Capítulo 1

Introducción

En este capítulo se introducirá de forma general el marco en el que se encuadra el presente trabajo de fin de máster (TFM). Concretamente se explicará que es la Visión Artificial y en particular el campo de la auto localización visual en entornos desconocidos mediante el uso del algoritmo de SLAM. Además se expondrá la estructura de este documento.

1.1. Visión Artificial

La Visión Artificial es una rama científica de la Inteligencia Artificial basada en la adquisición y extracción de características presentes en imágenes mediante el uso de computadores. Por tanto, los sistemas de visión artificial suelen estar compuestos por un sistema de adquisición de imágenes y un sistema capaz de procesarlas. La información obtenida permite, entre muchas otras, la detección y seguimiento de objetos, reconocimiento de caracteres (OCR), recreación de entornos tridimensionales a partir múltiples imágenes 2D o la toma de decisiones.

Los orígenes de este campo se remontan a la década de los 60, donde se conectó por primera vez una cámara a un computador con el fin de obtener información. Larry Roberts fue una de las primeras personas en desarrollar un experimento en el cual no solo se capturaban imágenes, sino que se analizaba su contenido (una estructura de bloques) para posteriormente reproducirlo desde otra perspectiva. Las técnicas de visión artificial estuvieron limitadas por la capacidad de cómputo de los ordenadores de la época. Sin embargo, en los últimos años, con el avance en este aspecto y la reducción en el coste del hardware, ha derivado en un gran avance de este campo. Esto ha permitido la aparición de algoritmos que trabajan en tiempo real y resuelven tareas tan complejas como la toma de decisiones críticas que dan lugar, por ejemplo, a la conducción autónoma de vehículos.

1.2. Visual SLAM

Localización y mapeado simultáneo o SLAM por sus siglas en inglés (*Simultaneous Localization And Mapping*) es el término empleado para describir el proceso por el cual, a partir de información obtenida de sensores, es posible generar un mapa del entorno que lo rodea al mismo tiempo que se localiza en él. SLAM tiene sus orígenes en el campo de la robótica siendo este uno de sus principales áreas de investigación, dónde cobra especial importancia en robots autónomos que operan en entornos desconocidos.

Visual SLAM hace referencia a sistemas SLAM cuyos sensores principales son una o más cámaras. Este tipo de sistemas emplean algoritmos de reconstrucción 3D, como la homografía, para realizar las tareas de construcción del mapa y el cálculo de la pose. Para ello, la mayoría de sistemas toman como entrada múltiples imágenes sucesivas de donde se extraen un conjunto puntos de interés para calcular la posición 3D de cada uno de ellos (generando el mapa) y simultáneamente dar una estimación de la pose (posición y orientación) de la cámara para cada una de las imágenes.

Pese a ser un algoritmo que aún está en desarrollo, dado que no hay una solución exacta para el problema, es empleado en un numerosas aplicaciones. Como por ejemplo:

- **Robot aspirador:** estos robots autónomos equipados con cámaras generan un mapa de los hogares o edificios para optimizar las rutas de limpieza al mismo tiempo que evitan obstáculos en su trayectoria.
- **UAV:** los vehículos aéreos no tripulados (UAV por sus siglas en inglés) como los drones pueden emplear el SLAM para visualizar el entorno que los rodea y tomar decisiones.
- **Realidad aumentada:** SLAM no solo tiene cabida en el mundo de la robótica, en el campo de la realidad aumentada también puede hacer uso de este algoritmo para relación el mundo real con el virtual.

1.2.1. Conceptos

Los algoritmos de localización como SLAM tienen asociados una serie de conceptos [1] que merece la pena detallar dado que se hará uso de ellos a lo largo del proyecto.

Calidad: la calidad del algoritmo dependerá de: la eficiencia temporal, la precisión espacial de la pose y la robustez.

Eficiencia temporal: medida como el tiempo de ejecución de cada iteración. Para considerar que el algoritmo es apto para trabajar en tiempo real deberá ser capaz de procesar al menos 30 fotogramas. (*frames*) por segundo.

Precisión de la posición: diferencia entre la pose estimada y la pose real, expresada como el error lineal y angular.

Robustez: entendida como la capacidad de recuperarse o seguir funcionando ante situaciones inesperadas, como oclusiones, imágenes borrosas, objetos dinámicos en la escena, etc.

Oclusiones: situación donde la cámara del sistema esté tapada total o parcialmente, de modo que no sea posible utilizar la parte de la imagen para obtener información.

Relocalización: capacidad para recuperarse de una pérdida por falta de información, siendo capaz de volver a estimar la posición de forma correcta dentro del mapa.

1.2.2. Problemas conocidos

Debido al coste computacional, falta de información y ciertas ambigüedades inherentes a los propios sensores de adquisición de imágenes, podemos encontrarnos una serie de problemas difíciles de abordar para Visual SLAM. Algunos de ellos son descritos a continuación:

1. **Inicialización:** Uno de los principales problemas consiste en la inicialización del mapa, debido a que no es posible conocer la distancia a la que se encuentran cada uno de los puntos de interés detectados con una simple observación. Es necesario emplear objetos 3D conocidos a modo referencia o bien tomar múltiples imágenes, desde distintos ángulos, de una misma zona para crear el mapa inicial. La calidad de este mapa inicial influirá en gran medida en el cálculo de los futuros puntos de interés.

2. **Ambigüedad de escala:** Relacionado con el primer problema, no es posible conocer la escala real a la cual se encuentran los objetos. Este es un problema inherente de los sistema monoculares, conocido como pérdida de información 3D al proyectar sobre un plano 2D. Es decir, no es trivial (para un computador) determinar si dos objetos que sobre el plano 2D tienen el mismo tamaño se encuentra a la misma distancia, o si uno de ellos, en el mundo real, es mucho más grande pero se encuentra a mayor distancia.

   Una manera de solventar el problema sería la incorporación de objetos 3D de medidas conocidas, como un tablero de ajedrez, aunque esta escala inicial comenzaría a variar en el tiempo. Otra forma de abordar este problema es añadir otros sensores que aporten información del mundo real de manera constante. Como por ejemplo el uso de sensores tipo LIDAR o cámaras RGBD para obtener estas distancias reales.

3. **Extracción de características:** Como se ha comentado anteriormente, para la creación del mapa que genera Visual SLAM se emplean puntos de interés. Estos puntos deben ser

lo suficientemente significativos como para distinguirlos y emparejarlos entre distintos fotogramas. Por norma general, las zonas que cumplen estos requisitos suelen ser esquinas, bordes de objetos o texturas relevantes.

El problema surge en el momento que la imagen proporciona pocos puntos de interés, o por el contrario, una cantidad muy similar de ellos (producto de una alta monotonía de texturas) que dificultaría la labor de emparejamiento. Otro problema son los objetos dinámicos en la escena. Estos suelen generar múltiples características relevantes que varían su posición en el espacio a lo largo del tiempo, induciendo a errores en el cálculo de la homografía.

A consecuencia de este problema, el sistema puede derivar en un estado en que pierde la capacidad de calcular la pose de la cámara y generación del mapa, ya que la información proporcionada por los sensores es insuficiente. Por ello es tan importante que los algoritmos de SLAM sean capaces de relocalizarse.

4. **Cierre de bucle:** Este problema hace referencia al momento en el cual la cámara vuelve, tras pasar un periodo de tiempo, a una zona del mundo que ya haya visitado con anterioridad. La dificultad de esta tarea radica en cambios en la escala, iluminación, cálculo de la pose de la cámara u otros motivos similares. Por este motivo, Visual SLAM debe ser capaz de reconocer este entorno conocido y modificar su mapa para hacerlo coincidir con esta situación.

1.3. Estructura de la memoria

La memoria del proyecto está estructurado en 6 secciones con el fin de abordar de manera sencilla todos los apartados del trabajo realizado.

La primera sección introduce las tecnologías y áreas de investigación que serán la base para el desarrollo del proyecto, concretamente, el campo de la Visión artificial y el algoritmo de Visual SLAM, explicando en qué consisten así como sus aplicaciones y principales problemas. La segunda sección establece los objetivos específicos del proyecto Inertial SD-SLAM. La tercera sección aporta una visión del Estado del Arte de los algoritmos de Visual SLAM que hacen uso de diferentes sensores a parte de cámaras para mejorar sus resultados. La sección cuarta aborda todo el desarrollo del proyecto. En este capítulo se exponen ...*TODO*... La quinta sección presenta los resultados obtenidos, incluyendo la metodología empleada para la obtención de los datos. En último lugar, se expondrán las conclusiones del trabajo tras haber analizado los resultados obtenidos.

Bibliografía

[1] E. P. García. Técnicas para la localización visual robusta de robots en tiempo real con y sin mapas.