Motion Estimation, 3D Reconstruction and Navigation with Range Sensors

Autor: Mariano Jaimez Tarifa

Directores: Javier González Jiménez
Daniel Cremers

Tesis doctoral
Doctorado en Ingeniería Mecatrónica
Escuela de Ingenierías Industriales
Universidad de Málaga

2017

UNIVERSIDAD
DE MÁLAGA

UNIVERSIDAD
DE MÁLAGA

AUTOR: Mariano Jaimez Tarifa

http://orcid.org/0000-0002-1885-2821

EDITA: Publicaciones y Divulgación Científica. Universidad de Málaga

BY NC ND

Esta obra está bajo una licencia de Creative Commons Reconocimiento-NoComercial-SinObraDerivada 4.0 Internacional:
http://creativecommons.org/licenses/by-nc-nd/4.0/legalcode
Cualquier parte de esta obra se puede reproducir sin autorización
pero con el reconocimiento y atribución de los autores.
No se puede hacer uso comercial de la obra y no se puede alterar, transformar o hacer obras derivadas.

Esta Tesis Doctoral está depositada en el Repositorio Institucional de la Universidad de Málaga (RIUMA): riuma.uma.es

UNIVERSIDAD
DE MÁLAGA

UNIVERSIDAD DE MÁLAGA

El Dr. D. Javier González Jiménez y el Dr. D. Daniel Cremers, directores de la tesis titulada ”Motion Estimation, 3D Reconstruction and Navigation with Range Sensors” realizada por D. Mariano Jaimez Tarifa, certifican su idoneidad para la obtención del título de Doctor en Ingeniería Mecatrónica.

Málaga, 12 de julio de 2017

Dr. D. Javier González Jiménez

Munich, 18 de julio de 2017

Daniel Cremers

Dr. D. Daniel Cremers

UNIVERSIDAD
DE MÁLAGA

UNIVERSIDAD
DE MÁLAGA

Contents

UNIVERSIDAD DE MÁLAGA

UNIVERSIDAD DE MÁLAGA

Acknowledgments and Retrospection

Many people have guided, encouraged and motivated me during this long academic journey. I take the opportunity here to thank them all, not only those who have helped me during my PhD but also those who played an important role in my education during my childhood and adolescence.

I am enormously thankful to Prof. Javier González Jiménez. He gave me the chance to join the MAPIR group and pursue a PhD degree on robotics five years ago. This turn up to be a great (and rather international) adventure full of professional and personal challenges. He taught me how to write a scientific paper (which involves undergoing the processes of "lija gorda" and "lija fina"), and pushed me to publish my work when I was skeptical about it. I have always had concerns about the current research system based on the "publish or perish principle" (and I still have them). He has helped me to keep my feet on the ground, and thanks to that a significant percentage of the papers supporting this thesis exist. However, his main strength is, in my opinion, his capacity to motivate us. He is always involved in our projects, being supportive during the hard times and excited about the good results. As a PhD student, this is very rewarding.

These years were full of joyful and often hilarious moments thanks to my colleagues and good friends in MAPIR. Raúl Sarmiento (*aka* Joselito) was always an endless source of entertainment and unpredictable thoughts. He has also provided me with technical support during all these years, fact for which he claims I owe him hundreds of euros (although I paid that debt back long ago). Francisco Meléndez was the best comrade and teammate for the many events and parties we organized in MAPIR at Christmas, Halloween, etc. Javier G. Monroy was my most reliable co-worker; he is committed, honest and good-hearted, and I appreciate him for that. Rubén Gómez was my companion of adventures. We went together to three ICRAs (Seattle, Stockholm and Singapore) and we visited each other during our respective stays in Munich and Zurich. I remember all of these trips with joy and I am happy to have shared those many good moments with him. Besides, there were uncountable enriching conversations about politics, science, education or even philosophy, and for those I thank Eduardo Fernández, Jesús Briales, Carlos Sánchez, Manuel López, Francisco A. Moreno, Andrés Góngora, Juan A. Fernández, Cipriano Galindo, Ana Cruz and Vicente Arévalo. Last, I must mention José Luis Blanco, a

UNIVERSIDAD DE MÁLAGA

former member of MAPIR. His work opened doors for many of us, MRPT was an invaluable programming tool for the work presented here and his detailed answers to my uncountable emails were highly appreciated.

In 2014 I enjoyed a research stay in the computer vision group at the Technical University of Munich led by Prof. Daniel Cremers. After that, Daniel offered me to pursue a dual PhD under his supervision, which was an incredible opportunity I could not refuse and I am very grateful for. As a supervisor, Daniel has taught me about optimization and the state-of-the-art in computer vision. He has been very supportive and flexible during these years with my changing plans and temporal visits to Munich. The computer vision group at TUM is a melting pot with very talented people from different nationalities and varied research interests. It is often visited by renowned researchers in the field, which is a strongly positive aspect and a great chance for us to grow. I acknowledge Daniel for all of this, and also for fostering a good atmosphere at work and for promoting out-of-work leisure activities like hiking, skiing, etc.

My life in Munich would have been sadder and boring without my colleagues and friends from the lab. Mohamed Souiai (*aka* Mo) taught me the basics of variational methods and also introduced me to many people and places in the city. Robert Maier was always willing to offer me the sofa of his flat when I ran into logistic problems (which happened more than once). Christian Kerl kindly lent me the teddy bear of his son to carry out my experiments (I am indebted to Moritz). Vladyslav Usenko was my most faithful companion for karaoke every week. My office mates, Christiane Sommer and Virginia Estellers, offered me endless and interesting discussions about the world, the people, politics and the right way to phrase ideas in English (on which we normally disagreed). Many others have contributed to turn my time in Munich into a pleasant and exciting experience, and I want to thank them for that: Lingni Ma, John Chiotellis, David Schubert, Laura Leal-Taixé, Jürgen Sturm, Jörg Stückler, Martin Oswald, Jakob Engel, Raluca Scona, Yvain Queau, Matthias Vestner, Zorah Lähner, Emanuel Laude, Rui Wang, Thomas Möllenhoff, Thomas Windheuser, Thomas Frerix, Benedikt Löwenhauser, Caner Hazirbas, Philip Häusser, Björn Häfner, Nikolaus Demmel, Tim Meinhardt, Vladimir Golkov, Tao Wu, Rudolph Triebel, Emanuele Rodola, Michael Möller, Csaba Domokos, Frank Schmidt, Sabine Wagner and Quirin Lohr.

In 2015, barely recovered from a back surgery, I spent two months working in Microsoft Research Cambridge under the supervision of Dr. Andrew Fitzgibbon. Honestly I regard Andrew as my third "unofficial" supervisor. Working with him was a thrilling and uplifting experience. I was lucky to spend hours with him developing and discussing different mathematical models, and I was very surprised by the fact he did not only had a vast theoretical background but also knew about implementation details, libraries, hardware, etc. Tom Cashman was another talented researcher I had the chance to work with in Cambridge. I appreciate much that he was always willing to give me a hand when I needed it, both during these months and afterwards. I thank both for those exciting months of work. Even though my health was far from good by that time, I remember it as a very fulfilling experience.

The doctoral degree is the final step of a long academic journey that also encompasses many years at school, high school and the university (bachelor and masters degrees). I believe that, at this point, it is fair to recall and thank those teachers that made an impact on me before I

UNIVERSIDAD DE MÁLAGA

started my PhD. I spent most of my childhood at the school Cerrado de Calderón, a place I remember with affection. There the first important teacher was Teresa Alba, who realized when I was 7 years old that I was a special pupil and proposed my parents to promote me to a higher course/class (one year). I barely remember her face but she changed my life with that decision, and I am grateful for it. A few years after, when I was 10 years old, I had a special tutor called Juan José Méndez. He always minded me and encouraged me to go further and learn more than the others if I had the potential to do so. Later at high school there were several influential figures. I remember Enrique Salinas, an exceptional sport teacher. José Luis Espejo taught me Spanish and its history; by that time I was unaware of how important one's language skills are but now I know it (in fact I notice that I automatically become a sillier person when I have to communicate or write in English. That is the price non-native English speakers pay to be part of this globalized world). Juan Carlos Rodriguez introduced me to the marvelous world of physics, and did it brilliantly. Nevertheless, the teacher who influenced me above all, and the one I admired the most, was Pedro Hormigo. He taught us math, and managed to do so in a very inspirational way (if I dig in my memories now after 12 years it feels almost like Harry Potter learning magic at Hogwarts). He is a once-in-a-lifetime teacher and for him is my more nostalgic thank you.

I started to get interested in research during the years of my bachelor. That interest was boosted by Prof. Juan A. Cabrera Carrillo. He offered me to participate in his projects at a very early stage of my studies, which led to subsequent years of collaboration, learning and research in the field of mechanical engineering. I was lucky to share these years with my closest friend (and also my strongest competitor by that time) Pablo Giner Abad. By working together we complemented each other but also pushed each other to the limit to keep up in our personal race. Together with Juan Castillo, the four of us formed a team always eager to address new challenges with imaginative solutions. I thank them all for those many enriching and also amusing moments.

It is time to talk about those who have been there since the day I was born: my family. My father showed me the value of hard working and willpower through sports. We enjoyed endless hours of football, tennis, ping pong, skiing and running together, and with them (and with him) I learned that training and enduring normally comes with a reward. As a pragmatic person, he has questioned the purpose of my work during these years, and I have to concede here that I have also sometimes found it pointless. During my childhood, my mother spent hundreds of hours teaching me and asking me the lessons I should memorize while I was running around our flat with my scooter or my football (or both). By that time she was concerned that I could become a lone, asocial and ruthless person, and she did her best to soften me in order to avoid that. Knowing myself, I can say that she was successful, and I thank her for that. My grandparents, the four of them, have lived exemplary lives even though their circumstances were at times dramatic. They have been references to look up to since I was a child. To finish I simply have to say that I have the best possible relatives in many many senses: my sisters Alejandra and Iria, my uncles and aunts, my cousins, and many others whose affection and tenderness make Málaga and Loja my true homes. But I must confess there is room for improvement in one particular aspect: no matter how many times I explain the topic I work on, they all always forget it almost instantly

UNIVERSIDAD DE MÁLAGA

(and demand me to explain it again and again and again). I assume I am not the only PhD student suffering this though...

And there is one last and very special person who truly deserves my gratitude: my partner Julia Nagel. She has experienced more than any other the consequences of my PhD during these years. Sometimes these consequences were positive, e.g. when my internship and my dual PhD brought me closer to her (she lives in Germany). Other times, during periods of high workload and deadlines, she had to cope with a less social and often distant me. Our particular issue during these years was travelling: she was always keen on travelling often, to far destinations and for long, and I was always busy with some project. This led to frequent conversations about work-life balance, and I admit now she was right when she said I should schedule time for myself and be able to disconnect from work. I have followed that advice lately. Most importantly, she has provided me with the peace and personal support I need to feel balanced and be able to concentrate on my work. She has also blindly believed in me and celebrated each of my achievements as hers. Moreover, I highly value that she has happily accepted my mother tongue as "our language". For these and many other reasons this thesis would not exist without her, and for her are the only Spanish words here: muchas gracias por todo.

Mariano Jaimez Tarifa
Munich
June 2017

I want to thank the Government of Spain (grant program FPI-MICINN 2012, project *Taroth* DPI2011-25483, project *Promove* DPI2014-55826-R) and the European Union (project *GiraffPlus*, ERC grant *Convex Vision*, ERC consolidator grant *3D Reloaded*) for funding these years of research, internships and attendances to international conferences. I also gratefully acknowledge the support of NVIDIA Corportation with the donation of hardware used for this research.

The Government of Spain has invested a significant sum of money in this thesis during a period when the Spanish economy was in crisis. Unfortunately this investment is not going to be profitable for the Spanish State because the lack of a powerful technological industry in our country pushes me to seek for jobs abroad. I sincerely feel this is a pitty and a detrimental situation for Spain but it is not in my hands to change it. The Spanish government should promote a much tighter collaboration between companies and researchers in order to avoid this. How to do that in an efficient way is beyond my knowledge, but it might be worth thinking about all the work developed by hundreds of PhD students which falls into oblivion after they finish. I believe that huge amount of effort should be exploitable.

UNIVERSIDAD
DE MÁLAGA

Resumen

Introducción

Desde su orígenes, la robótica ha perseguido el desarrollo de máquinas que sean capaces de desempeñar tareas de forma autónoma, con el objetivo de aumentar la eficiencia de ciertos procesos, disminuir los costes o liberar a las personas de trabajos/tareas tediosas o peligrosas. Dada su complejidad, distintas disciplinas de la ingeniería y la física han convergido en lo que hoy denominamos "mecatrónica" para abordar la conceptión, el diseño y la fabricación de estos sistemas complejos (los robots). De entre los muchos retos que todavía quedan por resolver, uno de los más importantes, si no el que más, es el de dotar de autonomía o inteligencia a los robots. Para poder sustituir a una persona, un robot debe ser capaz de conocer su entorno, interpretar sus circunstancias y decidir cómo actuar en cada caso para lograr su objetivo. Así, la autonomía está directamente relacionada con la capacidad de percepción del robot y la forma en la que procesa la información que percibe. Al igual que los humanos, los robots están dotados de sensores (sus "sentidos") que les permiten ver, oír, palpar e incluso oler. Y al igual que para los humanos, el sentido más poderoso es el de la vista o, como comúnmente se denomina en el argot científico, la visión. Sin embargo, los robots no sólo cuentan con sistemas de visión pasiva (aquellos que observan el entorno tal cual es sin alterarlo, iluminarlo o "irradiarlo") sino que también pueden incorporar sistemas de visión activa. En el ámbito de los sensores, la visión activa consiste en la emisión de algún tipo de patrón de iluminación (o radiación) que se refleja en el entorno y es posteriormente detectado por un determinado sensor receptor. Gracias a ello, los sistemas de visión activa son capaces de captar la geometría de los objetos que observan, lo cual resulta de enorme utilidad en muchos ámbitos tecnológicos. En robótica, los sistemas de visión activa, también denominados "sensores de rango", se utilizan comúnmente para:

- Conocer la distribución de los obstáculos que rodean a un robot.
- Conocer la posición y orientación de objetos que se van a manipular.
- Construir mapas 2D o 3D del entorno en el que opera un robot móvil.
- Estimar el movimiento 2D o 3D de un robot para conocer su posición en cada instante.
- Segmentar la escena observada en los distintos objetos que la componen.

UNIVERSIDAD
DE MÁLAGA

Además, los sensores de rango encuentran un sinfín de aplicaciones más allá de la robótica, como por ejemplo en:

- La interacción con ordenadores o videoconsolas.
- El modelado 3D de objetos.
- El análisis de movimiento, tanto deportivo como terapéutico.
- La estimación de movimiento para aplicaciones de realidad virtual/aumentada.
- Sistemas de seguridad y ayuda en la conducción de vehículos.

Sin embargo, todas estas aplicaciones requieren algoritmos complejos capaces de procesar las medidas geométricas de los sensores de forma precisa y rápida. Hasta la fecha se han propuesto numerosas soluciones particulares para cada problema, pero estas soluciones frecuentemente se ajustan a casuísticas concretas y entornos controlados, o bien adolecen de practicidad y presentan resultados satisfactorios solo desde un punto de vista teórico. En esta tesis se proponen algoritmos nuevos para la resolución de muchos de los ejemplos mencionados anteriormente. En general, el objetivo ha sido el encontrar soluciones con base teórica sólida que a su vez sean directamente aplicables al problema considerado, no solo en teoría o en simulación sino también en la práctica. Ello implica no solo preocuparse por la formulación sino también por la implementación, y en muchos casos ambos aspectos deben ser considerados a la vez para escoger una formulación precisa que pueda ser implementada de forma eficiente. Siguiendo este mismo espíritu utilitarista, el código asociado a los trabajos que aquí presentamos es público y puede ser utilizado por la comunidad científica.

Motivación

Los sensores de rango han estado presentes desde principios del siglo XX pero han evolucionado mucho en las últimas décadas. Tras la aparición del sónar durante la I Guerra Mundial, los primeros sensores de rango basados en emisión de luz fueron los escáneres láser (o *lidars*) a principios de la década de 1960. La precisión de estos sensores era muy alta al igual que su precio, y durante años fueron utilizados solo en aplicaciones militares o espaciales. Con el tiempo los costes se redujeron y las versiones más simples (escáneres 2D) empezaron a equiparse en robots móviles para la detección de obstáculos, la localización y el mapeado 2D. Sin embargo, la revolución llegó con la aparición de la cámara Kinect de Microsoft [1] a finales de 2010. La Kinect fue la primera cámara de bajo coste (150 euros) que proporcionaba no solo imágenes de color sino también de profundidad, y a una frecuencia razonablemente alta (30 Hz). Así, fue el primer sensor de bajo coste que permitía "conocer" la geometría del entorno con precisión, y su impacto en la comunidad robótica y en otros muchos campos fue, y aún sigue siendo, enorme. No obstante, el sensor Kinect también suponía un reto puesto que la cantidad de datos a procesar era muy alta (imágenes de color y profundidad con resolución VGA a 30 Hz), y un porcentaje significativo de trabajos publicados en robótica y en visión por computador en los últimos años tratan sobre su uso y aplicación con distintos fines.

UNIVERSIDAD DE MÁLAGA

Por otra parte, la mejora incesante de los procesadores ha permitido abordar en los últimos años problemas que hasta hace poco eran computacionalmente intratables. Las CPU's actuales contienen múltiples unidades de procesamiento (núcleos) con cachés más grandes. También incluyen registros especiales y conjuntos de instrucciones para realizar operaciones sobre datos vectorizados de forma mucho más eficiente (SSE, AVX). Las GPU's han dejado de ser meras unidades de procesamiento de gráficos (como su propio nombre indica) para convertirse en unidades de procesamiento masivo de datos. Han incrementado enormemente su potencia de cómputo, su memoria y su ancho de banda a la vez que han reducido su consumo energético, ampliando así su rango de aplicación. Además, los compiladores modernos (por ejemplo MSVS, GCC o Intel para C++) generan código cada vez más optimizado sin que ello requiera un esfuerzo extra por parte del programador. Las nuevas versiones de CUDA incorporan la "memoria unificada" para permitir al programador acceder a direcciones de memoria de la CPU y la GPU indistintamente. Muchas otras librerías incluyen funcionalidades para explotar el paralelismo del hardware moderno, ya sea mediante vectorización, programación multi-thread o implementación en GPU: STL (C++11), OpenCL, Eigen, OpenGL, etc.

Estos son los dos aspectos principales que motivan el trabajo presentado en esta tesis. De forma resumida, el reto que afrontamos es el de utilizar los sensores de rango modernos y la potencia de los ordenadores actuales para resolver algunos problemas fundamentales de robótica o de visión que siguen sin ser resueltos (o que han sido solo parcialmente resueltos).

Objetivos

A pesar de los grandes avances conseguidos en visión por computador en las últimas décadas, todavía existen muchos problemas abiertos. Entre ellos, los que resultan más relevantes para la robótica normalmente requieren un cierto conocimiento de la geometría del entorno observado, y por tanto pueden abordarse mediante el uso de sensores de rango. En esta tesis nos centramos principalmente (pero no exclusivamente) en la estimación de movimiento 2D y 3D mediante distintos tipos de sensores de rango. En algunos casos el objetivo es la estimación del movimiento del propio sensor, y en otros la del movimiento de los objetos que observa. Además, presentamos nuevos algoritmos que explotan los datos geométricos de cámaras o escáneres láser para la reconstrucción 3D de objetos o la navegación autónoma de robots.

Por tanto esta tesis no versa sobre un único tema sino que aborda varios problemas de distinta índole. El objetivo común es proponer soluciones a dichos problemas que sean más rápidas o más precisas que las existentes, o bien que permitan a un robot o dispositivo operar en condiciones más extremas (por ejemplo en la oscuridad o en entornos muy dinámicos). A continuación explicamos en más detalle los temas abordados, sus posibles aplicaciones prácticas y las dificultades inherentes a cada uno de ellos:

- **Odometría basada en sensores de rango**. La odometría consiste en la estimación incremental del movimiento mediante uno o varios sensores. Es una parte fundamental de muchos sistemas robóticos o de realidad virtual/aumentada. En robótica, por ejemplo, permite un conocimiento preciso y continuo de la posición de un robot, lo cual es vital para que éste pueda llevar a cabo tareas de forma autónoma. En el ámbito de la realidad virtual o aumentada, conocer el

UNIVERSIDAD DE MÁLAGA

movimiento del dispositivo que genera las imágenes virtuales (y por tanto, el punto de vista del usuario que las visualiza) es fundamental para renderizar dichas imágenes con la perspectiva adecuada.

Nuestro objetivo es crear nuevos algoritmos de odometría basados en sensores de rango. Teniendo en cuenta los casos de aplicación descritos, es esencial que estos algoritmos funcionen en tiempo real y de forma precisa incluso en entornos muy dinámicos.

- Estimación del **flujo de escena**. Se denomina flujo de escena al campo vectorial que representa de forma independiente el movimiento 3D (o velocidad) de todos y cada uno de los puntos observados por una cámara. Sus aplicaciones son múltiples: el modelado 3D de cuerpos no rígidos, la manipulación de objetos móviles, la interacción hombre-máquina o el análisis de movimiento son algunos ejemplos.

  El estado del arte en este ámbito no ha alcanzado todavía un nivel de madurez equiparable al de la odometría o la estimación de flujo óptico. Por esta razón, nuestro objetivo es superar o paliar las principales limitationes que actualmente impiden el uso del flujo de escena en casos reales. La más importante es su alto coste computacional, y en ello centramos nuestros esfuerzos.

- **Modelado 3D** de objetos. Consiste en obtener una representación matemática (en este caso una superficie) que describa la forma de un objeto dado a partir de un conjunto de imágenes. Esto resulta útil a la hora de medir dimensiones, insertar objetos reales en mundos virtuales, o combinar objetos y entornos reales que no están juntos físicamente (por ejemplo, utilizando un modelo virtual de un mueble para ver cómo quedaría en una habitación).

  Más concretamente, nuestro objetivo es explotar la "información de fondo" en imágenes de profundad (aquellas partes de la imagen que no observan el objeto a modelar sino los alrededores o el fondo) para guiar el proceso de reconstrucción 3D.

- **Navegación reactiva** con sensores de rango. Como su propio nombre indica, los algoritmos de navegación reactiva permiten a un robot móvil "reaccionar" ante distintas distribuciones de obstáculos para alcanzar un destino prefijado. Al contrario que los sistemas deliberativos, no necesitan conocer previamente el mapa del entorno, pero solo permiten la navegación a destinos locales (no muy lejanos a la pose del robot).

  La mayoría de las soluciones existentes simplifican el problema de navigación representando en 2D tanto al robot como a los obstáculos que lo rodean. Nuestro objetivo es superar dicha simplificación y considerar la distribución tridimensional de obstáculos y la forma 3D del robot para obtener algoritmos de navegación más precisos.

Contribuciones

En este apartado se enumeran las contribuciones que se presentan en esta tesis. Se incluye un breve resumen de cada una y se indican los artículos en las que han sido publicadas. Tanto el código como los vídeos asociados a dichas publicaciones pueden encontrarse en:

`http://mapir.isa.uma.es/mjaimez`

UNIVERSIDAD DE MÁLAGA

Odometría visual 3D a partir de imágenes de profundidad

Presentamos un método que alinea imágenes de profundidad consecutivas para estimar el movimiento de la cámara. Demostramos que este método es mucho más rápido y más preciso que otros métodos que utilizan, al igual que el nuestro, solo información geométrica para la estimación [2]. Por otro lado, demostramos que nuestro método es tan preciso como aquellos del estado del arte que utilizan información tanto geométrica como fotométrica (color) en el proceso de estimación [3]. Además, funciona en tiempo real (30Hz o más) ejecutándose en un único núcleo de la CPU.

- M. Jaimez and J. Gonzalez-Jimenez, “Fast Visual Odometry for 3D Range Sensors”, *IEEE Transactions on Robotics*, vol. 31, no. 4, pp. 809 - 822, 2015.

Odometría visual 2D con escáneres láser

Desarrollamos un método para alinear barridos láser consecutivos y así poder estimar el movimiento plano de robots. Dicho método se basa en la ecuación de flujo de rango, aplicada por primera vez para escáneres láser en [4]. Demostramos que nuestro método es más eficiente computacionalmente y más preciso que algunos de los métodos de referencia [5, 6]. Con un tiempo de ejecución de apenas 1 milisegundo, resulta ideal para todas aquellas aplicaciones robóticas que consuman muchos recursos computacionales.

- M. Jaimez, J. G. Monroy, J. Gonzalez-Jimenez, “Planar Odometry from a Radial Laser Scanner. A Range Flow-based Approach”, *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 4479-4485, 2016.

Posteriormente extendimos dicha formulación para obtener resultados más precisos a partir de una representación simétrica del flujo de rango y del alineamiento de múltiples barridos láser. Este método mezcla alineamiento de barridos consecutivos con alineamiento frente a "barrido de referencia" (o *keyscan*), aunando las ventajas de ambas estrategias. También se propone una nueva técnica para la elección de los *keyscans* basada en modelar el error del método en función de la traslación y rotación existente entre los barridos alineados. Así, se puede establecer el nivel de error deseado y a partir de él decidir cuándo añadir nuevos *keyscans* durante la estimación. Además, presentamos una sección de resultados mucho más extensa con comparativas cuantitativas y cualitativas tanto en simulación como con datos reales.

- M. Jaimez, J.G. Monroy, M. Lopez-Antequera and J. Gonzalez-Jimenez, “Robust Planar Odometry Based on Symmetric Range Flow and Multi-Scan Alignment”, *IEEE Transactions on Robotics*. **Under Review**.

Estimación del flujo de escena en tiempo real con cámaras RGB-D

Desarrollamos el primer método capaz de estimar el flujo de escena en tiempo real con cámaras RGB-D. Utilizamos el algoritmo Primal-Dual [7, 8] para resolver el problema ya que es fácilmente paralelizable y por tanto idóneo para una implementación GPU. Con ello conseguimos

UNIVERSIDAD
DE MÁLAGA

tiempos de cómputo 2 o 3 órdenes de magnitud más bajos que los métodos existentes, los cuales requieren de media varios segundos (o incluso minutos) para ejecutarse.

- M. Jaimez, M. Souiai, J. Gonzalez-Jimenez and D. Cremers, "A Primal-Dual Framework for Real-Time Dense RGB-D Scene Flow", *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 98-104, 2015.

Estimación conjunta del flujo y de la segmentación de una escena con cámaras RGB-D

Inspirándonos en [9], presentamos un algoritmo para estimar los distintos sólidos rígidos que componen una escena y el movimiento asociado a cada uno de ellos. La contribución principal de este artículo consistió en utilizar una segmentación suave (no-binaria) que permitiese estimar adecuadamente el movimiento de aquellas zonas u objetos de la escena que no fuesen rígidas. Para ello, el movimiento de cada punto observado por la cámara (flujo de escena) se calcula por interpolación lineal de las velocidades/transformaciones asociadas a cada sólido rígido.

- M. Jaimez, M. Souiai, J. Stückler, J. Gonzalez-Jimenez, D. Cremers, "Motion Cooperation: Smooth Piece-Wise Rigid Scene Flow from RGB-D Images", *International Conference on 3D Vision (3DV)*, pp. 64-72, 2015.

Estimación conjunta del flujo de escena y de la odometría visual con cámaras RGB-D

Este es un problema de gran relevancia y complejidad, puesto que un gran número de aplicaciones necesitan conocer tanto el movimiento de la cámara/sensor considerado como el de los objetos que observa. La dificultad estriba en el hecho de que, cuando la cámara se mueve, todo está en "movimiento aparente" respecto a ella, y por tanto es muy difícil distinguir entre los cambios en la imagen causados por el movimiento de la cámara y aquellos causados por el movimiento propio de los objetos de la escena. En este trabajo describimos una estrategia para segmentar la imagen en partes fijas y partes móviles. Una vez hecho esto, las partes fijas se utilizan para hacer odometría visual y para las partes móviles se estima el flujo de escena. Además, todo esto se hace a una frecuencia de 10 Hz, lo que permite que sea aplicable en casos reales.

- M. Jaimez, C. Kerl, J. Gonzalez-Jimenez and D. Cremers, "Fast Odometry and Scene Flow from RGB-D Cameras based on Geometric Clustering", *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3992-3999, 2017.

Formulación de un nuevo término de fondo para el modelado y el seguimiento (tracking) de objetos a partir de imágenes de profundidad

Frecuentemente los modelos 3D de objetos se construyen a partir de conjuntos o secuencias temporales de imágenes. En dichas imágenes el objeto es observado desde distintas perspectivas, pero también son visibles el resto de objetos que los rodean, comúnmente considerados como "fondo". En este trabajo describimos una formulación basada en superficies de subdivisión para el modelado, y presentamos una nueva estrategia para imponer que el modelo que se estima no sea visible desde aquellos píxeles que observan el fondo y no el objeto a modelar. Esta estrategia

UNIVERSIDAD DE MÁLAGA

permite que el algoritmo converja más fácilmente a la forma real del objeto, ya que mantiene el modelo dentro de las siluetas del objeto (una por imagen) e impide que crezca o se expanda de forma indeseada.

- M. Jaimez, T. Cashman, A. Fitzgibbon, J. Gonzalez-Jimenez, D. Cremers, "An Efficient Background Term for 3D Reconstruction and Tracking with Smooth Surface Models", *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7177-7185, 2017.

Navegación reactiva basada en la determinación exacta de colisiones en 3D

Este trabajo generaliza el algoritmo presentado en [10] al "mundo 3D", considerando tanto la forma real del robot (en vez de su proyección en planta) como la distribución espacial de los obstáculos que lo rodean. El método resultante funciona con cualquier combinación de sensores de rango, ya sean escáneres 2D, escáneres 3D o cámaras RGB-D, y es capaz de generar comandos de movimiento a una frecuencia de 200 Hz. Además, fue testeado durante casi 20 kilómetros de navegación con distintos robots demonstrando su eficacia tanto en hogares como en entornos de oficina/laboratorio.

- M. Jaimez, J. L. Blanco and J. Gonzalez-Jimenez, "Efficient Reactive Navigation with Exact Collision Determination for 3D Robot Shapes", *International Journal of Advanced Robotic Systems*, vol. 12, no. 63, 2015.

Marco y evolución de la tesis

Mi trabajo de investigación comienza a mediados del 2012 en el grupo MAPIR (MAchine Perception and Intelligent Robotics)[1], dentro del Departamento de Ingeniería de Sistemas y Automática de la Universidad de Málaga. Este grupo tiene amplia experiencia investigadora en robótica móvil, robótica olfativa y visión, y desde el comienzo de mi tesis yo me he situado a caballo entre la visión y la robótica, o mejor dicho, he desarrollado algoritmos de visión que son directamente aplicables en robótica. Dado el boom de Kinect tras su reciente aparición en 2010, decidimos orientar la tesis a explorar las ventajas que este tipo de cámaras ofrecían frente a las cámaras RGB tradicionales.

Inicialmente dedicamos más de medio año a explotar la información geométrica que proporcionan las cámaras de profundidad para la navegación autónoma de robots. Tras desarrollar un algoritmo de navegación eficiente y efectivo, decidimos abordar el problema de la odometría visual, a la que hemos dedicado gran parte del tiempo y esfuerzo de esta tesis. En este contexto, tuve la oportunidad de realizar una estancia de 4 meses en el grupo de visión[2] de la Universidad Técnica de Munich, liderado por el catedrático Daniel Cremers. Ello me permitió aprender sobre optimización y cálculo variacional, conocer mejor el estado del arte y explorar una nueva temática que acabó convirtiéndose en el segundo pilar de mi tesis: la estimación del flujo de escena. Además, la colaboración resultó positiva para ambas partes y desde entonces soy también miembro de dicho grupo y realizo mi doctorado en régimen de cotutela.

[1] http://mapir.isa.uma.es
[2] http://vision.in.tum.de

UNIVERSIDAD DE MÁLAGA

Posteriormente, tuve la suerte de hacer otra estancia en el centro de investigación de Microsoft en Cambridge (Reino Unido), bajo la supervisión del investigador principal Andrew Fitzgibbon. Esta estancia resultaba idónea puesto que Microsoft fue quien desarrolló el sensor Kinect, en el que se basa gran parte de mi tesis, y también un sinfín de trabajos de investigación y aplicaciones relacionadas. Por otra parte, mis meses allí coincidieron con el lanzamiento de las Hololens [11] (equipadas con 2 cámaras de profundidad), y por tanto estuve en contacto con algunos de los proyectos que en el futuro (o quizás ya en la actualidad) permitirán el uso y la interacción con dicho dispositivo.

Gracias a mi actividad investigadora he podido conocer y compartir conversaciones con algunas de las figuras más importantes a nivel mundial en visión y en robótica, lo cual agradezco enormemente. Además de las estancias ya mencionadas, he asistido a congresos internacionales en Seattle (Estados Unidos), Lyon (Francia), Estocolmo (Suecia) y próximamente Singapur. También he tenido la oportunidad en estos últimos meses de impartir, conjuntamente con mi director Javier González Jiménez, la asignatura de "Visión por Computador" en la E.T.S.I. Informática de la Universidad de Málaga. En resumen, han sido unos años vibrantes, de constante cambio y aprendizaje, de ilusión y de estrés, de derrotas y de victorias.

Estructura de la tesis

Dado que esta tesis se realiza en régimen de cotutela entre la Universidad de Málaga y la Universidad Técnica de Munich, ha sido redactada en inglés. Para cumplir con la reglamentación de la Universidad de Málaga, y también para obtener el doctorado con mención internacional, se incluye un amplio resumen de la misma escrito en español.
El bloque principal (el redactado en inglés) se divide en los siguientes capítulos:

- **Capítulo 1: Introducción.** Se introduce la temática de la tesis, se describe el contexto, se presentan las contribuciones y se detalla la estructura de la misma.
- **Capítulo 2: Sensores de rango.** Se presenta el concepto de "sensor de rango". Se enumeran los distintos tipos de sensores de rango existentes y se explican sus principios de funcionamiento. Se ilustran sus ventajas y desventajas respecto a otros tipos de sensores comúnmmente utilizados en robótica o en visión, y se describen sus principales aplicaciones.
- **Capítulo 3: Odometría con sensores de rango.** Se introduce el concepto de odometría, centrándonos en la odometría visual. Se analiza el estado del arte y las distintas estrategias existentes. Posteriormente se presentan los métodos desarrollados para la estimación del movimiento de sensores de rango, tanto en 3D para cámaras de profundidad como en 2D para escáneres láser.
- **Capítulo 4: Estimación del flujo de escena con cámaras RGB-D.** Se presenta el concepto de flujo de escena como extensión del flujo óptico al mundo 3D. Se incluyen dos contribuciones: la estimación del flujo de escena en tiempo real, y la estimación conjunta de la segmentación y del flujo de escena basada en una partición suave del entorno en distintos sólidos rígidos. Además, en la última sección se aborda un problema ambicioso que no ha sido muy explorado

UNIVERSIDAD DE MÁLAGA

hasta la fecha: la estimación conjunta del movimiento de una cámara RGB-D y del movimiento de los objetos que dicha cámara observa.

- **Capítulo 5: Modelado y seguimiento de objetos a partir de imágenes de profundidad.** Dado que la literatura en este tema es muy extensa, ofreceremos una perspectiva general resumida del estado del arte para luego centrarnos en analizar las distintas estrategias que existen para explotar la información de los píxeles que ven el "fondo" a la hora de modelar o seguir un objeto. Presentaremos un nuevo "término de fondo" para explotar dicha información y demostraremos sus ventajas al hacer modelado o seguimiento de objetos con superficies de subdivisión.
- **Capítulo 6: Navegación reactiva de robots móviles con sensores de rango.** En este capítulo describimos el concepto de navegación reactiva, la cual, junto al planificador, gobierna el movimiento autónomo de un robot. Explicamos los distintos algoritmos de navegación reactiva existentes, normalmente basados en proyecciones 2D de los obstáculos y de la forma del robot. Proponemos superar esa simplificación y considerar la forma tridimensional del robot y la distribución espacial real de los obstáculos, detectados con distintos tipos de sensores de rango.
- **Capítulo 7: Conclusiones.** Terminamos la tesis resumiendo el trabajo realizado y el impacto de las contribuciones que se presentan. Miramos también hacia el futuro y analizamos tanto las tecnologías venideras como algunos de los grandes problemas todavía sin resolver.

Conclusiones

En esta tesis se han abordado distintos problemas relacionados con la estimación del movimiento, el modelado de objetos y la navegación de robots. Todos ellos han tenido como denominador común el uso de sensores de rango para su resolución. Este tipo de sensores han demostrado ser una alternativa eficaz a las tradicionales cámaras RGB monoculares o estéreo. Conocer la geometría del entorno es fundamental en muchos problemas de visión, y en dichos casos resulta ventajoso el uso sensores de rango (las cámaras RGB necesitan estimar la profundidad de la escena observada, consumiendo recursos computacionales en el proceso). Además, los sensores de rango pueden funcionar en condiciones de iluminación cambiantes o incluso en total oscuridad, lo cual amplía su rango de aplicación. Como contrapartida, los sensores de rango solo pueden detectar objetos por debajo de una cierta distancia umbral y tienden a verse afectados por la radiación solar, lo cual restringe a veces su uso en exteriores.

A continuación presentamos las conclusiones individuales de cada trabajo, destacando sus pros y contras y las posibles líneas de actuación futuras.

1. En odometría visual, hemos demostrado que a partir de imágenes de profundidad se pueden obtener estimaciones precisas y rápidas del movimiento 3D de una cámara. La principal limitación de nuestro método es que no es robusto frente a objetos móviles, lo cual puede solventarse incluyendo estimadores-M de los residuos geométricos o extendiendo la formulación existente a un sistema *multiframe*. Desde un punto de vista de su aplicación, este método podría ser usado para estimar el movimiento de dispositivos de realidad virtual/aumentada

UNIVERSIDAD DE MÁLAGA

equipados con cámaras de rango (Hololens), o de forma más general, como "front-end" de sistemas de mapeado y SLAM.

2. Por otra parte, hemos aplicado las estrategias de alineamiento denso desarrolladas en los últimos años en odometría 3D a la odometría plana con escáneres láser, obteniendo resultados más precisos y rápidos que los métodos existentes. La principal limitación de nuestro método es que necesita que los barridos láser observados sean diferenciables al menos "a trozos", lo cual limita su usabilidad en entornos de exteriores compuestos principalmente por árboles, plantas, objetos dispersos, etc. Sin embargo, esta aparente limitación no lo es tanto puesto que el movimiento en exteriores rara vez es plano, y por tanto la odometría visual con escáneres láser 2D deja de ser aplicable. Dadas sus características, este método es ideal para estimar el movimiento de robots de servicios o de telepresencia que operen en hogares, entornos de oficinas, museos u hoteles.

3. Hemos presentado la primera implementación en tiempo real para la estimación del flujo de escena con cámaras RGB-D. Así, permitimos que todos aquellos robots o sistemas equipados con cámaras RGB-D (y con tarjeta gráfica NVIDIA) puedan utilizar el flujo de escena en aplicaciones que requieran respuesta online. Las principales limitaciones de este método son que solo es capaz de estimar movimientos pequeños (lo cual no es problema funcionando en tiempo real) y que no considera las oclusiones. En base a lo comentado, puede resultar un método ideal para la interacción entre personas y ordenadores/videoconsolas, o también para facilitar el funcionamiento de robots manipuladores en entornos dinámicos con objetos móviles.

4. Un enfoque totalmente opuesto se describe en el artículo "Motion Cooperation: Smooth Piece-Wise Rigid Scene Flow from RGB-D Images". En este caso se presentó un algoritmo que obtiene resultados muy precisos al estimar conjuntamente los distintos segmentos rígidos que componen la escena y sus velocidades asociadas. Como contrapartida, su carga computacional es mucho más alta: 30 segundos por imagen utilizando la CPU y la GPU, lo cual impide su aplicación directa en muchos casos prácticos. En este trabajo también demostramos que la elección de una segmentación "suave" puede ser beneficiosa frente a la técnicas de segmentación binaria tradicionales cuando los objetos observados no sean totalmente rígidos, es decir, cuando haya personas, animales, juguetes u otros cuerpos flexibles en la escena. Dada la precisión de este método, y el hecho de que representa el movimiento de cada punto no como un vector sino como una transformación rígida, puede ser de utilidad para renderizar imágenes "virtuales" que emulen lo que la cámara vería durante el intervalo transcurrido entre dos imágenes RGB-D consecutivas. Si en vez de considerar solo 2 imágenes consideramos una secuencia completa, entonces podríamos generar una versión "slow motion" de dicha secuencia original.

5. Hemos abordamos la estimación conjunta de la odometría y del flujo de escena con el objetivo de derivar un método fiable pero que también fuese aplicable en la práctica. Para ello utilizamos un esquema doble de segmentación: por un lado segmentamos la escena en partes rígidas y partes móviles, y por otro la dividimos en pequeños bloques que consideramos como

UNIVERSIDAD DE MÁLAGA

sólidos rígidos para aliviar la carga computacional. Como resultado obtuvimos un método que es robusto frente a objetos móviles a la hora de estimar el movimiento de la cámara, y que además calcula el flujo de escena varios órdenes de magnitud más rápido que la mayoría de métodos existentes. Además, hasta donde sabemos, es el único método capaz de hacer ambas cosas con precisión y a una frecuencia suficientemente alta como para que pueda aplicarse en casos reales. Este método sería de utilidad en un gran número de aplicaciones en las que el dispositivo en cuestión (ya sea un robot móvil, unas gafas de realidad virtual, u otros) funcione en entornos dinámicos y cambiantes.

6. También hemos formulado el problema del modelado 3D y el seguimiento (o *tracking*) de objetos a partir de imágenes de profundidad. La formulación de dichos problemas casi siempre viene expresada como un proceso de optimización, y nuestra contribución principal consistió en añadir un nuevo "término de fondo" para que todos aquellos píxeles que no observan el objeto a modelar contribuyan a que el modelo estimado adquiera una silueta similar a la del objeto real. Los resultados demuestran las múltiples ventajas de dicha formulación frente a la estrategia tradicional basada en la transformada de la distancia. Sin embargo, el sistema global de reconstrución y tracking no es suficientemente maduro para competir con los trabajos del estado del arte. Esto se debe, en parte, a que la topología del objeto a modelar es desconocida y el proceso de reconstrucción 3D siempre parte de una simple esfera que engloba los datos. Dicha esfera se va adaptando y refinando durante la optimización pero su topología nunca cambia, por lo que la reconstrucción se vuelve extremadamente compleja. Como posible solución se podría formular un problema de optimización continua-discreta que alterne deformaciones de la malla con modificaciones en su topología. Esta y otras alternativas deben explorarse para mejorar la calidad de las reconstrucciones obtenidas.

7. Por último, hemos extendido el algoritmo de navegación reactiva presentado en [10] para que tuviera en cuenta la forma real del robot y la distribución real de los obstáculos que lo rodean. Al contrario que la mayoría de los sistemas existentes, que simplifican el problema de navegación al caso 2D, nosotros propusimos un sistema de cálculo de colisiones exacto que no sobrelimita el movimiento del robot. El sistema fue testeado en tres plataformas robóticas diferentes y, en estos últimos años, ha permitido a multitud de robots recorrer muchas decenas (si no cientos) de kilómetros de forma autónoma. La aplicación en este caso es obvia: es un método útil para cualquier robot móvil que deba ser autónomo y que se desplace sobre suelo plano.

Panorama Futuro

Podemos decir que, a día de hoy, tanto la navegación autónoma de robots como la estimación de su movimiento son problemas prácticamente resueltos si el entorno en el que dichos robots operan es estático. Como bien apuntó el catedrático Dieter Fox en su charla "The 100-100 tracking challenge" [3], el reto que ahora afrontamos es el conseguir el mismo grado de precisión y eficacia en entornos cambiantes en los que el robot esté rodeado de personas y objetos que se

[3] International Conference on Robotics and Automation (ICRA), Estocolmo (Suecia), 2016.

UNIVERSIDAD DE MÁLAGA

muevan constantemente. En consecuencia, los trabajos que aborden la estimación del flujo de escena tanto por separado como conjuntamente con la odometría visual acabarán siendo cada vez más frecuentes y relevantes tanto en robótica como en visión.

A pesar del progreso realizado en los últimos años en la estimación del flujo de escena, incluyendo algunos de los trabajos que aquí presentamos, existen todavía dos grandes problemas a resolver:

- El flujo de escena es computacionalmente pesado. El hecho de que algunos algoritmos (como el PD-Flow [12] propuesto en esta tesis) se ejecuten en tiempo real no implica que este problema esté resuelto. Esos algoritmos son rápidos porque utilizan potentes GPUs o hardware dedicado como FPGAs. La mayoría de teléfonos, tablets, dispositivos de realidad virtual o robots de consumo no están equipados con hardware tan potente, y si lo estuvieran no podrían sobrecargarlo con la estimación de flujo de escena porque hay decenas de procesos adicionales que deben ejecutar. Por tanto, debemos encontrar métodos más inteligentes para calcular el flujo de escena. Las estrategias de clusterización propuesta aquí y en otros trabajos [13] representan un paso adelante en esta dirección. La estimación de movimientos independientes por cluster reduce significativamente el número de incógnitas (en dos o tres órdenes de magnitud) y frecuentemente mejora la precisión. Puede ser que la siguiente gran mejora consista en aprender a calcular las transformaciones asociadas a los clusters de manera más eficiente.

- No existe ningún método capaz de estimar flujo de escena con cámaras RGB monoculares. Esta limitación es muy importante porque las cámaras RGB son baratas, tienen un consumo energético muy bajo y están equipadas en muchos dispositivos electrónicos. El principal problema radica en que la estimación de profundidad se realiza asumiendo que la disparidad entre imágenes es causada por el movimiento de la cámara. Cuando los objetos se mueven esta hipótesis no se cumple y no existe entonces un procedimiento claro para estimar su posición en el espacio (siempre asumiendo que la escala no es conocida). La solución a este problema debe pasar por la detección de aquellas partes de la escena que no son estáticas (residuos altos después del alineamiento de acuerdo a la transformación de la cámara) y la estimación conjunta de la posición y del movimiento de dichos elementos.

Desde un punto de vista tecnológico, asistimos en los últimos años (o en las últimas décadas) a un desarrollo espectacular de los sensores de rango, acompañado también a veces de una caída notable de su precio. En esta tesis hemos trabajado principalmente con dos tipos de sensores: las cámaras de profundidad, y los escáneres láser 2D. Conscientemente, y quizás erróneamente, hemos dejado de lado los escáneres 3D. La razón principal es que, a causa de su elevado precio, no contamos con este tipo de sensor en nuestro laboratorio. Es cierto que existen varios datasets con datos de lidars, o que se pueden utilizar simuladores para trabajar con ellos, pero en todos estos años hemos querido trabajar con sensores reales que nos permitieran comprobar que los algoritmos desarrollados funcionaban en la práctica, y esto no era posible en el caso de los lidars. Sin embargo, se atisba un cambio de paradigma (quizás parecido al que ocurrió con el sensor Kinect) porque varias compañías han anunciado que próximamente pondrán a la venta lidars 3D de estado sólido, a precios significativamente más bajos que los actuales. Esta innovación

UNIVERSIDAD DE MÁLAGA

está motivada por la inminente llegada de los coches autónomos y el hecho de que dependen en gran medida de este tipo de sensores para poder funcionar. Así, la robótica, al igual que ocurrió con el sensor Kinect, podrá verse beneficiada de un avance tecnológico que no se concibió específicamente para ella (aunque, al fin y al cabo, un coche autónomo no es ni más ni menos que un robot).

UNIVERSIDAD
DE MÁLAGA

UNIVERSIDAD
DE MÁLAGA

1 Introduction

Since its origins, robotics has aimed to create autonomous machines to increase the efficiency of industrial processes, reduce manufacturing costs and free people from tedious and dangerous tasks. Given its complexity, different disciplines of engineering, physics and maths have converged to what we nowadays call *mechatronics* to address the conception, design and manufacture of these complex systems called robots. Among the many challenges still unsolved, one of upmost importance is that of endowing robots with intelligence and autonomy. In order to emulate a person, a robot must be able to perceive its surroundings, interpret its circumstances and decide how to act to achieve its goal. Thus, autonomy is directly related to the robot's capacity to perceive the environment and its ability to process that sensorial data. Like humans, robots are provided with sensors (their "senses") which allow them to see, hear, touch or even smell. And also, like for humans, the most powerful of all senses is vision. Nevertheless, robots are not only equipped with passive sensory systems (those which measure the ambient energy) but also with active sensory systems. In vision, these systems work by emitting a pattern of light which is reflected back from the environment and subsequently detected by a specific sensor. By virtue of this mechanism, active sensory systems are able to infer the geometry of the surrounding obstacles, which is extremely useful in many technological fields.

In robotics, active visual sensory systems, also known as "range sensors", are commonly employed for:

- Knowing the spatial distribution of obstacles around a robot.
- Knowing the location and orientation of objects to be manipulated.
- Building 2D or 3D maps of the environment where the robot operates.
- Estimating the 2D or 3D trajectory of a robot.
- Segmenting the observed scene into the different objects it is composed of.

Additionally, range sensors find a myriad of applications beyond robotics, for example in:

- Human-computer interaction and gaming.
- 3D modelling of objects.
- Motion analysis, both for professional sport training or for therapeutic treatment.

UNIVERSIDAD DE MÁLAGA

- Motion estimation for devices of virtual/augmented reality.
- Driver assistance systems for cars and other vehicles.

All these applications require complex algorithms able to process the geometric data provided by this kind of sensors. Up to date, numerous solutions have been proposed for each particular problem. However, these solutions are often tuned to work for particular scenarios and under controlled conditions, or present good theoretical results that lack practicality. In this thesis we propose novel algorithms to tackle/solve many of the aforementioned examples. In general, our goal has been the development of methods with a solid theoretical basis which, in turn, could be directly applicable to the addressed problem, not only in theory or in simulation but also in real-world scenarios. To this end, one must pay attention to both the formulation and the implementation of a given algorithm in order to select precise formulations that could be efficiently implemented. Following this utilitarian spirit, we have opted for publishing the code of our works so that the scientific community can use and test them.

1.A Motivation

Range sensors have existed since the beginning of the XX century but they have significantly evolved in the last decades. After the invention of the *sonar* during the I World War, the first range sensors based on light emission were the laser scanners or *lidars*, developed in the 1960s. The accuracy of these sensors was very high but so was their price, and for many years they were mostly utilized for military or spatial applications. Over time their costs dropped and simple versions (2D scanners) started to be equipped in mobile robots for obstacle detection, localization and 2D mapping. Nevertheless, the real revolution came with the advent of Microsoft's Kinect camera by the end of 2010. Kinect was the first low-cost camera (150 euros) able to provide not only colour but also depth images, and at a decently high frame rate (30 Hz). Thus, it was the first low-cost sensor that allowed to "sense" the geometry of the environment with accuracy, and its impact on the robotics community and in many other fields was, and still is, enormous. However, the Kinect sensor also posed a challenge since it provides a huge amount of data to process (colour and depth images with VGA resolution at 30 Hz). In consequence, a significant percentage of the published papers in robotics and computer vision in the last years deals with its use and application to different ends.

Additionally, the steady development of computers has enabled the approach of problems that were, until recently, computationally intractable. CPU's now offer multiple processing units (cores) with larger cache sizes. They also include special registers and instruction sets to efficiently perform certain operations on vectorized data (SSE, AVX). GPU's are no longer mere graphic processing units (as their name indicates) but massive units for parallel computation. They have significantly increased their power, memory and bandwidth and, at the same time, they have become energetically more efficient, broadening the prospects of potential applications. Simultaneously, the appearance of new libraries and compilers has eased the implementation of code that can run on multiple cores of a CPU or a GPU. If we focus on C++, modern compilers (MSVS, GCC, Intel...) generate highly optimized code without requiring much programming

UNIVERSIDAD DE MÁLAGA

effort. Newer versions of CUDA incorporate the unified memory abstraction that allows programmers to access both GPU and CPU memory within a single memory address space. Many other libraries include functionalities to exploit parallel programming, either through vectorization, multi-thread or GPU implementations: STL (C++11), OpenCL, Eigen, OpenGL, etc.

These two key aspects have motivated the work presented in this thesis. In short, the challenge is about how to exploit the newly available range sensors and the increasing computational power of modern PCs to solve fundamental problems that remain unsolved (or only partially solved) in computer vision and robotics.

1.B Goals

Despite the great advances seen in the last decades in computer vision, there are still many open problems. Among those, the ones that are more relevant in robotics often require certain geometric knowledge of the environment, and therefore can benefit from the use of range sensors. In this thesis we mostly (but not uniquely) address the estimation of 2D and 3D motion with different kinds of range sensors. In some cases the interest is in the motion of the sensor itself, while in other cases the goal is to estimate the translations and rotations of the observed objects. Besides motion estimation, we present new algorithms that also exploit geometric data for 3D reconstruction and autonomous navigation. Thus, this thesis does not tackle a single subject but rather covers a variety of topics. The common goal is to come up with novel solutions that are more precise and faster than existing approaches, or that allow a robot or system to operate under more extreme conditions (e.g. in the dark or in very dynamic environments).

Next, we explain in more detail the addressed topics, their potential applications and the inherent difficulties associated to them:

- **Range-based odometry**. Odometry consists in estimating the motion of a sensor or a set of sensors in an incremental way, i.e. by accumulating pose increments. It is a fundamental component of many robotic systems and virtual/augmented reality devices. In robotics, for instance, it permits us to know the pose of a robot precisely and continuously, which is crucial if the robot must perform an autonomous task. Regarding virtual/augmented reality, knowing the pose of the goggles in real time (and therefore the point of view of the user) is fundamental to render virtual images from the right perspective and with low latency.

  Our goal is to create new odometry algorithms based on range sensors. Given their potential applications, it is essential that these algorithms work in real time and are accurate even in dynamic environments where moving objects are often observed.

- **Scene flow estimation**. Scene flow refers to the vector field that represents the independent 3D motion (or velocity) of each point observed by a camera. It has multiple applications: 3D modelling of non-rigid bodies, manipulation of moving objects, human-machine interaction and motion analysis are a few examples.

  The state of the art in this field is less mature if compared to odometry or optical flow estimation. For that reason, our goal is to overcome (or alleviate) the main limitations that currently prevent its use in real-world scenarios. Among those, the most important one is the

UNIVERSIDAD DE MÁLAGA

high computational load associated to the estimation process, and that is what we put the focus on.

- **3D Modelling** and **tracking** of objects from a set (or sequence) of images. 3D modelling is useful to measure dimensions, insert real objects in virtual worlds, or virtually combine real objects and environments which are not in the same location (e.g. using a virtual 3D model of a piece of furniture to see how it fits in a given room). Tracking, on the other hand, can be employed for human-computer interaction, face reenactment or character animation for movies or videogames.

  In this field, our goal is to exploit background information (i.e. those image regions that do not observe the object in question) to better constrain the reconstruction or tracking problem, thereby improving the basin of convergence and the quality of the results obtained.

- **Reactive Navigation** with range sensors. This strategy allows a robot or vehicle to "react" to changing environments populated with moving obstacles while advancing towards a predefined target. In contrast to deliberative approaches, these methods do not require previous knowledge of the environment (map), but can only drive the robot towards local targets, i.e., not very far from its pose.

  Most existing approaches simplify the navigation problem by considering only 2D representations of the robot shape and the surrounding obstacles. Our goal is to overcome this conservative assumption and consider more realistic robot shapes and obstacle distributions in 3D. This is possible by virtue of the rich geometric data provided by modern range sensors.

1.C Contributions

In this section we enumerate the contributions of this thesis. A brief summary of each is included, together with the related published papers. Both the code and the demonstration videos associated to them can be found in:

`http://mapir.isa.uma.es/mjaimez`

3D visual odometry from depth images

We present a new method to register consecutive depth images in order to estimate the camera motion. We demonstrate that this method is much faster and more precise than other existing techniques that are also based on geometric alignment [2]. Moreover, we show that our approach is as accurate as other state-of-the-art methods which require both geometric and photometric data to estimate the camera motion [3]. Last, it works in real time (30 Hz or more) running on a single CPU core.

- M. Jaimez and J. Gonzalez-Jimenez, "Fast Visual Odometry for 3D Range Sensors", *IEEE Transactions on Robotics*, vol. 31, no. 4, pp. 809 - 822, 2015.

UNIVERSIDAD DE MÁLAGA

2D visual odometry from laser scans

We have developed a method to register consecutive laser scans to estimate the planar motion of a robot. This method is based on the range flow equation, applied for the first time to laser scans in [4]. We demonstrate that our approach is more efficient and precise than some of the most prominent works in scan matching [5, 6]. With a runtime of barely 1 millisecond, this methods is suitable for those robotic applications that are computationally demanding and require planar odometry.

- M. Jaimez, J. G. Monroy, J. Gonzalez-Jimenez, "Planar Odometry from a Radial Laser Scanner. A Range Flow-based Approach", *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 4479-4485, 2016.

Such formulation was later improved by including multi-scan alignment and a new symmetric range flow constraint. We also proposed a new technique to select keyscans based on modelling the estimate error as a function of the translations and rotations between the aligned scans. This strategy allows us to set a threshold directly on the error domain and use it to obtain the 2D frontier on the translation-rotation plane which would trigger the selection of a new keyscan. Moreover, we present a thorough experimental section with qualitative and quantitative comparisons both in simulation and with real data.

- M. Jaimez, J.G. Monroy, M. Lopez-Antequera, D. Cremers and J. Gonzalez-Jimenez, "Robust Planar Odometry Based on Symmetric Range Flow and Multi-Scan Alignment", *IEEE Transactions on Robotics*. **Under Review**.

Scene flow estimation in real-time with RGB-D cameras

We have developed the first method able to estimate the scene flow in real time with RGB-D cameras. It imposes photometric and geometric consistency between consecutive images and uses the Primal-Dual algorithm [7, 8] as a solver. This choice is optimal since the Primal-Dual algorithm can be easily parallelized and is therefore straightforward to implement on a GPU. As a result, we achieve runtimes two or three orders of magnitude lower than those of state-of-the-art methods, which typically require several seconds (or even a few minutes) to run.

- M. Jaimez, M. Souiai, J. Gonzalez-Jimenez and D. Cremers, "A Primal-Dual Framework for Real-Time Dense RGB-D Scene Flow", *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 98-104, 2015.

Joint segmentation and scene flow estimation with RGB-D cameras

Inspired by [9], we present a new algorithm to segment the scene into the different rigid bodies that compose it and estimate their underlying rigid motions. The main contribution of this work is the use of a smooth (non-binary) segmentation that allows us to interpolate motions along the transitions between rigid parts. We show this strategy provides better results than traditional binary segmentations when the observed scene contains nonrigid parts/objects (e.g. people).

UNIVERSIDAD DE MÁLAGA

- M. Jaimez, M. Souiai, J. Stückler, J. Gonzalez-Jimenez, D. Cremers, "Motion Cooperation: Smooth Piece-Wise Rigid Scene Flow from RGB-D Images", *International Conference on 3D Vision (3DV)*, pp. 64-72, 2015.

Visual odometry and scene flow estimation with RGB-D cameras

This is a problem of great interest and complexity because a high number of applications need to know both the trajectory of a camera and the motion of the objects it observes. The difficulty lies on the fact that, when the camera moves, the whole scene is in "apparent motion" and therefore changes in the image caused by the camera motion and those caused by the own motion of objects are hard to distinguish. In this work we describe a specific strategy to segment the image into static and moving parts. After that, the visual odometry is computed from the static parts and the scene flow is estimated for the moving objects. Furthermore, the whole algorithm runs at a frequency of 10 Hz, which makes it applicable to real-world scenarios.

- M. Jaimez, C. Kerl, J. Gonzalez-Jimenez and D. Cremers, "Fast Odometry and Scene Flow from RGB-D Cameras based on Geometric Clustering", *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3992-3999, 2017.

New background term for object reconstruction and tracking from depth images

Commonly, a 3D model of an object is computed from a set or temporal sequence of images. In those images the object is seen from different perspectives, but the surrounding objects are also observed (and are normally referred to as "background"). In this work we present a framework based on subdivision surfaces for 3D reconstruction and tracking, and describe a novel strategy to penalize projections of the 3D model onto the background regions of the images. This strategy widens the basis of convergence of the algorithm by keeping the model within the visual hull of the object.

- M. Jaimez, T. Cashman, A. Fitzgibbon, J. Gonzalez-Jimenez, D. Cremers, "An Efficient Background Term for 3D Reconstruction and Tracking with Smooth Surface Models", *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7177-7185, 2017.

Reactive navigation based on exact 3D collision determination

We generalize the algorithm presented in [10] to the "3D world", modelling the robot as a set of prisms (instead of just considering its 2D vertical projection) and exploiting the exact spatial distribution of the surrounding obstacles. The resulting method works with any possible combination of range sensors, being them 2D laser scanners, 3D laser scanners or RGB-D cameras, and it is able to generate motion commands at a frequency of 200 Hz. Moreover, it was tested with different robots for almost 20 km of autonomous navigation at different flats and office-like environments.

- M. Jaimez, J. L. Blanco and J. Gonzalez-Jimenez, "Efficient Reactive Navigation with Exact Collision Determination for 3D Robot Shapes", *International Journal of Advanced Robotic Systems*, vol. 12, no. 63, 2015.

UNIVERSIDAD DE MÁLAGA

1.D Framework and Timeline

My research work began in the middle of 2012 at the MAPIR (MAchine Perception and Intelligent Robotics) group[1], which is part of the Department of System Engineering and Automation, at the University of Málaga. This group has extensive experience in mobile robotics, computer vision and robotic olfaction. Given my background on mechanics, control and maths, the natural choice for my research topic was mobile robotics. Nonetheless, and probably because of the boom of Kinect after its appearance in 2010, I ended up working at the frontiers between robotics and computer vision or, more specifically, developing vision algorithms that are directly applicable to robotics.

Initially, we dedicated more than half a year to exploit the geometric information provided by depth cameras for autonomous navigation of wheeled robots. After that, we opted for addressing the visual odometry problem, to which we have dedicated a high percentage of the time and effort of this thesis. In this context, I had the chance to do a 4-month internship at the computer vision group[2] of the Technical University of Munich, led by Prof. Daniel Cremers. This gave me an deeper insight into optimization and variational calculus, and let me explore a new topic which would later become the second pillar of my thesis: scene flow estimation. Beyond that, this collaboration was positive for both sides and, from then on, I have also been a member of that group and pursued a dual PhD doctorate.

Subsequently, I enjoyed another 2-month internship at Microsoft Research Cambridge (UK), under the supervision of Dr. Andrew Fitzgibbon. This stay was suitable because Microsoft is the company that commercialized the Kinect sensor [1], on which most of my thesis is based, and they have also authored an endless number of related publications and applications. Additionally, my months there coincided with the launch of the Hololens [11], a device for augmented reality equipped with two depth cameras. Thanks to this good timing I could also be in contact with some of the projects that will allow the use and interaction with such device in the future.

Through all these years I have had the chance and the pleasure to meet some of the most prominent figures in computer vision and robotics. Apart from the aforementioned internships, I have attended international conferences in Seattle (USA), Lyon (France), Stockholm (Sweden), Singapore and Hawaii (USA). Aside from research, I have also participated in teaching, imparting the "Computer Vision" course together with my supervisor Prof. Javier González Jiménez at the University of Málaga. In summary, these have been vibrant years of constant change and learning, enthusiasm and stress, defeats and victories.

1.E Outline

This thesis is divided into the following chapters:

- **Chapter 1: Introduction.** We present the theme of the thesis, describe its context, list the contributions and detail its structure.

[1] http://mapir.isa.uma.es

[2] http://vision.in.tum.de

UNIVERSIDAD DE MÁLAGA

- **Chapter 2: Range Sensors.** The concept of *range sensing* is introduced. The different types of range sensors are enumerated and their working principles explained. We describe their advantages and disadvantages if compared to other types of sensors commonly employed in robotics and computer vision, and present their potential applications.

- **Chapter 3: Odometry with range sensors.** We introduce the concepts of odometry and visual odometry. We describe the mathematical tools used to represent rigid transformations and to warp images and scans according to them. Afterwards, we present our works on odometry based on range sensors, both with depth cameras and with 2D laser scanners.

- **Chapter 4: Scene flow estimation with RGB-D cameras.** We introduce the concept of scene flow as a 3D extension of the well-known optical flow. We present two different contributions: the scene flow estimation in real time, and the joint segmentation and scene flow estimation based on a smooth division of the scene into its rigidly moving parts. Moreover, we tackle an ambitious problem that is not often addressed in the literature: the joint estimation of the camera motion and the motion of the objects it observes.

- **Chapter 5: Reconstruction and tracking of objects from depth images.** Since the literature on this topic is very extensive, we give a brief overview of the state-of-the-art focusing on the different strategies that use the "background" information to constraint the model to the visual hull of the object. We present a new background term to exploit such information and demonstrate its advantages over existing approaches when used for modelling or tracking of objects with subdivision surfaces.

- **Chapter 6: Reactive navigation of mobile robots equipped with range sensors.** In this chapter we introduce the concept of reactive navigation which, together with path planning, governs the autonomous motion of mobile robots. In contrast to most existing approaches, which normally use a two-dimensional representation of the robot and the obstacles, we propose to overcome that simplification and tackle the collision determination problem in 3D. Hence, the proposed algorithm works precisely for robots with non-uniform shapes and for any arbitrary combination of different range sensors.

- **Chapter 7: Conclusions.** We finish this thesis by summing up the presented works and their impact. Moreover, we look towards the future, analyzing the upcoming technologies and some of the challenges that still remain unsolved.

UNIVERSIDAD
DE MÁLAGA

2
Range sensing

2.A Introduction

The term *range sensing* refers to a set of technologies that allow the measure of distances between a given device and its surroundings. These technologies are usually (but not exclusively) based on active sensory systems, i.e. systems that probe the environment by emitting energy and measuring how or when it is reflected. There exists a number of different working principles on which these systems are based, but most resort to the emission and reception of sound, light or radio waves.

The ones based on sound are often called *sonars* and are mostly used for marine navigation and submarine military applications. They are also equipped in mobile robots and cars to work as proximity sensors for collision prevention (*e.g.* in the system that beeps when a car gets too close to another car while parking). However, their limited accuracy and resolution prevents them from being used for more complex tasks. Alternatively, systems based on radio waves (*radars*) are a suitable solution to scan the environment under bad visibility conditions since radio waves can penetrate dust, rain, fog or snow. Moreover, this technology has the advantage of allowing the detection of multiple objects at the same bearing. However, radars have a low angular resolution (for small-size antennas) if compared to light-based sensors and their measurements are affected by specular reflections and depend on the different materials of the surrounding objects. For these reasons, radars are often used in middle-to-large scale systems but are not very common in robotics or other applications involving small devices.

This thesis focuses on the range sensors more frequently used in robotics, human-machine interaction and visual/augmented reality: laser scanners and depth-sensing cameras. A more detailed description of their working principles and their characteristics is given next.

2.B Laser Scanners

A laser scanner, also known as *lidar*, is a device that senses distance by emitting laser light and measuring the time it takes to return to the sensor. Since light pulses or waves can only be used to measure distance for a specific bearing, lidars normally incorporate oscillating mirrors to be able to scan in multiple directions. Depending on the degrees of freedom of this oscillating mechanism, lidars will provide 2D or 3D scans of their surroundings (see Figure 2.1). In robotics,

UNIVERSIDAD DE MÁLAGA

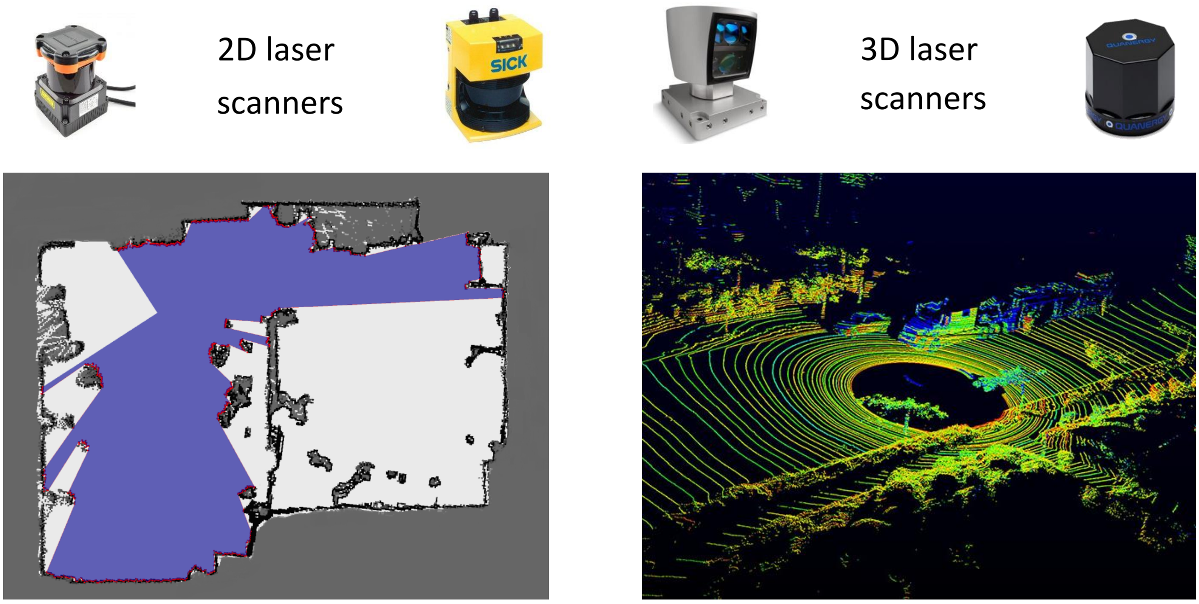

Figure 2.1: Left: Hokuyo UTM-30LX and SICK S3000 laser rangefinders, together with an example of a 2D scan (270°) overlapped with a previously-built map of the environment. **Right**: Velodyne HDL-64E and Quanergy laser rangefinders, with and example of a 3D scan (360° × 27°) taken with a Velodyne sensor mounted on a car.

2D lidars are commonly employed to obtain horizontal "slices" or "cuts" of the environment, which is very useful for indoor localization, 2D mapping and obstacle avoidance. On the other hand, 3D lidars are mostly employed for outdoor applications. In these cases, the motion of a robot (or a vehicle) is not always planar and the 3D spatial distribution of both close and distant obstacles becomes more relevant. Aside from robotics and autonomous navigation, the 3D lidar technology is also of interest in geodesy, geology, meteorology and military applications, but this is out of the scope of this thesis.

The main advantages of lidars are:

- Long range. They can provide measurements from few centimeters to more than 100 meters, depending on the model. As a consequence, they are suitable for indoor operation in cluttered environments but also for outdoor operation at high speeds where detecting distant objects/obstacles becomes crucial.
- Wide field of view. Their horizontal aperture typically ranges from 90 degrees to 360 degrees, which is much wider than the standard field of view of digital cameras.
- Good angular resolution, normally below 1 degree.
- High accuracy. They exhibit low measurement errors that are either constant (for short ranges) or linear with the distance.
- Medium-high sampling rate. This is necessary to operate in dynamic environments and react in time to changing conditions.

If compared to other sensors, their main disadvantage is their relatively high price. The cheapest 2D laser scanners on the market cost around 400€ - 500€, but that price rises up to many

UNIVERSIDAD
DE MÁLAGA

thousands of euros for low-to-middle quality 3D laser scanners. Moreover, lidars have a high power consumption, are not robust to shaky motions and their performance deteriorates under the presence of fog, heavy rain or dust. Nowadays, this prevents 3D lidars from being a basic component of mobile robots and vehicles, but advances in solid state-based technologies might help to bring them to the consumer markets.

2.C Depth-sensing Cameras

In contrast to standard digital cameras, which capture the colour (RGB) of the scene, depth-sensing cameras store the distance to the points they observe. With a high (spatial) resolution, they can provide a fine-grained representation of the observed objects. Regarding their accuracy, they tend to be precise for close distances but become inaccurate for mid-long range measurements, with an error that typically grows quadratically with the distance [14]. On the other hand, their field of view is normally equivalent to that of a digital camera equipped with a 25-30 mm lens (e.g. 58° × 45° for Kinect 1). In this sense, depth cameras are inferior to laser scanners because they do not provide a comprehensive view of the scene. Last, they can work at 30 Hz - 60 Hz, which is sufficient for most applications.

Next, we present the two different working principles these cameras rely on, as well as their particular advantages and disadvantages.

2.C.1 Time-of-flight Cameras

This type of cameras emit light pulses and measure their time of flight (ToF) until they are received back at the sensor. Unlike lidars, these cameras do not incorporate rotating parts, flashing the scene only once per image and capturing the reflection by an image sensor / pixel matrix. Examples of these cameras are the Microsoft Kinect 2.0 and the Heptagon Swiss Ranger 4000 (Figure 2.2). Since they normally emit and detect infrared light, they are affected by the sun radiation. However, they can still provide decent measurements in outdoor scenarios if there is no direct sunlight. As can be seen in Figure 2.2, one of their drawbacks is the fact that they are prone to creating spurious points around the silhouettes of objects or, more precisely, at depth

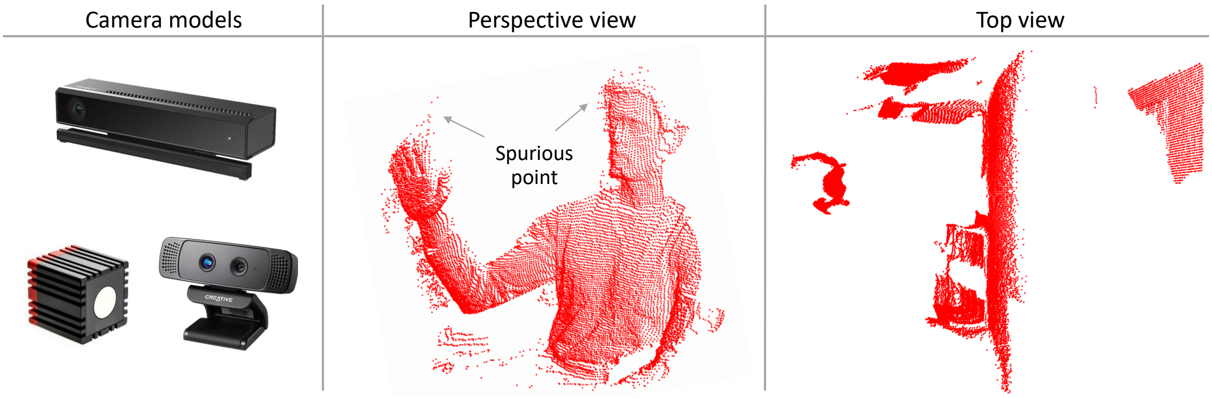

Figure 2.2: Left: Kinect 2, Heptagon Swiss Ranger and Creative Gesture cameras. **Middle**: Example of the 3D point cloud computed from a depth image obtained with a Kinect 2. **Right**: Top view of a scene observed with Kinect 2.

UNIVERSIDAD DE MÁLAGA

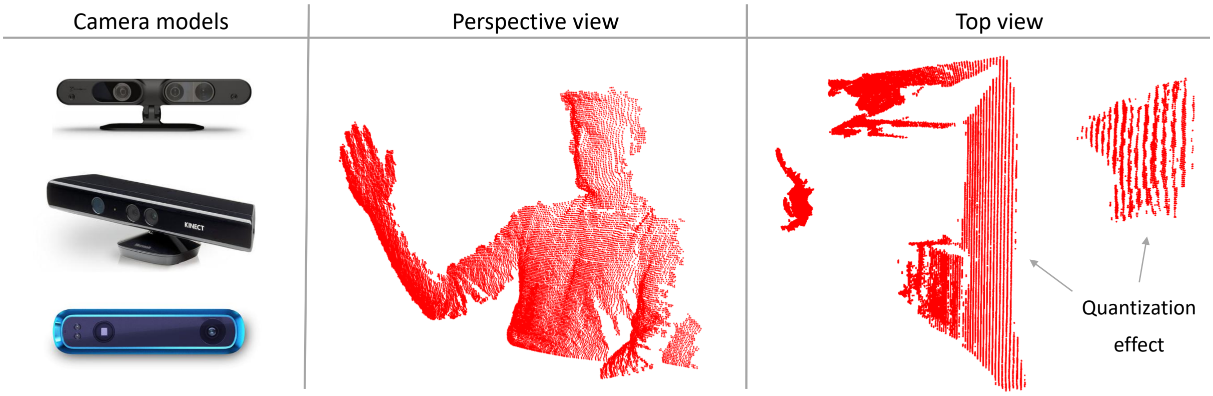

Figure 2.3: Left: PrimeSense Carmine, Kinect 1 and Structure sensor. **Middle**: Example of the 3D point cloud computed from a depth image obtained with a PrimeSense Carmine. **Right**: Top view of a scene observed with the same camera.

discontinuities. These spurious points can be removed at a post-processing stage, at the expense of consuming computational resources.

2.C.2 Structured light Cameras

Structured light cameras work by projecting a known pattern of light on the scene and inferring the depth field from the way that pattern deforms. The main example of this kind is the Microsoft Kinect 1 (although the technology behind Kinect 1 was developed by PrimeSense, which later also produced their own depth cameras: the PrimeSense Carmine). Structured light cameras typically employ an infrared pattern of light and, hence, become completely blind under the presence of sunlight. The reason is that the sun radiation is much more intense than the pattern emitted by the camera, and therefore the latter is masked and cannot be perceived by the sensor. Commonly, they exhibit another downside: a strong quantization of the measured depth. As this quantization actually affects the inverse depth, it is not noticeable for short distances but becomes extreme for distant regions, as can be seen in Figure 2.3.

2.D RGB-D Cameras

Both structured light and ToF cameras can be found as independent devices on the market, but they are often combined with RGB cameras, leading to the so-called *RGB-D cameras*. This synergy is suitable for robotics and many other fields because it combines both photometric and geometric data to provide a four-dimensional representation of the observed scene. An issue when jointly exploiting RGB and depth images is that of properly registering them. Since RGB and depth images are captured by different lenses, there is no direct correspondence between their pixels. In other words, a point observed by a certain pixel in the RGB image does not coincide with the point observed by that same pixel in the depth image. In fact, there will often be points of the scene which are visible from the RGB lens but not from the infrared lens, and vice versa. Although RGB-D cameras often incorporate a function to automatically register colour and depth images, that registration is far from perfect, as illustrated in Figure 2.4. There exist methods to improve that registration by aligning the edges (regions of high gradients) of

UNIVERSIDAD DE MÁLAGA

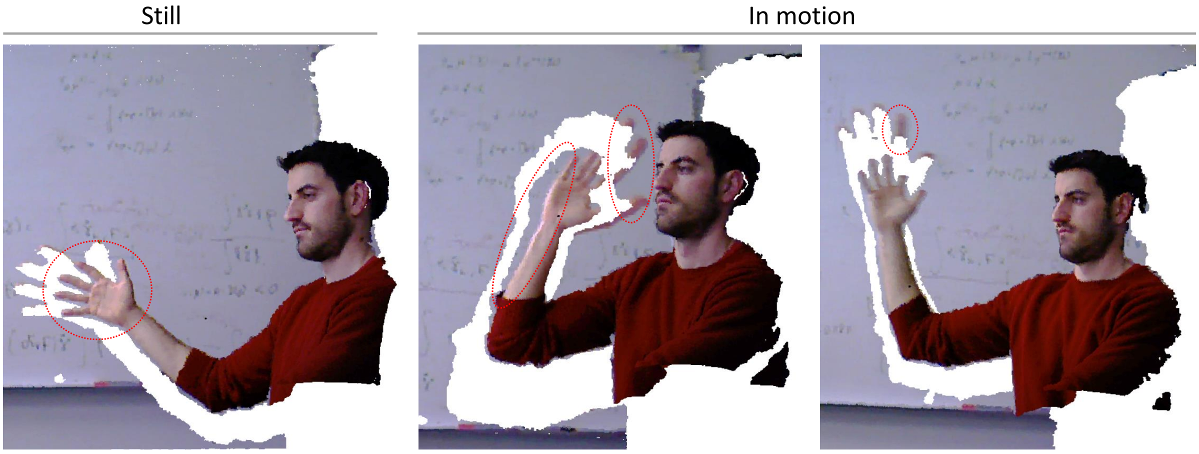

Figure 2.4: Examples of 3D coloured point clouds built from automatically registered RGB-D images taken with a PrimeSense Carmine.

the two images or warping them according to the rigid transformation existing between the two lenses (calibration). Both approaches require extra computation and will always imply at least a small increment in the runtime of the proposed algorithm. For this reason, the solution to this problem often consists in developing methods which are robust to the RGB-D misalignment.

UNIVERSIDAD DE MÁLAGA

UNIVERSIDAD
DE MÁLAGA

3 Range-based Odometry

3.A Introduction

The term *odometry* resulted from the combination of two greek words[1]: *odos*, which means road or path, and *metros*, a measure. Thus, odometry is related to the act of measuring the length of a certain path or trajectory. Originally, the term odometry was specifically used to refer to what we nowadays call *wheel odometry*, and consisted in counting the number of turns of one or several wheels to estimate the planar trajectory of a vehicle. This process is inherently incremental, since the trajectory can only be estimated as a sum of small pose increments (i.e. one cannot estimate the whole trajectory of a vehicle from the final count of wheel turns, this problem is highly underconstraint). In fact, wheel odometry can only be exact if these increments are infinitesimally small and integrated through time, which is impossible in practice. Alternatively, *visual odometry* resorts to align consecutive images to estimate the motion of the sensor. To clarify that we employ range measurements in our work, we will also coin the term *range-based odometry*, although both might be equivalent in many cases (e.g. when working with depth images). Like wheel odometry, visual and range-based odometry work incrementally as they require a certain degree of overlap between the incoming data to be able to align them. In both cases, since odometry continuously builds upon previous estimates, it accumulates and propagates the estimation errors through time. For this reason, having accurate motion estimates becomes crucial.

In general, there exist a number of technologies to track the motion of a given vehicle or device:

- The *Global Positioning System*, or GPS, was developed by the Defense Department of the USA during the Cold War. It consists of a set of satellites that are permanently orbiting the Earth and emitting electromagnetic signals. These signals can be received on earth to pinpoint, through triangulation, the position of the receiver with errors of just a few centimeters. However, GPS's are unable to provide information about orientation and they do not work in indoor scenarios where the satellites' signals cannot be detected.

- Similarly, but at a lower scale, indoor systems use arrays of sensors to track the 3D motion of a special marker. Some are based on infrared cameras and reflectors [15], others on RFID

[1] Etymological reference from https://en.wikipedia.org/wiki/Odometry

UNIVERSIDAD DE MÁLAGA

tags [16], or even on ultrasounds [17]. The common limitation of all these approaches is the need of a controlled environment, which is not possible in many applications.

- *Inertial Measuring Units*, or IMUs, measure accelerations, rotational velocities and sometimes the magnetic field of the Earth. Through temporal integration, and by detecting the direction of gravity, these devices can be used to estimate translations and orientation. On the one hand, orientation can be obtained quite accurately by leveraging the different measurements that IMUs provide. On the other hand, translations are hard to obtain given the second-order nature of the estimation process and the difficulty associated with canceling the gravitational component of the measurements [18].

- Encoders are utilized to measure wheel turns (wheel odometry) or to know the relative positions of robotic arms, legs, etc. The main and significant drawback associated to wheel odometry is slippage, since in that case the rotation of the wheels does not correspond to the motion of the vehicle/robot. More generally, estimates based on encoders require a precise geometric model of the limbs or parts in motion, which is not always known with such precision.

- Visual and range-based odometries are very flexible solutions since they can be particularized to work with different sensors (RGB or depth cameras, stereo systems, laser scanners) and configurations (2D, 3D or any particular combination of translations and rotations). As a drawback, they are sensitive to the illumination conditions or the solar radiation.

Each of the aforementioned technologies are suitable for some specific tasks, but in recent years visual odometry has demonstrated to be the most precise and versatile alternative. For instance, it is used to provide continuous estimates of the 3D pose of a drone [19], which is mandatory if the drone is to carry out any autonomous task. Similarly, planar odometry based on scan matching is employed to know the position of service robots equipped with laser scanners (e.g. telepresence robots [20]). Planar odometry does not always rely on lidars though, and other platforms like the Roomba vacuum cleaner [21] incorporate solutions based on RGB cameras (probably combined with wheel odometry and proximity sensors). The Microsoft augmented-reality device Hololens [11] includes a set of RGB and depth cameras to perform 3D SLAM, and to do so it must align the incoming photometric and geometric data provided by these cameras in real time. In general, almost any application related to 2D or 3D mapping includes an odometry module to be able to integrate the data from images or scans correctly. As a last illustrative case, visual odometry is even employed for video processing, for example to render smoother versions of time-lapse videos [22].

3.B Rigid Transformations and Lie Algebra

In this section we introduce the mathematical tools utilized throughout this thesis to handle geometric transformations in Euclidean spaces, particularly 2D and 3D rigid transformations.

UNIVERSIDAD DE MÁLAGA

A rigid transformation is the one that preserves distances between the transformed points, hence moving them as if they were part of a rigid body. Rigid transformations encompass translations and rotations, and are normally expressed as

$$T = \left(\begin{array}{c|c} R & \boldsymbol{t} \\ \hline 0 & 1 \end{array}\right) , \tag{3.1}$$

where R is a rotation matrix and $\boldsymbol{t}$ is a translation vector. In the three-dimensional case, $R \in \mathbb{R}^{3\times 3}$ and $\boldsymbol{t} \in \mathbb{R}^3$, while in a 2D space $R \in \mathbb{R}^{2\times 2}$ and $\boldsymbol{t} \in \mathbb{R}^2$. This matrix can be used to transform points from one reference system to another, or equivalently, to move them according to a certain rigid body motion. If $\boldsymbol{p}$ is a vector with the coordinates of a given point P, its transformed coordinates $\boldsymbol{p}'$ can be computed as

$$\boldsymbol{p}'_h = T\boldsymbol{p}_h, \qquad \boldsymbol{p}'_h = \begin{pmatrix} \boldsymbol{p}' \\ 1 \end{pmatrix}, \qquad \boldsymbol{p}_h = \begin{pmatrix} \boldsymbol{p} \\ 1 \end{pmatrix}, \tag{3.2}$$

where $\boldsymbol{p}_h$ and $\boldsymbol{p}'_h$ are the homogeneous coordinates of the point before and after the transformation, respectively. Using homogeneous coordinates is advantageous because they allow us to concatenate multiple transformations as a simple product of matrices.

The only inconvenience associated to this representation of rigid motions is the fact that it is not minimal because the size of rotation matrices (R) is larger than the actual degrees of freedom (e.g. rotations in 3D can be encoded by a vector of 3 elements but R is 3×3). Fortunately, there exists a theory that connect spatial transformations (Lie groups) and their minimal representation (Lie algebras). A Lie group is a group that is also a differentiable manifold, with the property that the group operations are compatible with the smooth structure. Every Lie group has an associated Lie algebra, which is a vector space generated by differentiating the group transformations along specific directions in the space. This vector space (or tangent space) has the same dimensions as the number of degrees of freedom of the group transformations and is an optimal space to represent differential quantities related to the group [23, 24]. In this thesis we will work with the Lie algebras $\mathfrak{se}(2)$ and $\mathfrak{se}(3)$ associated to 2D and 3D rigid transformations, respectively. The corresponding Lie groups are commonly denoted as SE(2) and SE(3). If $\boldsymbol{\xi} = (\boldsymbol{\nu} \;\; \boldsymbol{\omega})^T$ is a vector of the Lie algebra where $\boldsymbol{\nu}$ encodes the translational component and $\boldsymbol{\omega}$ the rotational one, the rigid transformation associated to it is given by the following exponential map:

$$T = \exp(\hat{\boldsymbol{\xi}}) = \exp\left(\begin{array}{c|c} \omega_\times & \boldsymbol{\nu} \\ \hline 0 & 0 \end{array}\right) , \tag{3.3}$$

where $\omega_\times$ is a skew symmetric matrix built from the rotational components of $\boldsymbol{\xi}$:

$$\omega_\times = \begin{pmatrix} 0 & -\omega \\ \omega & 0 \end{pmatrix} , \quad \text{if } \boldsymbol{\xi} \in \mathfrak{se}(2) , \tag{3.4}$$

$$\omega_\times = \begin{pmatrix} 0 & -\omega_z & \omega_y \\ \omega_z & 0 & -\omega_x \\ -\omega_y & \omega_x & 0 \end{pmatrix} , \quad \text{if } \boldsymbol{\xi} \in \mathfrak{se}(3) . \tag{3.5}$$

UNIVERSIDAD
DE MÁLAGA

Similarly, the vector $\boldsymbol{\xi}$ associated to a given transformation T can be computed with the logarithm map:

$$\hat{\boldsymbol{\xi}} = \log\left(T\right) . \tag{3.6}$$

Last, we want to describe the physical meaning of the vector $\boldsymbol{\xi}$, which is normally missing in the literature. For clarity we will consider only the 3D case, but the explanation is equally valid for 2D. Let's assume $\boldsymbol{\xi}$ represents the velocity of a rigid body, and P is a point that belongs to that rigid body. Under these assumptions, the instant velocity of P can be expressed as

$$\boldsymbol{v}_P = \begin{pmatrix} \nu_x + z\omega_y - y\omega_z \\ \nu_y - z\omega_x + x\omega_z \\ \nu_z + y\omega_x - x\omega_y \end{pmatrix} , \tag{3.7}$$

where (x, y, z) are the coordinates of P. To obtain the position of P at a given time t we integrate its instant velocity:

$$\boldsymbol{p}(t) = \boldsymbol{p} + \int_0^t \boldsymbol{v}_p \, dt \, . \tag{3.8}$$

Note that this integral must be computed numerically because the coordinates are interdependent. It can be proved (although we do not do it in this thesis) that, if we compute the position of P at t=1, that position coincides with $\boldsymbol{p}'$ (3.2):

$$\boldsymbol{p'} = \boldsymbol{p}(1) = \boldsymbol{p} + \int_0^1 \boldsymbol{v}_p \, dt \, . \tag{3.9}$$

Therefore, the vector $\boldsymbol{\xi}$ contains the translational and rotational velocities of a rigid body according to which P moves (temporally normalized). If the transformation does not correspond to any motion but to a change of reference frame, then $\boldsymbol{\xi}$ represents the rigid velocity that would bring one of the reference frames towards the other in one second.

3.C Coarse-to-Fine and Theory of Warping

The problem of dense image alignment has extensively been studied in computer vision. The term *dense* means that all the image pixels must be aligned, not only those observing special features. The two-dimensional motion field that encodes the displacement of each individual pixel on the image plane is called *optical flow*. Virtually every problem related to image alignment involves the estimation of the optical flow implicitly or explicitly. In visual odometry, the relation between the camera motion and the optical flow is normally embedded in the formulation and the projection model (e.g. pinhole). Another case/example of interest is scene flow, which is often decomposed as the estimation of optical flow plus range flow (or disparity). All these problems are non-convex and require linear approximations, commonly based on the optical flow constraint [25], to be solved. However, these linearizations are quite restrictive. Since they rely on the image gradients, each constraint becomes valid only locally in a small region surrounding each pixel, providing very little basin of convergence for many algorithms. Fortunately, this limitation has been overcome in the past by imposing such linearizations in a coarse-to-fine scheme [26, 27].

A coarse-to-fine strategy consists in building a pyramid of images which are subsequently aligned from the coarsest to the finest level. The coarsest levels offer a wider basin of convergence

UNIVERSIDAD DE MÁLAGA

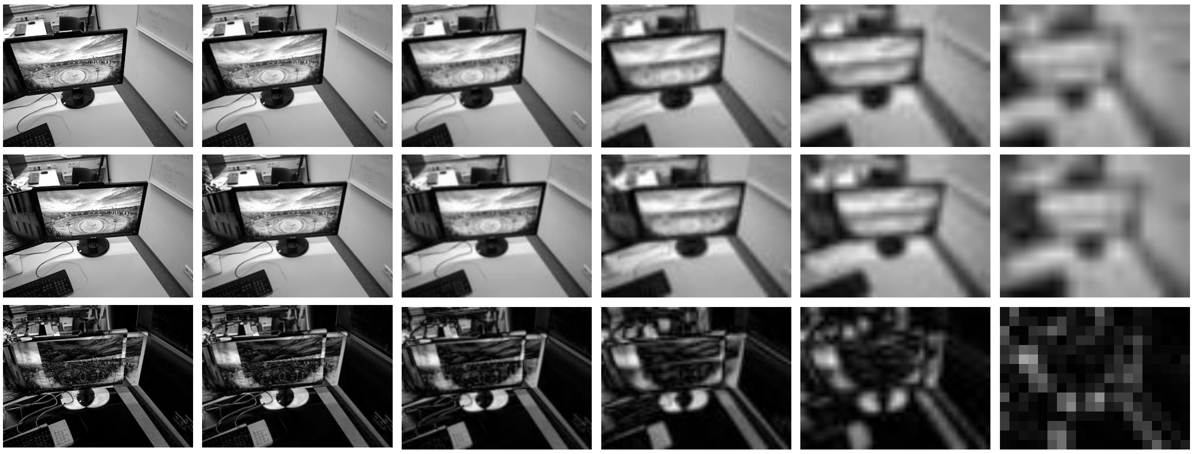

Figure 3.1: Example of an image pyramid. Resolution goes from VGA (left) to 15×20 (right). The first rows show two consecutive intensity images taken with an RGB-D camera, and the last row shows the residual image (absolute difference between the two). It can be observed that the misaligned areas cover many pixels in the original images, whereas they are only one or two pixels wide in the coarsest image. Hence, any linearization applied on the finest level would not help to align both images while the same process applied on the coarsest level would work (but not with precision).

because each pixel there covers a "larger area" of the scene (Figure 3.1), whereas the finest levels allow for precise alignment once the algorithm gets close to the solution. To avoid alliasing and information loss, image pyramids are built by combining downsampling and smoothing. The former typically divides the image resolution by two at every step, and the latter is computed by convolving the image with a Gaussian kernel. However, when it comes to geometric data, Gaussian smoothing generates spurious points at depth discontinuities and, for that reason, a bilateral filter [28] can be used instead.

Once the solution is obtained for a given level of the coarse-to-fine scheme, that solution is used to warp one of the two images to align it with the other. This step is fundamental because the linearizations at the following levels would not hold otherwise. There are different strategies to perform this warping, depending on the addressed problem. Next, we describe the basic warping strategy based on the optical flow, and another case of interest for this thesis: when both the scene geometry and the rigid transformation between the images are known.

3.C.1 Warping with optical flow

Let $I_1, I_2 : \Omega \to \mathbb{R}$ be two intensity images, where $\Omega \subset \mathbb{R}^2$ denotes the image domain. An image pyramid is computed for each input image; $I_{(.)}^L$ will be used to refer to the image of the L-th level of the pyramid. If $\boldsymbol{u}^L$ is the optical flow that maps I_2^L towards I_1^L then:

$$I_2^L(\boldsymbol{x} + \boldsymbol{u}^L) \approx I_1^L(\boldsymbol{x}) \,, \tag{3.10}$$

where $\boldsymbol{x}$ represents the coordinates of a given pixel. The warping in this case is performed by creating a new image $I_{2,w}^{L+1}$ for the next level which is as similar/close as possible to I_1^{L+1}, i.e.

$$I_{2,w}^{L+1}(\boldsymbol{x}) = I_2^{L+1}(\boldsymbol{x} + \hat{\boldsymbol{u}}^L) \,, \tag{3.11}$$

where $\hat{\boldsymbol{u}}^L$ is an upsampled version of $\boldsymbol{u}^L$. Subsequently, at level $L+1$, the new warped image $I_{2,w}^{L+1}$ will be aligned with I_1^{L+1} with a finer optical flow $\boldsymbol{u}^{L+1}$. This process is repeated for

UNIVERSIDAD DE MÁLAGA

each level, thereby concatenating the solutions obtained throughout the pyramid and bringing I_2 closer to I_1 at each new step.

3.C.2 Warping with rigid transformations

In this section we describe the warping for visual odometry with RGB-D cameras and therefore assume that the geometry of the scene is known. Let $I_1, I_2 : \Omega \to \mathbb{R}$ and $Z_1, Z_2 : \Omega \to \mathbb{R}$ be two registered pairs of intensity and depth images, respectively. In this case four image pyramids are built, and $I^L_{(.)}$ and $Z^L_{(.)}$ will denote the intensity and depth images of the L-th level of these pyramids. In visual odometry, the optical flow is not computed explicitly but it can be calculated from the estimated transformation and the camera projection model. Let $\pi : \mathbb{R}^3 \to \Omega$ be a function that projects 3D points onto the image plane and $\pi^{-1} : \Omega \times \mathbb{R} \to \mathbb{R}^3$ be the inverse projection function which provides the 3D coordinates of any observed point. If $T_L \in$ SE(3) is the rigid transformation that aligns the points of the two images at a given level L, then

$$I_2^L\left(\pi(T_L\,\pi^{-1}(\boldsymbol{x}, Z_1^L(\boldsymbol{x})))\right) \approx I_1^L(\boldsymbol{x}) \tag{3.12}$$

It must be remarked that T_L actually results from composing all the transformations obtained from the first until the L-th level of the pyramid. When working with RGB-D data, both intensity and depth images must be warped according to the estimated transformation:

$$I_{2,w}^{L+1}\left(\pi(T_L^{-1}\,\pi^{-1}(\boldsymbol{x}, Z_2^{L+1}(\boldsymbol{x})))\right) = I_2^{L+1}(\boldsymbol{x})\,, \tag{3.13}$$

$$Z_{2,w}^{L+1}\left(\pi(T_L^{-1}\,\pi^{-1}(\boldsymbol{x}, Z_2^{L+1}(\boldsymbol{x})))\right) = \left|T_L^{-1}\,\pi^{-1}(\boldsymbol{x}, Z_2^{L+1}(\boldsymbol{x}))\right|_z\,, \tag{3.14}$$

where $|\bullet|_z$ is the z-coordinate. In this case, the inverse transformation T_L^{-1} is employed to transform the whole coloured point cloud (computed from the second image) and render a new artificial image from the point of view of I_1, Z_1. This is the most accurate procedure to obtain the warped images but it is not the most efficient one because it involves averaging all the projections and/or keeping a z-buffer to take the ones which are visible from the camera. There exists a fast alternative warping, equivalent to the one presented in (3.11), which can be computed as

$$I_{2,w}^{L+1}(\boldsymbol{x}) = I_2^{L+1}\left(\pi(T_L\,\pi^{-1}(\boldsymbol{x}, Z_1^{L+1}(\boldsymbol{x})))\right)\,, \tag{3.15}$$

$$Z_{2,w}^{L+1}(\boldsymbol{x}) = Z_2^{L+1}\left(\pi(T_L\,\pi^{-1}(\boldsymbol{x}, Z_1^{L+1}(\boldsymbol{x})))\right)\,. \tag{3.16}$$

This method computes the warped image pixel-wise and hence it is easy to parallelize, but it might create artefacts because it combines observations from both RGB-D pairs which do not observe exactly the same points of the scene.

3.D Contributions

We have developed a new dense method for range-based odometry. This method expresses the range flow constraint [29, 30] as a function of the motion of the sensor, and provides accurate estimates by minimizing the resulting overconstrained system. The two main advantages of this method are its low computational runtime and the fact that it does not rely on photometric data. Consequently, it can be particularized to work with any range sensor.

UNIVERSIDAD
DE MÁLAGA

First, we present our initial work on 3D visual odometry with depth cameras (§3.E). The algorithm has a low computational cost if compared to other existing approaches and runs in real time on a single CPU core (QVGA resolution). Aside from the main formulation, the paper describes a new temporal filter for the camera pose based on the uncertainty of the estimates, and a special strategy to compute the image gradients. It does not incorporate robust functions in the minimization problem (weighted squared residuals are minimized) and, consequently, its performance deteriorates in the presence of moving objects.

Second, we adapt that formulation to estimate planar motion with 2D laser scanners (§3.F). In this case we minimize a robust function of the geometric residuals to improve accuracy against moving object. Thus, we include a two-fold strategy to robustify our algorithm: a pre-weighting stage (similar to §3.E) to downweight points for which the range flow constraint (linearization) does not hold, and a robust penalty function to handle the remaining outliers during the optimization. Results show that our method is very precise and much faster (0.9 millisecond) than other scan-matching algorithms.

Third, in §3.G we extend the work presented in §3.F by introducing a novel symmetric range flow constraint and aligning multiple scans at each iteration. The symmetric formulation equidistributes the estimated motion in both scans, which results in a lower linearization error. The multi-scan approach combines consecutive and keyscan-based alignment to reduce drift and increase robustness against moving objects. Moreover, we present a new keyscan-selection criterion that allows us to impose thresholds directly on the error domain (as opposed to traditional strategies which put limits to other related magnitudes like the maximum translation or rotational, average residual, etc.). We include a thorough experimental evaluation demonstrating that our method outperforms state-of-the-art algorithms in scan matching.

UNIVERSIDAD DE MÁLAGA

3.E Fast Visual Odometry for 3-D Range Sensors

Mariano Jaimez and Javier Gonzalez-Jimenez

Abstract:

This paper presents a new dense method to compute the odometry of a free-flying range sensor in real time. The method applies the range flow constraint equation to sensed points in the temporal flow to derive the linear and angular velocity of the sensor in a rigid environment. Although this approach is applicable to any range sensor, we particularize its formulation to estimate the 3-D motion of a range camera. The proposed algorithm is tested with different image resolutions and compared with two state-of-the-art methods: generalized iterative closest point (GICP) [2] and robust dense visual odometry (RDVO) [3]. Experiments show that our approach clearly overperforms GICP which uses the same geometric input data, whereas it achieves results similar to RDVO, which requires both geometric and photometric data to work. Furthermore, experiments are carried out to demonstrate that our approach is able to estimate fast motions at 60 Hz running on a single CPU core, a performance that has never been reported in the literature. The algorithm is available online under an open source license so that the robotic community can benefit from it.

Published in IEEE Transactions on Robotics, 2015.
DOI: 10.1109/TRO.2015.2428512

©IEEE (Revised layout)

UNIVERSIDAD
DE MÁLAGA

3.F Planar Odometry from a Radial Laser Scanner. A Range Flow-based Approach

Mariano Jaimez, Javier G. Monroy and Javier Gonzalez-Jimenez

Abstract:

In this paper we present a fast and precise method to estimate the planar motion of a lidar from consecutive range scans. For every scanned point we formulate the range flow constraint equation in terms of the sensor velocity, and minimize a robust function of the resulting geometric constraints to obtain the motion estimate. Unlike traditional approaches, this method does not search for correspondences but performs dense scan alignment based on the scan gradients, in the fashion of dense 3D visual odometry. The minimization problem is solved in a coarse-to-fine scheme to cope with large displacements, and a smooth filter based on the covariance of the estimate is employed to handle uncertainty in unconstraint scenarios (e.g. corridors). Simulated and real experiments have been performed to compare our approach with two prominent scan matchers and with wheel odometry. Quantitative and qualitative results demonstrate the superior performance of our approach which, along with its very low computational cost (0.9 milliseconds on a single CPU core), makes it suitable for those robotic applications that require planar odometry. For this purpose, we also provide the code so that the robotics community can benefit from it.

Published in Proc. International Conference on Robotics and Automation (ICRA), 2016.
DOI: 10.1109/ICRA.2016.7487647

©IEEE (Revised layout)

UNIVERSIDAD DE MÁLAGA

3.G Robust Planar Odometry based on Symmetric Range Flow and Multi-Scan Alignment

Mariano Jaimez, Javier G. Monroy, Manuel Lopez-Antequera,
Daniel Cremers and Javier Gonzalez-Jimenez

Abstract:

This paper presents a dense method for estimating planar motion with a laser scanner. Starting from a symmetric representation of geometric consistency between scans, we derive a precise range flow constraint and express the motion of the scan observations as a function of the rigid motion of the scanner. In contrast to existing techniques, which align the incoming scan with either the previous one or the last selected keyscan, we propose a combined and efficient formulation to jointly align all these three scans at every iteration. This new formulation preserves the advantages of keyscan-based strategies but is more robust against suboptimal selection of keyscans and the presence of moving objects.

An extensive evaluation of our method is presented with simulated and real data in both static and dynamic environments. Results show that our approach is one order of magnitude faster and significantly more accurate than existing methods in all the conducted experiments. With a runtime of about one millisecond, it is suitable for those robotic applications that require planar odometry with low computational cost. The code is available online as a ROS package.

Submitted to IEEE Transactions on Robotics, 2017.

©IEEE (Revised layout)

UNIVERSIDAD DE MÁLAGA

4 Scene Flow Estimation

4.A Introduction

The term *scene flow* refers to the dense 3D motion field of a scene between two instants of time, and can be regarded as a 3D extension of the well-known optical flow. In contrast to optical flow, the scene flow estimation is a problem that has not often been addressed in the literature since it demands a geometric knowledge of the environment. For this reason, the earliest works on scene flow estimation utilized stereo systems. These methods had to estimate disparity before or simultaneously to the scene flow, which noticeably increases the complexity of the problem. In this context, most scene flow algorithms were extensions of optical flow approaches, where the third component of the motion (shift in depth) was simply obtained as a by-product of the optical flow and the disparity field. However, the arrival of the RGB-D cameras changed this trend. This type of cameras makes it possible to separate scene flow from the disparity estimation problem, which has promoted research in this field.

Scene flow finds a myriad of potential applications. It can be very useful in dynamic environments as a tool to predict the future locations of moving objects. It can also be employed to enhance human-computer interaction by adding information about the velocity of the points of the scene. It is already an important component of many systems developed for tracking and non-rigid 3D reconstruction [91, 92, 93]. Alternatively, it can be exploited to analyze human motion, for both therapeutic and sport purposes. Lastly, it can also be a powerful tool for video processing and compression, but the fact that most video sequences are recorded with single RGB cameras prevents such application at present.

Nevertheless, scene flow expressed as a dense motion field may be hard to exploit because it provides much more data than most applications are able to process (e.g. if computed for QVGA images, it provides 76800 motion vectors for every aligned pair of images). Consequently, it is necessary to develop algorithms that are able to process and simplify such stream of data in a way that existing technologies can directly benefit from it. Moreover, most scene flow algorithms share two limitations. First, they only work with stereo systems or RGB-D cameras and, hence, scene flow cannot be exploited for standard RGB sequences (or at least, to our knowledge, there does not exist any scene flow algorithm for RGB images). Second, it is computationally very expensive, with runtimes that usually range between several seconds and a few minutes per

UNIVERSIDAD DE MÁLAGA

frame. For these reasons scene flow is still rarely used in robotics or visual&augmented reality, but recent advances suggest that this trend might change.

4.B Representations for Scene Flow

There are different ways to encode scene flow, each of which has distinct advantages and disadvantages. Next, we summarize the most common representations and describe their particular characteristics. [1]

1. Traditionally, scene flow has been expressed as the combination of optical flow $u : \Omega \rightarrow \mathbb{R}^2$ and depth displacement $w : \Omega \rightarrow \mathbb{R}$. This is ideal for those methods that build upon existing optical flow algorithms, and also for stereo-based scene flow where the problem is inherently decoupled into estimating dense correspondences (optical flow) and depth. Furthermore, it leads to a simple and concise data term $E_D(u, w)$ because photometric and geometric consistency can be formulated directly as a function of u and w:

   $$E_D(u,w) = \int_\Omega \|I_2(\boldsymbol{x} + u(\boldsymbol{x})) - I_1(\boldsymbol{x})\| + \mu \, \|Z_2(\boldsymbol{x} + u(\boldsymbol{x})) - Z_1(\boldsymbol{x}) - w(\boldsymbol{x})\| \, \boldsymbol{dx} \, , \tag{4.1}$$

   where μ is a weighting parameter. In general Z can be a depth image (from RGB-D cameras) or a depth field calculated from disparity (from stereo pairs). It must be emphasized that the data term $E_D(u, w)$ shown in (4.1) is a common choice but not the only one: other alternatives or variants of (4.1) have been proposed in the literature.

   On the other hand, this representation provides information about correspondences (endpoints) but not about the trajectory that each 3D point describes between the instants at which the two aligned frames were taken.

2. Alternatively, the motion of the observed points can be represented directly as a 3D displacement field $m : \Omega \rightarrow \mathbb{R}^3$. In this case the data term becomes more complex because it involves the projection of the 3D motion to the image plane Ω:

   $$\begin{aligned} E_D(m) = \int_\Omega & \left\| I_2\left(\pi\left(\pi^{-1}(\boldsymbol{x}, Z_1(\boldsymbol{x})) + m(\boldsymbol{x})\right)\right) - I_1(\boldsymbol{x}) \right\| \\ & + \mu \, \left\| Z_2\left(\pi\left(\pi^{-1}(\boldsymbol{x}, Z_1(\boldsymbol{x})) + m(\boldsymbol{x})\right)\right) - Z_1(\boldsymbol{x}) - m_z(\boldsymbol{x}) \right\| \, \boldsymbol{dx} \, . \end{aligned} \tag{4.2}$$

   As an advantage, regularization of the motion field can be directly applied on m. In the case of scene flow, a regularization term E_R is typically imposed to constrain the estimation problem and smooth the motion field. To that end, such a term would often try to minimize the gradients of the 3D motion vectors, thereby enforcing smoothness of the solution:

   $$E_R(m) = \int_\Omega \|J_m(\boldsymbol{x})\| \, \boldsymbol{dx} \tag{4.3}$$

   where $J_m(\boldsymbol{x})$ refers to the Jacobian of the motion field with respect to $\boldsymbol{x}$. This is not equivalent to minimizing the gradients of the optical and range flows because a constant optical flow

[1] The notation employed in this section was introduced in §3.B and §3.C, please revisit them if you feel unfamiliar with it.

UNIVERSIDAD DE MÁLAGA

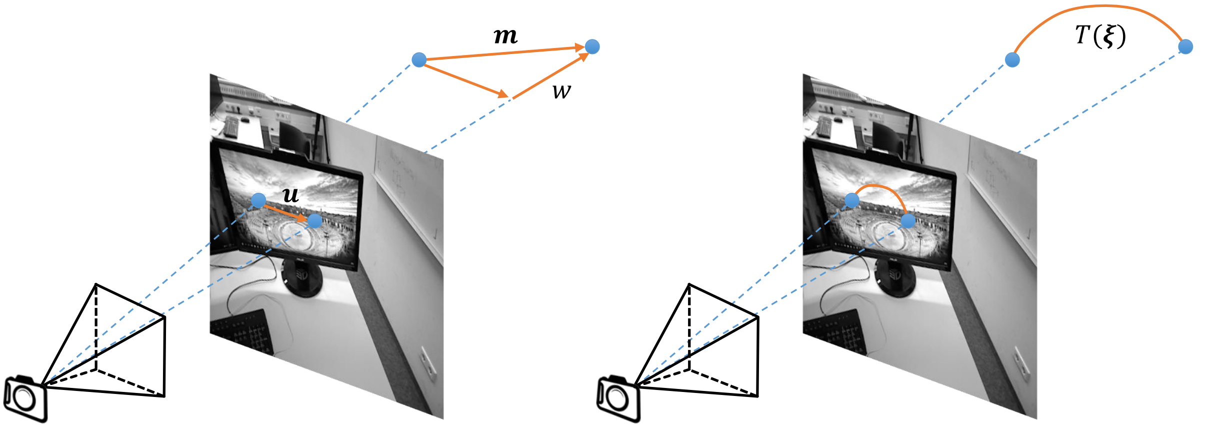

Figure 4.1: Different representations of scene flow[3]. **Left**: Both optical flow + depth shift $(\boldsymbol{u} + w)$ and 3D displacement vector $(\boldsymbol{m})$ provide information only about the final location of each point, but not about its trajectory. **Right**: Encoding scene flow as a rigid body motion makes it possible to recover the trajectory described by every point during the time elapsed between the aligned frames.

(global solution for the regularization term) does not correspond to a constant 3D motion field, and vice versa. In general:

$$E_R(u, w) = \int_\Omega \|J_u(\boldsymbol{x})\| + \mu \|\nabla w(\boldsymbol{x})\| \, \boldsymbol{dx} \neq E_R(m) \tag{4.4}$$

In order to minimize the gradients of the actual motion field using the optical flow-based representation one must include the projection camera model in the regularization term, which complicates the formulation.

Like the optical flow-based formulation, this representation provides information about the final positions of the 3D points of the scene but not about their trajectories.

3. A more elaborate choice consists in overparameterizing the scene flow as a dense field of rigid body motions $\boldsymbol{\xi}$. This parameterization endows each pixel with 6 DoF (in contrast to the 3 DoF of the previous alternatives) and, consequently, requires more constraints to be solved. However, this more complex formulation has some important advantages. First, regularization is more meaningful because rigidity is a more realistic assumption than uniform motion or uniform optical flow:

$$E_R(\xi) = \int_\Omega \|J_\xi(\boldsymbol{x})\| \, \boldsymbol{dx} \neq E_R(m) \neq E_R(u, w) \,. \tag{4.5}$$

Second, estimating the underlying rigid motions of the objects of the scene is a powerful tool that can be exploited for image segmentation. Third, regarding each point as part of a rigid body permits us to compute its 3D trajectory between the two aligned frames (see Figure 4.1).

4.C Contributions

We contribute several methods to estimate scene flow with RGB-D cameras. Next, we provide a summary of the papers where these methods are described.

[3] The camera icon was made by Freepik from www.flaticon.com.

UNIVERSIDAD DE MÁLAGA

In §4.D we present a GPU-based real-time implementation that minimizes an energy function with a primal-dual solver [8]. The data term imposes photometric and geometric consistency between consecutive RGB-D images, and the regularization term enforces smoothness of the motion field. Unlike many existing approaches, which impose regularization on the image plane, we present an alternative strategy to regularize the motion field on the 3D surface of the observed scene, which naturally handles discontinuities of the motion field at the object borders.

A different approach is proposed in §4.E. This paper takes inspiration from [9] and addresses the joint problem of segmenting the scene into the different rigid bodies that compose it and estimating their underlying rigid motions. As a main contribution, we propose a smooth labelling strategy to model the non-rigid motions typically present along the transitions between rigid parts (e.g. in the neck of a person). We incorporate an occlusion mask and include an outlier label to handle those pixels with high residuals for all the motion candidates (those associated to the segments). Results demonstrate that a smooth segmentation based on motion interpolation is more precise than the standard binary alternative, and often leads to a lower number of segments.

Finally, in §4.F we tackle the complex problem of estimating both the camera motion and the scene flow. Here the main difficulty lies in distinguishing static parts of the scene from those which are moving. This step would be straightforward if the camera was still but it becomes challenging when it also moves since all pixels are in apparent motion. We propose to divide the scene into geometric clusters which are, in turn, labelled as static or moving. This strategy also allows us to speed up the scene flow estimation process by two orders of magnitude (if compared to approaches that compute it pixel-wise). Results show that the resulting algorithm provides accurate visual odometry and scene flow with a runtime of 80 milliseconds, which is unprecedented in the literature.

UNIVERSIDAD
DE MÁLAGA

4.D A Primal-Dual Framework for Real-Time Dense RGB-D Scene Flow

Mariano Jaimez, Mohamed Souiai, Javier Gonzalez-Jimenez and Daniel Cremers

Abstract:

This paper presents the first method to compute dense scene flow in real time for RGB-D cameras. It is based on a variational formulation where brightness constancy and geometric consistency are imposed. Depth data provided by RGB-D cameras allows us to impose regularization of the motion field on the 3D surface (or set of surfaces) of the observed scene instead of on the image plane, leading to more geometrically consistent results. The minimization problem is efficiently solved by a primal-dual algorithm which is implemented on a GPU, achieving a previously unseen temporal performance. Several tests have been conducted to compare our approach with a state-of-the-art work (RGB-D flow) where quantitative and qualitative results are evaluated. Moreover, an additional set of experiments have been carried out to show the applicability of our work to estimate motion in real time. Results demonstrate the accuracy of our approach, which outperforms the RGB-D flow, and which is able to estimate heterogeneous and non-rigid motions at a high frame rate.

Published in Proc. International Conference on Robotics and Automation (ICRA), 2015.
DOI: 10.1109/ICRA.2015.7138986

©IEEE (Revised layout)

UNIVERSIDAD DE MÁLAGA

4.E Motion Cooperation: Smooth Piecewise Rigid Scene Flow from RGB-D Images

Mariano Jaimez, Mohamed Souiai, Jörg Stückler, Javier Gonzalez-Jimenez and Daniel Cremers

Abstract:

We propose a novel joint registration and segmentation approach to estimate scene flow from RGB-D images. Instead of assuming the scene to be composed of a number of independent rigidly-moving parts, we use non-binary labels to capture non-rigid deformations at transitions between the rigid parts of the scene. Thus, the velocity of any point can be computed as a linear combination (interpolation) of the estimated rigid motions, which provides better results than traditional sharp piecewise segmentations. Within a variational framework, the smooth segments of the scene and their corresponding rigid velocities are alternately refined until convergence. A K-means-based segmentation is employed as an initialization, and the number of regions is subsequently adapted during the optimization process to capture any arbitrary number of independently moving objects. We evaluate our approach with both synthetic and real RGB-D images that contain varied and large motions. The experiments show that our method estimates the scene flow more accurately than the most recent works in the field, and at the same time provides a meaningful segmentation of the scene based on 3D motion.

Published in Proc. International Conference on 3D Vision (3DV), 2015.
DOI: 10.1109/3DV.2015.15

©IEEE (Revised layout)

UNIVERSIDAD DE MÁLAGA

4.F Fast Odometry and Scene Flow from RGB-D Cameras based on Geometric Clustering

Mariano Jaimez, Christian Kerl, Javier Gonzalez-Jimenez and Daniel Cremers

Abstract:

In this paper we propose an efficient solution to jointly estimate the camera motion and a piecewise-rigid scene flow from an RGB-D sequence. The key idea is to perform a two-fold segmentation of the scene, dividing it into geometric clusters that are, in turn, classified as static or moving elements. Representing the dynamic scene as a set of rigid clusters drastically accelerates the motion estimation, while segmenting it into static and dynamic parts allows us to separate the camera motion (odometry) from the rest of motions observed in the scene. The resulting method robustly and accurately determines the motion of an RGB-D camera in dynamic environments with an average runtime of 80 milliseconds on a multi-core CPU. The code is available for public use/test.

Published in Proc. International Conference on Robotics and Automation (ICRA), 2017.
DOI: 10.1109/ICRA.2017.7989459

©IEEE (Revised layout)

UNIVERSIDAD
DE MÁLAGA

UNIVERSIDAD
DE MÁLAGA

5

3D Reconstruction and Tracking with Subdivision Surfaces

5.A Introduction to 3D Reconstruction

In computer vision, *3D reconstruction* consists in estimating the shape of a given object from one or multiple images in which the object is visible. Depending on the sensor and the configuration used for the reconstruction, the resulting model will have the exact dimensions of the real object (depth or RGB-D cameras [135, 136], laser scanners [137], calibrated RGB cameras [138]) or will keep its proportions with a random or previously-fixed size (monocular RGB cameras [139]). It is worth noting that, when the entity to be modelled is not an object but the environment, the process is referred to as *mapping*. These two commonly-separated topics are, in essence, the same. The main difference lies on the fact that the size and topology of the environment to be mapped are unknown (and probably quite large), whereas those of the object to be reconstructed might be delimited beforehand. For this reason, mapping algorithms must be able to accommodate new incoming data while the sensor explores the environment. On the other hand, reconstruction techniques should be able to segment individual objects and might require higher precision to incorporate fine-grained details.

3D models of objects are useful for many applications. In virtual reality, they allow to insert instances of real objects in virtual scenarios. Even more interesting, if models of people are available they can be used to create Skype-like virtual meetings where each person is represented by an avatar that bears their own true appearance. The same idea has been applied in the gaming industry (see Figure 5.1). In augmented reality, 3D modelling can be used to generate a model of a real object and place it at a given location of the environment, as shown in Figure 5.1. As a different example, 3D reconstruction algorithms can be exploited for reverse engineering, to obtain the dimensions of a product or some components of it. Likewise, it can be used to generate miniatures of people as described in [140].

5.B Introduction to Tracking

Tracking is the process of detecting the location and/or pose of an object throughout a sequence of images. Many algorithms aim at finding the object on an image and placing a bounding box around it. However, we focus on a more complex concept of tracking, where a previously tailored

UNIVERSIDAD DE MÁLAGA

Figure 5.1: Illustrative examples of real applications of 3D modelling[2]. **Left**: Willen Dafoe in the Playstation game "*Beyond: Two Souls*". **Middle**: Hololens users check how a blue couch fits in their living room. **Right**: Miniature figures of former PhD students of the computer vision group at TUM.

model of the object is continuously aligned with the incoming data to know the exact object pose during the whole sequence. That model is typically represented by a 3D mesh composed of triangles or quads, although other alternatives based on smooth spline-like surfaces are recently gaining importance.

Tracking finds numerous applications in computer vision and computer graphics. For instance, body and hand tracking are currently used for human-computer interface and gaming [141, 142]. Face capture and tracking can be employed to animate the facial expressions of an input video, process known as *face reenactment* [143]. From a medical point of view, body tracking can help to diagnose injuries, correct wrong postures and assist during rehabilitation [144]. It could also be exploited by the fashion industry for virtual clothing to promote online sales.

5.C Contributions

Images used for 3D reconstruction or tracking normally observe not only the object to be modelled or tracked but also parts of the environment where this object is present. As a consequence, their pixels must be segmented into two different categories: those from which the object to reconstruct is visible (often called *foreground*) and those which observe other objects of the scene (often referred to as *background*). The foreground pixels contain information that the 3D model must fit, be it colour, position, orientation, etc. The background pixels also impose the restriction that the model should not be visible from them. Our work focuses on this second type of constraints that try to keep the model within the visual hull of the object. To that end, we present a new background term which formulates ray casting as a differentiable energy function. More precisely, this term addresses a min-max problem by first solving ray casting for the background pixels and then deforming the model so that the rays of the background pixels do not intersect it. Aside from that, we describe a complete framework for 3D reconstruction and tracking with subdivision surfaces, and show that the proposed background term can be easily combined with different data terms into an overall optimization problem. Lastly, the experimental section demonstrates that our proposal has several advantages over the distance transform-based term which is commonly employed in the literature.

UNIVERSIDAD DE MÁLAGA

[2] Pictures from `https://blog.es.playstation.com` (left), `www.matrixinception.com` (middle) and [140] (right).

5.D An Efficient Background Term for 3D Reconstruction and Tracking with Smooth Surface Models

Mariano Jaimez, Thomas J. Cashman, Andrew Fitzgibbon,
Javier Gonzalez-Jimenez and Daniel Cremers

Abstract:

We present a novel strategy to shrink and constrain a 3D model, represented as a smooth spline-like surface, within the visual hull of an object observed from one or multiple views. This new "background" or "silhouette" term combines the efficiency of previous approaches based on an image-plane distance transform with the accuracy of formulations based on raycasting or ray potentials. The overall formulation is solved by alternating an inner nonlinear minimization (raycasting) with a joint optimization of the surface geometry, the camera poses and the data correspondences. Experiments on 3D reconstruction and object tracking show that the new formulation corrects several deficiencies of existing approaches, for instance when modelling non-convex shapes. Moreover, our proposal is more robust against defects in the object segmentation and inherently handles the presence of uncertainty in the measurements (e.g. null depth values in images provided by RGB-D cameras).

Published in Proc. Conference on Computer Vision and Pattern Recognition (CVPR), 2017.
DOI: Not available yet

©IEEE (Revised layout)

UNIVERSIDAD
DE MÁLAGA

UNIVERSIDAD
DE MÁLAGA

6 Reactive Navigation

6.A Introduction

Autonomous navigation is becoming a key aspect/technology for modern society. The development of machines, vehicles or robots, that can travel autonomously without the need of human intervention is compelling for multiple reasons. From the point of view of manufacturing, endowing robots with such capability would increase efficiency and flexibility, while probably reducing costs at the same time. Regarding the transport system, autonomous vehicles will represent a paradigm shift. They will improve our quality of life, freeing us from driving and parking, and will reduce the number of traffic accidents (which are mostly caused by human mistakes and negligence). Mobile robots currently allow us to explore and monitor remote places that humans cannot reach: other planets, oceanic trenches, caves, warfare scenarios, etc [35, 166]. Furthermore, mastering autonomous navigation is leading to the emergence of new products and services that were not conceivable before. There already exist robots that can clean the floor of our house [21] or mow the lawn autonomously [167]. There are also projects of robots working in museums or airports as assistants to visitors [68]. These are just a few examples (out of many) of how relevant autonomous navigation is becoming nowadays.

Autonomous navigation encompasses two distinct capabilities. First, it involves motion planning, which consists in finding a route from a departing point to a given destination. To do so, this process requires previous knowledge of the environment, some sort of map. Moreover, it is not only important to find a feasible route but to find the best one (either fastest, safest, etc.), since there are normally many possible ways to go from one location to another. Second, the vehicle or robot should be able to react to the obstacles it encounters during its journey (typically people or other vehicles which are in constant motion and therefore do not appear in maps). Thus, reactive navigation receives data from different sensors and modifies the original plan in real time to avoid collisions. Any advanced algorithm of autonomous navigation is composed of these two blocks, following the so-called *hybrid architecture* [168].

Autonomous navigation can be implemented for different types of robots, and would have different requirements to fulfill in each case. Underwater vehicles might take marine currents into account to increase their range and speed, and drones will need to incorporate more complex 3D reactive strategies to guarantee safe operation [169, 170]. Besides, it is also important to con-

UNIVERSIDAD DE MÁLAGA

sider the mechanical constraints of the robot, and whether these are holonomic or nonholonomic. Holonomic constraints are those that only involve position variables, while nonholonomic constraints involve velocities. For example, planar motion is a simple holonomic constraint while the motion of wheeled vehicles is nonholonomic since the wheels cannot move laterally (assuming there is no slippage).

6.B Contributions

In our work we focus on terrestrial mobile robots, and particularly on those moving on flat surfaces, i.e. with planar motion. In §6.C we present a 3D extension of the reactive navigation algorithm described in [10]. We propose to model the shape of a robot as a set of prisms sorted in height, and group the detected 3D obstacles in the corresponding height bands/levels. Thus, the 3D reactive navigation problem is solved by combining several 2D reactive navigators into a unique space of search where all potential collisions between the robot and the 3D obstacles are taken into account. Both holonomic and nonholonomic constraints are considered by defining path or trajectory families that implicitly fulfill these constraints (as described in [10]). The resulting algorithm is tested with different robotic platforms in various environments. Results demonstrate its effectiveness to drive the robot following a sequence of random destinations in dynamic (and sometimes tight) environments. The autonomous navigation tasks were completed without human intervention in most tests, and the incidents observed where mainly due to unnoticed obstacles or slippage of the wheels, which causes the robot to move in an undesirable way.

UNIVERSIDAD DE MÁLAGA

6.C Efficient Reactive Navigation with Exact Collision Determination for 3D Robot Shapes

Mariano Jaimez, Jose-Luis Blanco and Javier Gonzalez-Jimenez

Abstract:

This paper presents a reactive navigator for wheeled robots moving on a flat surface which takes both the actual 3D shape of the robot and the 3D surrounding obstacles into account. The robot volume is modelled by a number of prisms consecutive in height, and the detected obstacles, which can be provided by different kinds of range sensor, are segmented into these height bands. Then, the reactive navigation problem is tackled by a number of concurrent 2D navigators, one for each prism, which are consistently and efficiently combined to yield an overall solution. Our proposal for each 2D navigator is based on the concept of the "Parameterized Trajectory Generator" which models the robot shape as a polygon and embeds its kinematic constraints into different motion models. Extensive testing has been conducted in office-like and domestic environments, covering a total distance of 18.5 km, to demonstrate the reliability and effectiveness of the proposed method. Moreover, additional experiments are performed to highlight the advantages of a 3D-aware reactive navigator. The code is available under an open-source licence.

Published in the International Journal of Advanced Robotic Systems, 2015.
DOI: 10.5772/60563

©SAGE (Revised layout)

UNIVERSIDAD DE MÁLAGA

UNIVERSIDAD
DE MÁLAGA

7 Conclusions

Conclusions

In this thesis we have addressed distinct problems that lie at the interface between robotics and computer vision. As a common denominator, the proposed methods have exploited the geometric data provided by range sensors for motion estimation, reconstruction, tracking and navigation.

Range sensors have demonstrated to be a powerful alternative to traditional solutions based on monocular or stereo cameras. Knowing the geometry of the environment is often mandatory for many vision-related tasks and, as a consequence, range sensors present a genuine advantage over passive sensory systems. While the former directly provide such data, the latter entail depth estimation as a preliminar step, consuming valuable computational resources in the process. Moreover, range sensors are able to work under poor illumination conditions or even in complete darkness, where RGB cameras render useless. On the other hand, range sensors can only detect objects below a certain distance threshold and are prone to being affected by the solar radiation, which sometimes prevents their use for outdoor applications.

Next, we present individual summaries of the proposed methods, highlighting their pros and cons as well as possible lines of future work.

1. Regarding visual odometry, we have demonstrated that the camera motion can be estimated fast and precisely from depth images. The main drawback of the proposed method is its lack of robustness to moving objects, but this limitation could be overcome by minimizing the geometric residuals within a robust penalty function (as we have proposed in posterior works) or by extending the existing formulation to perform multi-frame alignment. From the point of view of its application, this method could be utilized to estimate the motion of those virtual&augmented reality devices equipped with depth cameras (e.g. Hololens) or, more generally, as the front-end of SLAM systems.

2. Likewise, inspired by the most recent advances in 3D visual odometry, we have developed a direct method based on dense scan alignment for planar odometry. Results show that our method outperforms the most popular techniques on scan-matching while having a lower runtime. Our approach requires scans to be at least piecewise differentiable, which prevents its use in outdoor environments composed of trees, plants and other scattered objects. Nevertheless, this limitation is not that relevant because motion in outdoor scenarios

UNIVERSIDAD DE MÁLAGA

is seldom planar and, hence, odometry based on 2D laser scanners is not really an option. Given its characteristics, this method is suitable for service or telepresence robots operating in office buildings, museums, hotels, homes, etc.

3. We have presented the first real-time algorithm for scene flow estimation with RGB-D cameras. We have made the code public so that all those robots or systems equipped with RGB-D cameras (and NVIDIA GPUs) can exploit it for different purposes. The main limitation of this method is that it works only for small displacements between the incoming frames. However, this is not a serious restriction in practice because it runs in real time and the image pairs to be aligned are very close/similar (save in the presence of very fast object). In order to widen the range of potential applications, this algorithm should be combined with additional post-processing strategies to simplify and refine the extensive information contained in the estimated motion field.

4. A completely different approach for scene flow estimation is described in §4.E. This method achieves very accurate results by jointly segmenting the rigid bodies that form the scene and their underlying rigid body motions. We demonstrate that a smooth motion-based segmentation is beneficial if compared to the standard binary approach when the scene contains non-rigid parts like people, animals, toys or other flexible objects. As a drawback, it is computationally very expensive (20 to 30 seconds running on GPU), which prevents its direct application in real-world scenarios. Given that scene flow is computed from rigid transformations, this method can be used to render "virtual images" for temporal interpolation between the aligned frames. Applying this process to an entire RGB-D sequence would result in a "slow motion" version of the original video.

5. We have tackled an uninvestigated problem: the joint estimation of odometry and scene flow. Our solution relies on a two-fold segmentation of the scene, dividing it into rigid geometric clusters that are, in turn, classified as static or moving elements. Identifying the static parts of the scene is paramount to compute a robust odometry, while the geometric clustering is essential to reduce the computational complexity of the problem. By virtue of this piecewise rigid formulation, our method achieves a runtime of about 80 milliseconds running on CPU, which is some orders of magnitude faster than most existing scene flow algorithms. Despite the promising results, there is still room for improvement since the method fails to distinguish static from moving parts when the former represent a small percentage of the scene ($< 50\%$). Possible solutions to this problem would involve multi-frame alignment or a more elaborate strategy for temporal regularization. This method would be useful for virtually any mobile system that requires some degree of autonomy and operates in a dynamic environment (since it works with RGB-D cameras, it would be constrained to indoor use).

6. We have also presented a new background term to enforce silhouette consistency within 3D reconstruction and tracking systems. This term overcomes important limitations of the popular formulation based on the distance transform, but it comes at the expense of a higher computational cost. We embed this term into an overall optimization framework to fit geometric data (obtained from sets of depth images) with subdivision surfaces. Results show the advantages

UNIVERSIDAD DE MÁLAGA

of this new background term, but the overall framework described for 3D reconstruction and tracking is not mature enough to compete with state-of-the-art algorithms. Concerning the reconstruction stage, one significant difficulty is the lack of an initial topology. In our work, the fitting process always starts with a sphere placed at the centroid of the geometric data, which is subsequently deformed and refined in a coarse-to-fine scheme. The topology of the control mesh is never modified (just refined) during the optimization, fact that renders the reconstruction process extremely complicated. A tentative solution consists in formulating a continuous-discrete optimization strategy which alternates between data fitting and topology rearrangement. This and other alternatives should be explored to improve the quality of the reconstructions.

7. We have generalized an existing reactive navigation algorithm [10] for robots moving on a flat surface. In contrast to most existing approaches, our algorithm takes the 3D shape of the robot into account, as well as the real distribution of obstacles detected by different range sensors. It has been tested with various robotic platforms operating in different environments, allowing robots in MAPIR to cover hundreds of kilometers autonomously. The main limitation of this approach is that it assumes the robot shape is fixed, which is not true for mobile platforms equipped with a manipulator. Therefore, a natural extension of this approach should incorporate the possibility to model changing 3D shapes.

Outlook

We can contend that, nowadays, both autonomous navigation and odometry are solved problems if the environment is static. As Prof. Dieter Fox pointed out in his talk "The 100-100 tracking challenge"[1], the next milestone is achieving the same level of accuracy and robustness in changing environments where robots are surrounded by people or other robots in permanent motion. As a consequence, scene flow, either by itself or combined with segmentation or odometry, will become a more frequent and relevant research topic in robotics and computer vision.

Despite the significant progress made in the last years, including the works presented here, there are still two main problems to solve:

- Scene flow is computationally too expensive. The fact that some algorithms (like the PD-Flow [12] proposed in this thesis) run in real time does not imply that this problem is solved. Those algorithms are fast because they run on powerful GPUs or other dedicated hardware like FPGAs. Most phones, tablets, virtual reality devices or consumer robots cannot afford to be equipped with such hardware, and even if they were, they could not overload it just for scene flow estimation because there are tens of other processes they must run. Therefore, we must find more intelligent strategies to compute scene flow. A positive step forward is the clustering strategy proposed here and in other works like [13]. Estimating motions per cluster greatly reduces the number of unknowns (by two or three orders of magnitude) and often improves accuracy. The next significant improvement might come from studying how to compute the motion of each individual cluster more efficiently.

[1] International Conference on Robotics and Automation (ICRA), Stockholm (Sweden), 2016.

UNIVERSIDAD
DE MÁLAGA

- It is not known how to estimate scene flow with monocular RGB cameras. This is a tremendous limitation since this kind of cameras are cheap, require low power and are equipped in many consumer products. The critical problem is that depth estimation with monocular cameras is performed assuming that disparity between images is explained by the camera motion. When objects move, this hypothesis is violated and there is no obvious way to estimate their 3D position (up to scale) in space. Solutions to this problem must involve the detection of parts of the scene that are not static (high residuals after warping according to the camera motion) and the joint estimation of the position and motion of those elements.

From a technological point of view, in the last decade we have witnessed a spectacular development of range sensors accompanied by a noticeable drop of their prices. In this thesis we have mainly worked with two types of sensors: depth (or RGB-D) cameras and 2D laser scanners. Consciously, and maybe mistakenly, we have disregarded 3D laser scanners. The main reason justifying that decision is that, because of their high prices, our research group does not own such sensor. Admittedly there are datasets with 3D lidar scans and simulators that could be used to generate synthetic data. Yet, during all these years we have focused on working with real sensors, which allowed us to test our algorithms under real conditions, and this was not possible for 3D lidars. Nonetheless, a paradigm shift seems to be near in time because a few manufacturers have announced the future release of solid-state 3D lidars at considerably lower prices than their mirror-based counterparts. This innovation is encouraged by the imminent arrival of autonomous cars and the fact that they rely to a great extent on this kind of sensors to work. As occurred with the advent of Kinect in 2010, robotics might benefit from a technological breakthrough that was not originally conceived for such purpose (although, anyway, an autonomous car is nothing but a mobile robot).

UNIVERSIDAD DE MÁLAGA

Bibliography

[1] Microsoft, “Kinect sensor.” [Online]. Available: https://en.wikipedia.org/wiki/Kinect

[2] A. Segal, D. Haehnel, and S. Thrun, “Generalized-ICP,” in *Proc. of Robotics: Science and Systems (RSS)*, vol. 2, no. 4, 2009.

[3] C. Kerl, J. Sturm, and D. Cremers, “Robust odometry estimation for RGB-D cameras,” in *Int. Conf. on Robotics and Automation (ICRA)*, 2013, pp. 3748–3754.

[4] J. Gonzalez, A. Stentz, and A. Ollero, “A mobile robot iconic position estimator using a radial laser scanner,” *Journal of Intelligent and Robotic Systems*, vol. 13, no. 2, pp. 161–179, 1995.

[5] A. Censi, “An ICP variant using a point-to-line metric,” in *Int. Conf. on Robotics and Automation (ICRA)*, 2008, pp. 19–25.

[6] A. Diosi and L. Kleeman, “Fast laser scan matching using polar coordinates,” *The International Journal of Robotics Research*, vol. 26, no. 10, pp. 1125–1153, 2007.

[7] T. Pock, D. Cremers, H. Bischof, and A. Chambolle, “An algorithm for minimizing the Mumford-Shah functional,” in *Proc. Int. Conference on Computer Vision (ICCV)*, 2009, pp. 1133–1140.

[8] A. Chambolle and T. Pock, “A first-order primal-dual algorithm for convex problems with applications to imaging,” *Journal of Mathematical Imaging and Vision*, vol. 40, no. 1, pp. 120–145, 2011.

[9] D. Cremers and S. Soatto, “Motion competition: A variational approach to piecewise parametric motion segmentation,” *International Journal of Computer Vision*, vol. 62, pp. 249–265, 2005.

[10] J. Blanco, J. Gonzalez-Jimenez, and J. Fernandez-Madrigal, “Extending obstacles avoidance methods through multiple parameter-space transformation,” *Autonomous Robots*, vol. 24, no. 1, pp. 29–48, 2008.

UNIVERSIDAD DE MÁLAGA

[11] Microsoft, “Hololens.” [Online]. Available: https://www.microsoft.com/microsoft-hololens

[12] M. Jaimez, M. Souiai, J. Gonzalez-Jimenez, and D. Cremers, “A primal-dual framework for real-time dense RGB-D scene flow,” in *Int. Conf. on Robotics and Automation (ICRA)*, 2015, pp. 98 – 104.

[13] C. Vogel, K. Schindler, and S. Roth, “Piecewise rigid scene flow,” in *Proc. Int. Conference on Computer Vision (ICCV)*, 2013, pp. 1377–1384.

[14] K. Khoshelham and S. O. Elberink, “Accuracy and resolution of kinect depth data for indoor mapping applications,” *Sensors*, vol. 12, no. 2, pp. 1437–1454, 2012.

[15] Vicon, “Infrared-based motion capture.” [Online]. Available: https://www.vicon.com/products/camera-systems

[16] R. Krigslund, S. Dosen, P. Popovski, J. L. Dideriksen, G. F. Pedersen, and D. Farina, “A novel technology for motion capture using passive UHF RFID tags,” *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 5, pp. 1453–1457, 2013.

[17] Nexonar, “Ultrasonic-based motion capture.” [Online]. Available: http://www.nexonar.com/en/products/beacons/single-beacon/

[18] O. J. Woodman, “An introduction to inertial navigation,” *University of Cambridge, Computer Laboratory, Tech. Rep. UCAMCL-TR-696*, vol. 14, 2007.

[19] J. Engel, J. Sturm, and D. Cremers, “Camera-based navigation of a low-cost quadrocopter,” in *Int. Conf. on Intelligent Robots and Systems (IROS)*, 2012, pp. 2815–2821.

[20] J. Gonzalez-Jimenez, C. Galindo, and J. R. Ruiz-Sarmiento, “Technical improvements of the Giraff telepresence robot based on users’ evaluation,” in *IEEE International Symposium on Robot and Human Interactive Communication*, 2012, pp. 827–832.

[21] iRobot, “Roomba vacuum cleaner.” [Online]. Available: http://www.irobot.com/For-the-Home/Vacuuming/Roomba.aspx

[22] N. Joshi, W. Kienzle, M. Toelle, M. Uyttendaele, and M. F. Cohen, “Real-time hyperlapse creation via optimal frame selection,” *ACM Transactions on Graphics (TOG)*, vol. 34, no. 4, p. 63, 2015.

[23] E. Eade, “Lie groups for computer vision,” 2014. [Online]. Available: http://ethaneade.com/

[24] P. J. Olver, *Applications of Lie groups to differential equations*. Springer Science & Business Media, 2000, vol. 107.

[25] B. K. Horn and B. G. Schunck, “Determining optical flow,” *Artificial intelligence*, vol. 17, no. 1-3, pp. 185–203, 1981.

UNIVERSIDAD DE MÁLAGA

[26] R. Battiti, E. Amaldi, and C. Koch, “Computing optical flow across multiple scales: An adaptive coarse-to-fine strategy,” *International Journal of Computer Vision*, vol. 6, no. 2, pp. 133–145, 1991.

[27] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert, “High accuracy optical flow estimation based on a theory for warping,” in *Proc. European Conference on Computer Vision (ECCV)*, 2004, pp. 25–36.

[28] M. Elad, “On the origin of the bilateral filter and ways to improve it,” *IEEE Transactions on Image Processing*, vol. 11, no. 10, pp. 1141–1151, 2002.

[29] H. Spies, B. Jähne, and J. L. Barron, “Range flow estimation,” *Computer Vision and Image Understanding*, vol. 85, no. 3, pp. 209–231, 2002.

[30] J. Gonzalez-Jimenez and R. Gutierrez, “Direct motion estimation from a range scan sequence,” *Journal of Robotic Systems*, vol. 16, no. 2, pp. 73–80, 1999.

[31] J. Gonzalez, “Recovering motion parameters from a 2D range image sequence,” in *Int. Conf. on Pattern Recognition*, 1996, pp. 433–440.

[32] J. Blanco *et al.* (2017) The Mobile Robot Programming Toolkit (MRPT). [Online]. Available: http://www.mrpt.org/

[33] D. Nistér, O. Naroditsky, and J. Bergen, “Visual odometry,” in *Proc. Int. Conference on Computer Vision and Pattern Recognition (CVPR)*, 2004, pp. 652 – 659.

[34] D. Scaramuzza and F. Fraundorfer, “Visual odometry [tutorial],” *IEEE Robotics and Automation Magazine*, vol. 18, no. 4, pp. 80–92, 2011.

[35] M. Maimone, Y. Cheng, and L. Matthies, “Two years of visual odometry on the Mars exploration rovers,” *Journal of Field Robotics*, vol. 24, no. 3, pp. 169–186, 2007.

[36] K. Konolige, M. Agrawal, and J. Sola, “Large-scale visual odometry for rough terrain,” in *International Symposium on Robotics Research*, 2011, pp. 201–212.

[37] M. A. Fischler and R. C. Bolles, “Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography,” *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[38] J. Engel, J. Sturm, and D. Cremers, “Semi-dense visual odometry for a monocular camera,” in *Proc. Int. Conference on Computer Vision (ICCV)*, 2013, pp. 1449–1456.

[39] P. J. Besl and N. D. McKay, “A method for registration of 3-D shapes,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, p. 239–256, 1992.

[40] D. M. Cole and P. M. Newman, “Using laser range data for 3D SLAM in outdoor environments,” in *Int. Conf. on Robotics and Automation (ICRA)*, 2006, pp. 1556–1563.

UNIVERSIDAD DE MÁLAGA

[41] J. S. Gutmann and K. Konolige, “Incremental mapping of large cyclic environments,” in *Proc. Int. Symposium on Computational Intelligence in Robotics and Automation*, 1999, pp. 318–325.

[42] E. B. Olson, “Real-time correlative scan matching,” in *Int. Conf. on Robotics and Automation (ICRA)*, 2009, pp. 4387–4393.

[43] A. S. Huang, A. Bachrach, P. Henry, M. Krainin, D. Maturana, D. Fox, and N. Roy, “Visual odometry and mapping for autonomous flight using an RGB-D camera,” in *International Symposium on Robotics Research*, 2011, pp. 235–252.

[44] M. Fiala and A. Ufkes, “Visual odometry using 3-dimensional video input,” in *Canadian Conf. on Computer and Robot Vision*, 2011, pp. 86–93.

[45] I. Dryanovski, R. G. Valenti, and J. Xiao, “Fast visual odometry and mapping from RGB-D data,” in *Int. Conf. on Robotics and Automation (ICRA)*, 2013, pp. 2305–2310.

[46] A. I. Comport, E. Malis, and P. Rives, “Accurate quadrifocal tracking for robust 3D visual odometry,” in *Int. Conf. on Robotics and Automation (ICRA)*, 2007, pp. 40–45.

[47] T. Tykkala, C. Audras, and A. I. Comport, “Direct iterative closest point for real-time visual odometry,” in *Proc. Int. Conference on Computer Vision Workshops (ICCV Workshops)*, 2011, p. 2050–2056.

[48] F. Steinbrücker, J. Sturm, and D. Cremers, “Real-time visual odometry from dense RGB-D images,” in *Proc. Int. Conference on Computer Vision Workshops (ICCV Workshops)*, 2011, p. 719–722.

[49] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers, “A benchmark for the evaluation of RGB-D SLAM systems,” in *Int. Conf. on Intelligent Robots and Systems (IROS)*, 2012, p. 573–580.

[50] V. Panwar, “Interest point sampling for range data registration in visual odometry,” Ph.D. dissertation, Queen’s University, Canada, 2011.

[51] R. A. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A. J. Davison, P. Kohi, J. Shotton, S. Hodges, and A. Fitzgibbon, “KinectFusion: Real-time dense surface mapping and tracking,” in *Int. Symp. on Mixed and Augmented Reality (ISMAR)*, 2011, pp. 127–136.

[52] D. R. Canelhas, T. Stoyanov, and A. J. Lilienthal, “SDF tracker: A parallel algorithm for on-line pose estimation and scene reconstruction from depth images,” in *Int. Conf. on Intelligent Robots and Systems (IROS)*, 2013, pp. 3671–3676.

[53] T. Whelan, M. Kaess, M. Fallon, J. Johannsson, H. amd Leonard, and J. McDonald, “Kintinuous: Spatially extended KinectFusion,” in *3rd RSS Workshop on RGB-D: Advanced Reasoning with Depth Cameras*, 2012.

UNIVERSIDAD DE MÁLAGA

[54] T. Whelan, H. Johannsson, M. Kaess, J. J. Leonard, and J. McDonald, “Robust real-time visual odometry for dense RGB-D mapping,” in *Int. Conf. on Robotics and Automation (ICRA)*, 2013, pp. 5724–5731.

[55] G. Guennebaud, B. Jacob *et al.* (2017) Eigen: A C++ template library for linear algebra. [Online]. Available: http://eigen.tuxfamily.org/

[56] N. Gelfand, L. Ikemoto, S. Rusinkiewicz, and M. Levoy, “Geometrically stable sampling for the ICP algorithm,” in *Int. Conf. on 3-D Digital Imaging and Modeling*, 2003, pp. 260–267.

[57] M. Quigley, K. Conley, B. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, and A. Y. Ng, “ROS: an open-source Robot Operating System,” in *ICRA workshop on open source software*, 2009, p. 5.

[58] R. B. Rusu and S. Cousins, “3D is here: Point Cloud Library (PCL),” in *Int. Conf. on Robotics and Automation (ICRA)*, 2011, pp. 1–4.

[59] S. Thrun, D. Fox, W. Burgard, and F. Dellaert, “Robust Monte Carlo localization for mobile robots,” *Artificial intelligence*, vol. 128, no. 1, pp. 99–141, 2001.

[60] J. L. Blanco, J. Gonzalez-Jimenez, and J. A. Fernandez-Madrigal, “Optimal filtering for non-parametric observation models: applications to localization and SLAM,” *The International Journal of Robotics Research*, vol. 29, no. 14, pp. 1726–1742, 2010.

[61] H. Durrant-Whyte and T. Bailey, “Simultaneous localization and mapping: part I,” *IEEE Robotics & Automation Magazine*, vol. 13, no. 2, pp. 99–110, 2006.

[62] A. Diosi and L. Kleeman, “Laser scan matching in polar coordinates with application to SLAM,” in *Int. Conf. on Intelligent Robots and Systems (IROS)*, 2005, pp. 3317 – 3322.

[63] M. Achtelik, A. Bachrach, R. He, S. Prentice, and N. Roy, “Stereo vision and laser odometry for autonomous helicopters in GPS-denied indoor environments,” in *Unmanned Systems Technology XI, SPIE*, 2009, pp. 733 219–733 219.

[64] T. Tomic, K. Schmid, P. Lutz, A. Domel, M. Kassecker, E. Mair, I. L. Grixa, F. Ruess, M. Suppa, and D. Burschka, “Toward a fully autonomous UAV: Research platform for indoor and outdoor urban search and rescue,” *IEEE Robotics and Automation Magazine*, vol. 19, no. 3, pp. 46–56, 2012.

[65] F. Pomerleau, F. Colas, R. Siegwart, and S. Magnenat, “Comparing ICP variants on real-world data sets,” *Autonomous Robots*, vol. 34, no. 3, pp. 133–148, 2013.

[66] M. Jaimez and J. Gonzalez-Jimenez, “Fast visual odometry for 3-D range sensors,” *IEEE Transactions on Robotics*, vol. 31, no. 4, pp. 809–822, 2015.

[67] A. Geiger, P. Lenz, and R. Urtasun, “Are we ready for autonomous driving? the KITTI vision benchmark suite,” in *Proc. Int. Conference on Computer Vision (ICCV)*, 2012, pp. 3354–3361.

UNIVERSIDAD
DE MÁLAGA

[68] R. Triebel, K. Arras, R. Alami, L. Beyer, S. Breuers, R. Chatila, M. Chetouani, D. Cremers, V. Evers, M. Fiore *et al.*, "Spencer: A socially aware service robot for passenger guidance and help in busy airports," in *Field and Service Robotics*, 2016, pp. 607–622.

[69] A. Orlandini, A. Kristoffersson, L. Almquist, P. Björkman, A. Cesta, G. Cortellessa, C. Galindo, J. González-Jiménez, K. Gustafsson, A. Kiselev, A. Loufti, F. Melendez-Fernandez *et al.*, "ExCITE Project: A review of forty-two months of robotic telepresence technology evolution," *Presence: Teleoperators and Virtual Environments*, 2017.

[70] S. Coradeschi, A. Cesta, G. Cortellessa, L. Coraci, C. Galindo, J. González-Jiménez, L. Karlsson *et al.*, *GiraffPlus: A System for Monitoring Activities and Physiological Parameters and Promoting Social Interaction for Elderly*, ser. Human-Computer Systems Interaction: Backgrounds and Applications 3. Springer, 2014, pp. 261–271.

[71] E. Ackerman, "Fetch robotics introduces Fetch and Freight: Your warehouse is now automated," IEEE Spectrum online, 2015.

[72] M. Jaimez, J. G. Monroy, and J. Gonzalez-Jimenez, "Planar odometry from a radial laser scanner. A range flow-based approach," in *Int. Conf. on Robotics and Automation (ICRA)*, 2016, pp. 4479–4485.

[73] P. Biber and W. Strasser, "The Normal Distributions Transform: A new approach to laser scan matching," in *Int. Conf. on Intelligent Robots and Systems (IROS)*, 2003, pp. 2743–2748.

[74] J. D. Fossel, K. Tuyls, and J. Sturm, "2D-SDF-SLAM: A signed distance function based SLAM frontend for laser scanners," in *Int. Conf. on Intelligent Robots and Systems (IROS)*, 2015, pp. 1949–1955.

[75] G. D. Tipaldi and K. O. Arras, "FLIRT - Interest regions for 2D range data," in *Int. Conf. on Robotics and Automation (ICRA)*, 2010, pp. 3616–3622.

[76] F. Kallasi, D. L. Rizzini, and S. Caselli, "Fast keypoint features from laser scanner for robot localization and mapping," *IEEE Robotics and Automation Letters*, vol. 1, no. 1, pp. 176–183, 2016.

[77] F. Lu and E. Milios, "Robot pose estimation in unknown environments by matching 2D range scans," *Journal of Intelligent and Robotic Systems*, vol. 18, no. 3, pp. 249–275, 1997.

[78] J. Minguez, L. Montano, and J. Santos-Victor, "Abstracting vehicle shape and kinematic constraints from obstacle avoidance methods," *Autonomous Robots*, vol. 20, no. 1, pp. 43–59, 2006.

[79] A. W. Fitzgibbon, "Robust registration of 2D and 3D point sets," *Image and Vision Computing*, vol. 21, no. 13, pp. 1145–1153, 2003.

[80] A. Censi, L. Iocchi, and G. Grisetti, "Scan matching in the Hough domain," in *Int. Conf. on Robotics and Automation (ICRA)*, 2005, pp. 2739–2744.

UNIVERSIDAD DE MÁLAGA

[81] A. Censi and R. M. Murray, "Bootstrapping bilinear models of simple vehicles," *The International Journal of Robotics Research*, vol. 34, no. 8, pp. 1087–1113, 2015.

[82] L. Alvarez, C. A. Castano, M. Garcia, K. Krissian, L. Mazorra, A. Salgado, and J. Sanchez, "Symmetric optical flow," in *Int. Conference on Computer Aided Systems Theory*, 2007, pp. 676–683.

[83] C. Zach, "Robust bundle adjustment revisited," in *Proc. European Conference on Computer Vision (ECCV)*, 2014, pp. 772–787.

[84] J. Stückler and S. Behnke, "Integrating depth and color cues for dense multi-resolution scene mapping using RGB-D cameras," in *IEEE Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*, 2012, pp. 162–167.

[85] M. Meilland, A. I. Comport, and P. Rives, "Dense visual mapping of large scale environments for real-time localisation," in *Int. Conf. on Intelligent Robots and Systems (IROS)*, 2011, pp. 4242–4248.

[86] C. Kerl, J. Sturm, and D. Cremers, "Dense visual SLAM for RGB-D cameras," in *Int. Conf. on Intelligent Robots and Systems (IROS)*, 2013, pp. 2100–2106.

[87] A. Howard and N. Roy, "The robotics data set repository (radish)," 2003. [Online]. Available: http://radish.sourceforge.net/

[88] J. R. Ruiz-Sarmiento, C. Galindo, and J. Gonzalez-Jimenez, "Robot@home, a robotic dataset for semantic mapping of home environments," *International Journal of Robotics Research*, vol. 36, no. 2, pp. 131 – 141, 2017.

[89] C. Stachniss, "Robotics datasets." [Online]. Available: http://www2.informatik.uni-freiburg.de/~stachnis/datasets.html

[90] C. Stachniss, G. Grisetti, W. Burgard, and N. Roy, "Analyzing Gaussian proposal distributions for mapping with rao-blackwellized particle filters," in *Int. Conf. on Intelligent Robots and Systems (IROS)*, 2007, pp. 3485–3490.

[91] R. A. Newcombe, D. Fox, and S. M. Seitz, "DynamicFusion: Reconstruction and tracking of non-rigid scenes in real-time," in *Proc. Int. Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 343–352.

[92] M. Innmann, M. Zollhöfer, M. Nießner, C. Theobalt, and M. Stamminger, "VolumeDeform: Real-time volumetric non-rigid reconstruction," *arXiv:1603.08161*, 2016.

[93] M. Slavcheva, M. Baust, D. Cremers, and S. Ilic, "KillingFusion: Non-rigid 3D reconstruction without correspondences," in *Proc. Int. Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[94] A. Bruhn, J. Weickert, and C. Schnörr, "Lucas/Kanade meets Horn/Schunck: Combining local and global optic flow methods," *International Journal of Computer Vision*, vol. 61, no. 3, pp. 211–231, 2005.

UNIVERSIDAD DE MÁLAGA

[95] E. Herbst, X. Ren, and D. Fox, “RGB-D flow: Dense 3-D motion estimation using color and depth,” in *Int. Conf. on Robotics and Automation (ICRA)*, 2013, pp. 2276–2282.

[96] S. Vedula, S. Baker, P. Rander, R. Collins, and T. Kanade, “Three-Dimensional scene flow,” in *Proc. Int. Conference on Computer Vision (ICCV)*, vol. 2, 1999, pp. 722–729.

[97] B. D. Lucas and T. Kanade, “An iterative image registration technique with an application to stereo vision,” in *Proc. Int. Joint Conference on Artificial Intelligence*, 1981, pp. 674–679.

[98] Y. Zhang and C. Kambhamettu, “On 3D scene flow and structure estimation,” in *Proc. Int. Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2, 2001.

[99] T. Basha, Y. Moses, and N. Kiryati, “Multi-view scene flow estimation: A view centered variational approach,” *International Journal of Computer Vision*, vol. 101, no. 1, pp. 6–21, 2013.

[100] F. Huguet and F. Devernay, “A variational method for scene flow estimation from stereo sequences,” in *Proc. Int. Conference on Computer Vision (ICCV)*, 2007, pp. 1–7.

[101] A. Wedel, T. Brox, T. Vaudrey, C. Rabe, U. Franke, and D. Cremers, “Stereoscopic scene flow computation for 3D motion understanding,” *International Journal of Computer Vision*, vol. 95, no. 1, pp. 29–51, 2011.

[102] C. Vogel, K. Schindler, and S. Roth, “3D scene flow estimation with a rigid motion prior,” in *Proc. Int. Conference on Computer Vision (ICCV)*, 2011, pp. 1291–1298.

[103] J. M. Gottfried, J. Fehr, and C. S. Garbe, “Computing range flow from multi-modal Kinect data,” in *Advances in Visual Computing*. Springer, 2011, pp. 758–767.

[104] A. Letouzey, B. Petit, and E. Boyer, “Scene flow from depth and color images,” in *Proc. British Machine Vision Conference (BMVC)*, 2011, pp. 46.1–46.11.

[105] J. Quiroga, F. Devernay, J. L. Crowley *et al.*, “Local/global scene flow estimation,” in *Proc. Int. Conference on Image Processing*, 2013, pp. 3850–3854.

[106] J. Cech, J. Sanchez-Riera, and R. Horaud, “Scene flow estimation by growing correspondence seeds,” in *Proc. Int. Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011, pp. 3129–3136.

[107] S. Hadfield and R. Bowden, “Kinecting the dots: Particle based scene flow from depth sensors,” in *Proc. Int. Conference on Computer Vision (ICCV)*, 2011, pp. 2290–2295.

[108] ——, “Scene particles: Unregularized particle based scene flow estimation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 3, pp. 564–576, 2014.

[109] M. Hornacek, A. Fitzgibbon, and C. Rother, “SphereFlow: 6 DoF scene flow from RGB-D pairs,” in *Proc. Int. Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 3526–3533.

UNIVERSIDAD
DE MÁLAGA

[110] A. Wedel, T. Pock, C. Zach, H. Bischof, and D. Cremers, “An improved algorithm for TV-L1 optical flow,” in *Statistical and Geometrical Approaches to Visual Motion Analysis*. Springer, 2009, pp. 23–45.

[111] Z. Ma, K. He, Y. Wei, J. Sun, and E. Wu, “Constant time weighted median filtering for stereo matching and beyond,” in *Proc. Int. Conference on Computer Vision (ICCV)*, 2013, pp. 49–56.

[112] D. Scharstein and R. Szeliski, “A taxonomy and evaluation of dense two-frame stereo correspondence algorithms,” *International Journal of Computer Vision*, vol. 47, no. 1-3, pp. 7–42, 2002.

[113] H. Hirschmuller and D. Scharstein, “Evaluation of cost functions for stereo matching,” in *Proc. Int. Conference on Computer Vision and Pattern Recognition (CVPR)*, 2007, pp. 1–8.

[114] D. Sun, E. B. Sudderth, and H. Pfister, “Layered RGBD scene flow estimation,” in *Proc. Int. Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.

[115] J. Quiroga, T. Brox, F. Devernay, and J. Crowley, “Dense semi-rigid scene flow estimation from RGBD images,” in *Proc. European Conference on Computer Vision (ECCV)*, 2014, pp. 567–582.

[116] T. Brox, A. Bruhn, and J. Weickert, “Variational motion segmentation with level sets,” in *Proc. European Conference on Computer Vision (ECCV)*, 2006, pp. 471–483.

[117] M. Unger, M. Werlberger, T. Pock, and H. Bischof, “Joint motion estimation and segmentation of complex scenes with label costs and occlusion modeling,” in *Proc. Int. Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012, pp. 1878–1885.

[118] A. Roussos, C. Russell, R. Garg, and L. Agapito, “Dense multibody motion estimation and reconstruction from a handheld camera,” in *Int. Symp. on Mixed and Augmented Reality (ISMAR)*, 2012, pp. 31 – 40.

[119] G. Zhang, J. Jia, and H. Bao, “Simultaneous multi-body stereo and segmentation,” in *Proc. Int. Conference on Computer Vision (ICCV)*, 2011, pp. 826 – 833.

[120] J. Stückler and S. Behnke, “Efficient dense rigid-body motion segmentation and estimation in RGB-D video,” *International Journal of Computer Vision*, vol. 113, no. 3, pp. 233–245, 2015.

[121] T. Pock and A. Chambolle, “Diagonal preconditioning for first order primal-dual algorithms in convex optimization,” in *Proc. Int. Conference on Computer Vision (ICCV)*, 2011.

[122] L. I. Rudin, S. Osher, and E. Fatemi, “Nonlinear total variation based noise removal algorithms,” *Physica D*, vol. 60, pp. 259–268, 1992.

UNIVERSIDAD DE MÁLAGA

[123] T. F. Chan and J. Shen, “Variational image deblurring - a window into mathematical image processing,” *Lecture Note Series, Institute for Mathematical Sciences, National University of Singapore*, 2004.

[124] A. Chambolle, D. Cremers, and T. Pock, “A convex approach to minimal partitions,” *SIAM Journal on Imaging Sciences*, vol. 5, no. 4, pp. 1113–1158, 2012.

[125] M. Nikolova, S. Esedoglu, and T. F. Chan, “Algorithms for finding global minimizers of image segmentation and denoising models,” *SIAM Journal on Applied Mathematics*, vol. 66, no. 5, pp. 1632–1648, 2006.

[126] D. J. Butler, J. Wulff, G. B. Stanley, and M. J. Black, “A naturalistic open source movie for optical flow evaluation,” in *Proc. European Conference on Computer Vision (ECCV)*, ser. Part IV, LNCS 7577, 2012, pp. 611–625.

[127] S. Baker, D. Scharstein, J. Lewis, S. Roth, M. J. Black, and R. Szeliski, “A database and evaluation methodology for optical flow,” *International Journal of Computer Vision*, vol. 92, no. 1, pp. 1–31, 2011.

[128] D. Katz, M. Kazemi, J. A. Bagnell, and A. Stentz, “Interactive segmentation, tracking, and kinematic modeling of unknown 3D articulated objects,” in *Int. Conf. on Robotics and Automation (ICRA)*, 2013, pp. 5003–5010.

[129] D. Gutiérrez-Gómez, W. Mayol-Cuevas, and J. J. Guerrero, “Inverse depth for accurate photometric and geometric error minimisation in RGB-D dense visual odometry,” in *Int. Conf. on Robotics and Automation (ICRA)*, 2015, pp. 83–89.

[130] J. Stückler and S. Behnke, “Integrating depth and color cues for dense multi-resolution scene mapping using RGB-D cameras,” in *Multisensor Fusion and Integration for Intelligent Systems (MFI)*, 2012, pp. 162–167.

[131] M. Jaimez, M. Souiai, J. Stückler, J. Gonzalez-Jimenez, and D. Cremers, “Motion Cooperation: Smooth piece-wise rigid scene flow from RGB-D images,” in *Int. Conf. on 3D Vision (3DV)*, 2015, pp. 64–72.

[132] H. A. Alhaija, A. Sellent, D. Kondermann, and C. Rother, “Graphflow – 6D large displacement scene flow via graph matching,” in *German Conference on Pattern Recognition*, 2015, pp. 285–296.

[133] M. Dou, S. Khamis, Y. Degtyarev, P. Davidson *et al.*, “Fusion4D: real-time performance capture of challenging scenes,” *ACM Transactions on Graphics (TOG)*, vol. 35, no. 4, 2016.

[134] C. Vogel, K. Schindler, and S. Roth, “3D scene flow estimation with a piecewise rigid scene model,” *International Journal of Computer Vision*, vol. 115, no. 1, pp. 1–28, 2015.

[135] S. Izadi, D. Kim, O. Hilliges, D. Molyneaux, R. Newcombe, P. Kohli, J. Shotton, S. Hodges, D. Freeman, A. Davison *et al.*, “KinectFusion: real-time 3D reconstruction and interaction

UNIVERSIDAD DE MÁLAGA

using a moving depth camera," in *Proc. of the 24th annual ACM symposium on User Interface Software and Technology*, 2011, pp. 559–568.

[136] S. Choi, Q.-Y. Zhou, S. Miller, and V. Koltun, "A large dataset of object scans," *arXiv preprint arXiv:1602.02481*, 2016.

[137] J. C. Carr, R. K. Beatson, J. B. Cherrie, T. J. Mitchell, W. R. Fright, B. C. McCallum, and T. R. Evans, "Reconstruction and representation of 3D objects with radial basis functions," in *Proc. of the 28th annual conference on Computer Graphics and Interactive Techniques*, 2001, pp. 67–76.

[138] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, "A comparison and evaluation of multi-view stereo reconstruction algorithms," in *Proc. Int. Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, 2006, pp. 519–528.

[139] P. Ondrúška, P. Kohli, and S. Izadi, "MobileFusion: Real-time volumetric surface reconstruction and dense tracking on mobile phones," *IEEE Transactions on Visualization and Computer Graphics*, vol. 21, no. 11, pp. 1251–1258, 2015.

[140] J. Sturm, E. Bylow, F. Kahl, and D. Cremers, "CopyMe3D: Scanning and printing persons in 3D," in *German Conference on Pattern Recognition*, 2013, pp. 405–414.

[141] J. Shotton, T. Sharp, A. Kipman, A. Fitzgibbon, M. Finocchio, A. Blake, M. Cook, and R. Moore, "Real-time human pose recognition in parts from single depth images," *Communications of the ACM*, vol. 56, no. 1, pp. 116–124, 2013.

[142] J. Taylor, L. Bordeaux, T. Cashman, B. Corish, C. Keskin, T. Sharp, E. Soto, D. Sweeney, J. Valentin, B. Luff *et al.*, "Efficient and precise interactive hand tracking through joint, continuous optimization of pose and correspondences," *ACM Transactions on Graphics (TOG)*, vol. 35, no. 4, p. 143, 2016.

[143] J. Thies, M. Zollhofer, M. Stamminger, C. Theobalt, and M. Nießner, "Face2Face: Real-time face capture and reenactment of RGB videos," in *Proc. Int. Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 2387–2395.

[144] F. A. Moreno, J. A. Merchán-Baeza, M. González-Sánchez, J. González-Jiménez, and A. I. Cuesta-Vargas, "Experimental validation of depth cameras for the parameterization of functional balance of patients in clinical tests," *Sensors*, vol. 17, no. 2, p. 424, 2017.

[145] V. Ganapathi, C. Plagemann, D. Koller, and S. Thrun, "Real-time human pose tracking from range data," in *Proc. European Conference on Computer Vision (ECCV)*, 2012, pp. 738–751.

[146] M. R. Oswald, E. Töppe, and D. Cremers, "Fast and globally optimal single view reconstruction of curved objects," in *Proc. Int. Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012, pp. 534–541.

UNIVERSIDAD
DE MÁLAGA

[147] T. Cashman and A. Fitzgibbon, “What shape are dolphins? Building 3D morphable models from 2D images,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 232–244, 2013.

[148] W. Gander, G. H. Golub, and R. Strebel, “Least-squares fitting of circles and ellipses,” *BIT Numerical Mathematics*, vol. 34, no. 4, pp. 558–578, 1994.

[149] E. Töppe, M. R. Oswald, D. Cremers, and C. Rother, “Image-based 3D modeling via Cheeger sets,” in *Proc. Asian Conference on Computer Vision (ACCV)*, 2010, pp. 53–64.

[150] N. Savinov, L. Ladicky, C. Hane, and M. Pollefeys, “Discrete optimization of ray potentials for semantic 3D reconstruction,” in *Proc. Int. Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 5511–5518.

[151] D. Cremers and K. Kolev, “Multiview stereo and silhouette consistency via convex functionals over convex domains,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 6, pp. 1161–1174, 2011.

[152] M. Prasad, A. Zisserman, and A. W. Fitzgibbon, “Single view reconstruction of curved surfaces,” in *Proc. Int. Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2, 2006, pp. 1345–1354.

[153] S. Vicente and L. Agapito, “Balloon shapes: Reconstructing and deforming objects with volume from images,” in *Int. Conf. on 3D Vision (3DV)*, 2013, pp. 223–230.

[154] J. Taylor, R. Stebbing, V. Ramakrishna, C. Keskin, J. Shotton, S. Izadi, A. Hertzmann, and A. Fitzgibbon, “User-specific hand modeling from monocular depth sequences,” in *Proc. Int. Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 644–651.

[155] H. Hoppe, T. DeRose, T. Duchamp, M. Halstead, H. Jin, J. McDonald, J. Schweitzer, and W. Stuetzle, “Piecewise smooth surface reconstruction,” in *Proc. Annual Conference on Computer Graphics and Interactive Techniques*, 1994, pp. 295–302.

[156] N. Litke, A. Levin, and P. Schröder, “Fitting subdivision surfaces,” in *Proc. Conference on Visualization*, 2001, pp. 319–324.

[157] K.-S. Cheng, W. Wang, H. Qin, K.-Y. Wong, H. Yang, and Y. Liu, “Fitting subdivision surfaces to unorganized point data using SDM,” in *Proc. Pacific Conference on Computer Graphics and Applications*, 2004, pp. 16–24.

[158] S. Ilic, “Using subdivision surfaces for 3-D reconstruction from noisy data,” in *Workshop on Image Registration in Deformable Environments (DEFORM)*, 2006, pp. 1–10.

[159] S. Ivekovic and E. Trucco, “Fitting subdivision surface models to noisy and incomplete 3-D data,” in *Proc. Computer Vision/Computer Graphics Collaboration Techniques: MIRAGE*, 2007, pp. 542–554.

[160] A. Blake and A. Zisserman, *Visual Reconstruction*. Cambridge, USA: MIT Press, 1987.

UNIVERSIDAD DE MÁLAGA

[161] O. Sorkine and M. Alexa, "As-rigid-as-possible surface modeling," in *Symposium on Geometry processing*, vol. 4, 2007.

[162] A. L. Yuille and A. Rangarajan, "The concave-convex procedure (CCCP)," in *Advances in neural information processing systems (NIPS)*, vol. 2, 2002, pp. 1033–1040.

[163] P. D. Tao and L. T. H. An, "Convex analysis approach to DC programming: Theory, algorithms and applications," *Acta Mathematica Vietnamica*, vol. 22, no. 1, pp. 289–355, 1997.

[164] E. Bylow, J. Sturm, C. Kerl, F. Kahl, and D. Cremers, "Real-time camera tracking and 3D reconstruction using signed distance functions." in *Robotics: Science and Systems*, 2013.

[165] R. Maier, J. Stückler, and D. Cremers, "Super-resolution keyframe fusion for 3D modeling with high-quality textures," in *International Conference on 3D Vision (3DV)*, 2015, pp. 536–544.

[166] C. Kunz, C. Murphy, R. Camilli, H. Singh, J. Bailey *et al.*, "Deep sea underwater robotic exploration in the ice-covered arctic ocean with AUVs," in *Int. Conf. on Intelligent Robots and Systems (IROS)*, 2008, pp. 3654–3660.

[167] Husqvarna, "Automower." [Online]. Available: http://www.husqvarna.com/us/products/robotic-lawn-mowers

[168] R. Brooks, "A robust layered control system for a mobile robot," *IEEE Journal on Robotics and Automation*, vol. 2, no. 1, pp. 14–23, 1986.

[169] D. Scaramuzza, M. C. Achtelik, L. Doitsidis, F. Friedrich, E. Kosmatopoulos *et al.*, "Vision-controlled micro flying robots: from system design to autonomous navigation and mapping in GPS-denied environments," *IEEE Robotics & Automation Magazine*, vol. 21, no. 3, pp. 26–40, 2014.

[170] A. Giusti, J. Guzzi, D. Cireşan, F. L. He, J. P. Rodríguez, F. Fontana *et al.*, "A machine learning approach to visual perception of forest trails for mobile robots," *IEEE Robotics and Automation Letters*, vol. 1, no. 2, pp. 661–667, 2016.

[171] R. C. Arkin, *Behavior-based robotics*. MIT press, 1998.

[172] W. Walter, *The living brain*. WW Norton, 1953.

[173] J. Borenstein and Y. Koren, "Real-time obstacle avoidance for fast mobile robots," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 19, no. 5, pp. 1179–1187, 1989.

[174] ——, "The vector field histogram-fast obstacle avoidance for mobile robots," *IEEE Transactions on Robotics and Automation*, vol. 7, no. 3, pp. 278–288, 1991.

[175] L. Tychonievich, D. Zaret, J. Mantegna, R. Evans, E. Muehle, and S. Martin, "A maneuvering-board approach to path planning with moving obstacles," in *Proc. of the International Joint Conference on Artificial intelligence*, vol. 2, 1989, pp. 1017–1021.

UNIVERSIDAD DE MÁLAGA

[176] R. Simmons, “The curvature-velocity method for local obstacle avoidance,” in *Int. Conf. on Robotics and Automation (ICRA)*, vol. 4, 1996, pp. 3375–3382.

[177] D. Fox, W. Burgard, and S. Thrun, “The dynamic window approach to collision avoidance,” *IEEE Robotics and Automation Magazine*, vol. 4, no. 1, pp. 23–33, 1997.

[178] P. Fiorini and Z. Shiller, “Motion planning in dynamic environments using velocity obstacles,” *The International Journal of Robotics Research*, vol. 17, no. 7, pp. 760–772, 1998.

[179] H. Surmann, A. Nüchter, and J. Hertzberg, “An autonomous mobile robot with a 3D laser range finder for 3D exploration and digitalization of indoor environments,” *Robotics and Autonomous Systems*, vol. 45, no. 3, pp. 181–198, 2003.

[180] D. Holz, C. Lorken, and H. Surmann, “Continuous 3D sensing for navigation and SLAM in cluttered and dynamic environments,” in *Proc. International Conference on Information Fusion*, 2008, pp. 1–7.

[181] E. Marder-Eppstein, E. Berger, T. Foote, B. Gerkey, and K. Konolige, “The office marathon: Robust navigation in an indoor office environment,” in *Int. Conf. on Robotics and Automation (ICRA)*, 2010, pp. 300–307.

[182] J. Gonzalez-Jimenez, J. Ruiz-Sarmiento, and C. Galindo, “Improving 2D reactive navigators with Kinect,” in *Proc. International Conference on Informatics in Control, Automation and Robotics (ICINCO)*, 2013, pp. 393–400.

[183] B. Morisset, R. B. Rusu, A. Sundaresan, K. Hauser, M. Agrawal, J. C. Latombe, and M. Beetz, “Leaving Flatland: Toward real-time 3D navigation,” in *Int. Conf. on Robotics and Automation (ICRA)*, 2009, pp. 3786–3793.

[184] K. Nishiwaki, J. Chestnutt, and S. Kagami, “Autonomous navigation of a humanoid robot over unknown rough terrain using a laser range sensor,” *The International Journal of Robotics Research*, vol. 31, no. 11, pp. 1251–1262, 2012.

[185] J. S. Gutmann, M. Fukuchi, and M. Fujita, “3D perception and environment map generation for humanoid robot navigation,” *The International Journal of Robotics Research*, vol. 27, no. 10, pp. 1117–1134, 2008.

[186] R. Kümmerle, R. Triebel, P. Pfaff, and W. Burgard, “Monte Carlo localization in outdoor terrains using multilevel surface maps,” *Journal of Field Robotics*, vol. 25, no. 6-7, pp. 346–359, 2008.

[187] T. Lozano-Perez, “A simple motion-planning algorithm for general robot manipulators,” *IEEE Journal on Robotics and Automation*, vol. 3, no. 3, pp. 224–238, 1987.

[188] J. Minguez and L. Montano, “Nearness diagram (ND) navigation: collision avoidance in troublesome scenarios,” *IEEE Transactions on Robotics and Automation*, vol. 20, no. 1, pp. 45–59, 2004.

UNIVERSIDAD DE MÁLAGA

[189] S. Coradeschi, A. Cesta, G. Cortellessa, L. Coraci, J. Gonzalez, L. Karlsson, F. Furfari, A. Loutfi, A. Orlandini, F. Palumbo *et al.*, “Giraffplus: Combining social interaction and long term monitoring for promoting independent living,” in *Proc. International Conference on Human System Interaction (HSI)*, 2013, pp. 578–585.

UNIVERSIDAD DE MÁLAGA