos en SLAM visual mediante el seguimiento de objetos en movimiento. En: Actas de la conferencia internacional IEEE sobre robótica y automatización, págs. 375–380

Williams B (2009) Localización y mapeo simultáneos usando una sola cámara. Doctorado, tesis, Universidad de Oxford Versity, Inglaterra

Williams B, Klein G, Reid I (2007) Reubicación de SLAM en tiempo real. En: Proceedings of the IEEE international conferencia sobre visión artificial

Williams B, Cummins M, Neira J, Newman P, Reid I, Tardós JD (2009) Una comparación de cierre de bucle técnicas en SLAM monocular. Robot Autonom Syst 57 (12): 1188-1197 Willson (1995) Software de calibración de cámara Tsai. http://www.cs.cmu.edu/~rgw/TsaiCode.html . Consultado el 06 de marzo de 2012

Yilmaz A, Javed O, Shah M (2006) Seguimiento de objetos: una encuesta. ACM Comput Surv 38 (4): 1–45

Zhang Z (2000) Una nueva técnica flexible para la calibración de la cámara. IEEE Trans Pattern Anal Mach Intell 22 (11): 1330-1334

Zhang W, Kosecka J (2006) Localización basada en imágenes en entornos urbanos. En: Actas del tercero simposio internacional sobre procesamiento, visualización y transmisión de datos en 3D

Zhang Z, Deriche R, Faugeras O, Luong Q (1994) Una técnica robusta para hacer coincidir dos imágenes sin calibrar mediante la recuperación de la geometría epipolar desconocida. J Artif Intell. Volumen especial sobre Computer Vision 78 (1): 87–119

Zhang H, Li B, Yang D (2010) Detección de fotogramas clave para SLAM visual basado en apariencia. En: Actas de la conferencia internacional IEEE sobre robots y sistemas inteligentes, págs. 2071–2076