Actas de las XXXIX Jornadas de Automática, Badajoz, 5-7 de Septiembre de 2018
https://doi.org/10.17979/spudc.9788497497565.0822

IDENTIFICACION, NAVEGACION E INTERACCION HUMANO-ROBOT EN ENTORNOS PARCIALMENTE ESTRUCTURADOS MEDIANTE FUSIÓN SENSORIAL

E. M. Ortega Vázquez, emortega@ujaen.es; J. A. Moral Parras, joseantoniomoralparras@gmail.com;
E. Estevez, eestevez@ujaen.es; J. Gómez Ortega, juango@ujaen.es; J. Gámez García, jggarcia@ujaen.es

Grupo de Robótica, Automática y Visión por Computador (GRAV),
Universidad de Jaén, Jaén, 23071 España

Resumen

La creación de sistemas inteligentes conscientes de un entorno parcialmente desconocido o en su totalidad ha sido objeto de estudio en una gran cantidad de investigaciones. Además, la fusión sensorial se utiliza en una gran cantidad de tareas robóticas, como localización, construcción de mapas, transporte de mercancías, interacción con el entorno, inspección, etc... En el presente documento, se propone un caso de estudio donde se integran diferentes sensores y robots para llevar a cabo tareas de detección e identificación de personas en conjunción con herramientas de mapeado, e interacción persona-máquina, cuyos sistemas estén interconectados entre sí mediante el Middleware de código libre ROS. Para ello, se hace uso de diferentes técnicas para el posicionamiento y detección de humanos, navegación autónoma, así como su identificación a corta distancia.

Palabras clave: Navegación autónoma, fusión sensorial, interacción hombre-robot, TOF, LIDAR, clasificador de cascada Haar, ROS, FNN, Kinect, IMU

1 INTRODUCCIÓN

Uno de los mayores requerimientos en las interacciones hombre-robot es que el robot se comporte de forma inteligente, siendo consciente del entorno que le rodea, tanto obstáculos como personas [1].

Por otro lado, la navegación autónoma ha sido resuelta en múltiples ocasiones utilizando diversos sensores, como LIDAR (*Light Detection and ranging*) [2] o visión por computador [3], basados en la construcción del mapa del entorno y posterior navegación sobre él. La navegación autónoma puede estar basada en aprendizaje supervisado, en el caso de Gao *et al*.[4], donde compara el trazado de trayectorias tradicional o utilizando una red neuronal profunda para el mismo.

El objetivo que se pretende en este caso es crear un sistema inteligente consciente de su entorno parcialmente desconocido y capaz de actuar en él, orientado a la interacción con seres humanos. Para tal fin, se ha estudiado detectar a personas mediante fusión sensorial y ser capaz de posicionarlas de forma precisa en dicho entorno. Posteriormente, el robot se traslada a las cercanías del individuo, para interactuar con él, con lo que debe contar con navegación autónoma y detección de obstáculos. Finalmente, una identificación de la persona gracias a sus características fisionómicas ha sido integrada, lo cual permite identificar a la persona con visión nula o parcial de su rostro. Algunos usos de este sistema podrían ser atención personalizada o guías robotizadas para personas discapacitadas.

La mayor prioridad es automatizar todo el proceso, especialmente la calibración de los elementos sensores, en este caso, visión artificial fija en el entorno donde se pretende detectar la persona.

El resto del documento se estructura como sigue. La sección 2 describe el sistema y se definen las funciones desarrolladas para abarcar los problemas que comprende llevar a cabo el objetivo descrito anteriormente. La sección 3 describe la combinación de todos los sistemas, así como los resultados obtenidos. Finalmente, una sección de conclusiones.

2 MATERIALES Y MÉTODOS

Con el motivo de integrar movilidad junto con una mayor capacidad de interacción, se acopla la base móvil autónoma PMB2 y el robot humanoide Meka con dos brazos de 7 DOF (*Degrees Of Freedom*).

Para implementar y conectar ambos robots entre sí, se ha utilizado la plataforma ROS (*Robot Operating System*) [5], ya que se trata de una infraestructura flexible y abierta pensada para programar el software de cualquier robot, la cual funciona como middleware de comunicación entre todos los dispositivos.

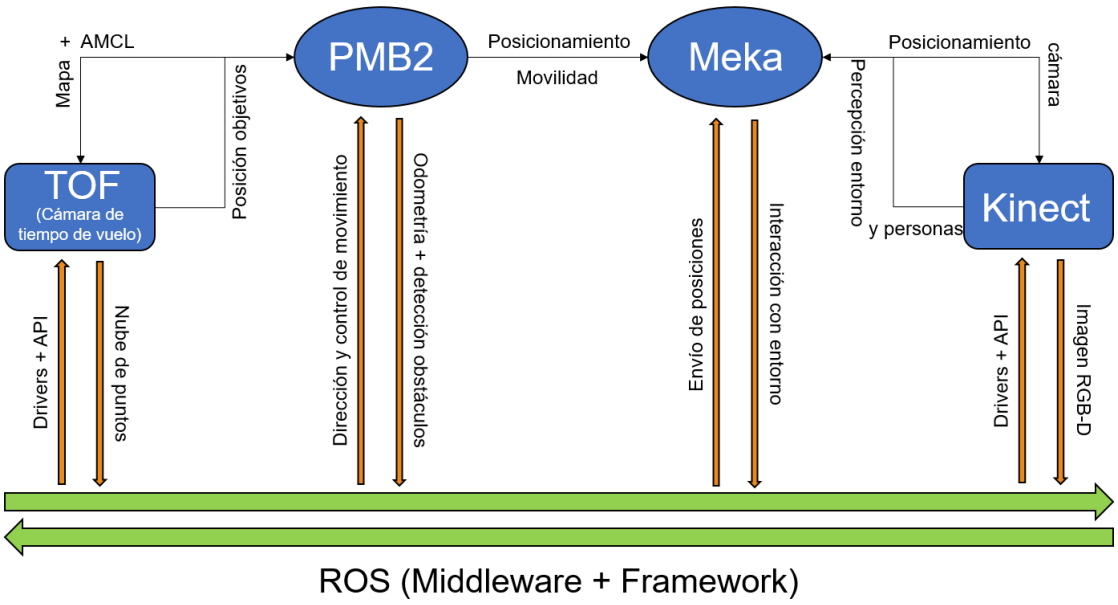


Figura 1: Arquitectura del sistema conectado e implementación presentada en el artículo

También permite hacer una unión precisa en el entorno virtual, para las transformaciones vectoriales entre los diferentes objetos, haciendo uso de los modelos URDF (*Unified Robot Description Format*) y la herramienta TF [6].

En los siguientes subapartados se plantean los diferentes problemas y su resolución, cuyo esquema general viene reflejado en la figura 1. El sistema se basa en cuatro sistemas principales: cámara TOF que ofrece posición de personas, base autónoma PMB2 que mapea el entorno desconocido sobre el que se referencia el sistema y da movilidad al Meka (permite interacción) y la cámara RGB-D Kinect, que percibe el entorno desde el punto de vista del robot y, además, identifica las personas.

2.1 NAVEGACIÓN AUTÓNOMA

El robot PMB2 es un robot diferencial de dos ruedas que integra sensor LIDAR (*Light Detection and Ranging*) que permite trazar la distancia entre el emisor y el objeto no reflectante/transparente en dos dimensiones, IMU (*Inertial Measurement Unit*) y tracción por ruedas controladas, por lo que incluye el *hardware* necesario para realizar navegación autónoma.

Con el fin de posicionar todos los sistemas en un entorno ciberfísico, se necesita crear un mapa del entorno desconocido. Partiendo de las cinemáticas inversas de este tipo de robot [7] y su odometría, junto con el LIDAR, se puede mapear el entorno y localizarse haciendo uso de la técnica SLAM (*Simultaneous Localization and Mapping*) [8]. La navegación autónoma a partir del mapa construido, se resume la figura 2, el cual hace uso del algoritmo de localización ACML (*Adaptative Monte Carlo Localization*), un sistema de localización probabilístico descrito en [9], basado en un filtro de partículas: primero, inicia con una distribución de partículas aleatorio y distribuido en el mapa, posteriormente se filtran según los estados probables que detecte con sensores.

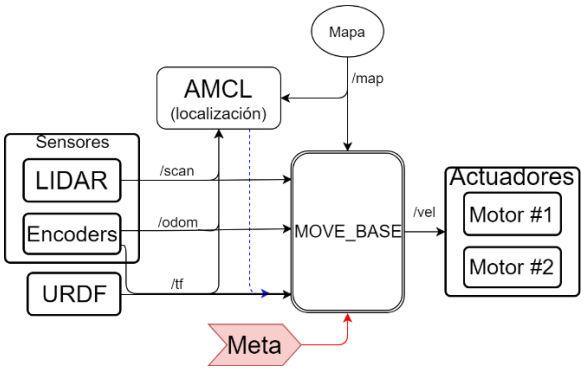


Figura 2: Navegación autónoma en PMB2

La navegación autónoma utilizada hace uso de mapas de inflado o también llamado mapa de coste [10], a partir del mapa reconstruido mediante SLAM, separando el problema en navegación global y local visibles en la figura 3. La navegación local con DWA (*Dynamic Window Approach*) es la que permite evitar colisiones, así como una serie de obstáculos, en la medida de la capacidad del robot [11], mientras que la navegación global permite realizar recorridos entre puntos a larga distancia.

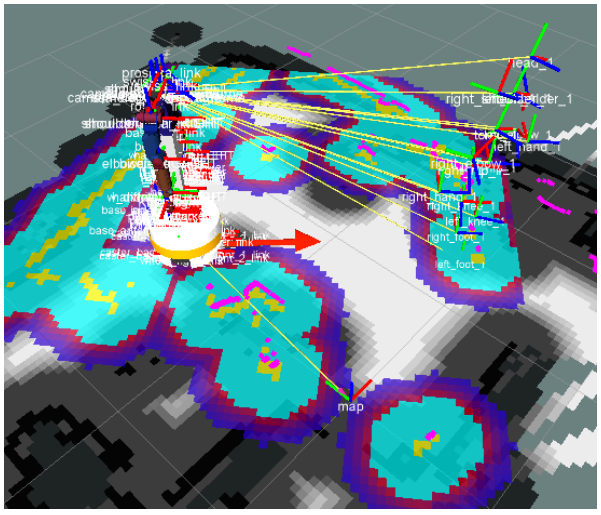


Figura 3: Mapas de coste global (gris) y local (azul), representación de los robots PMB2, Meka y coordenadas de la esqueletización de un humano por Kinect, referenciado al mapa

2.2 DETECCIÓN Y LOCALIZACIÓN DE PERSONAS EN ESCENA

El sensor utilizado para detectar y localizar personas en una región determinada de la escena consiste en un sistema de visión artificial que además obtiene información de profundidad del entorno. Por lo tanto, se utiliza la cámara TOF (*Time Of Flight*) [12], capaz de proporcionar información tridimensional del entorno en el que opera. En este caso, se integra la cámara de tiempo de vuelo de emisión de luz activa SwissRanger 4000 de Mesa Imaging o SR4000, con resolución de 176x144 píxeles, longitud de onda de 850nm, 54FPS y precisión de ± 1cm.

El problema se divide en los dos siguientes subapartados, en uno se describe el proceso de referencia de la cámara en el sistema y en el otro la detección de las personas.

2.2.1 Referencia de la cámara 3D en el sistema

Referenciar correctamente todos los elementos en el espacio, tanto posición como orientación (6D: 3 ángulos de Euler y sistema cartesiano) es de gran importancia en robótica para establecer ubicaciones de actuadores y sensores [13], tanto posición absoluta como relativa, por tanto, si se sitúa correctamente la cámara en el mapa, se puede referenciar de forma absoluta la posición de una persona. La metodología empleada es la de resolver cada una de las seis incógnitas anteriormente mencionadas, para situar la cámara en el mapa trazado por el robot PMB2 (figura 4).

Para resolver dos de los tres ángulos de Euler, se realiza una fusión sensorial colocando un IMU de 9 DOF con giroscopio en dicha cámara. Esto permite conocer la inclinación en los ángulos β y α, sin embargo, el ángulo γ se obtiene mediante brújula terrestre, menos preciso y más complejo de automatizar en el proceso de mapeo. Por lo tanto, se propone obtener el ángulo γ por un método diferente a la orientación magnética, descrito más adelante.

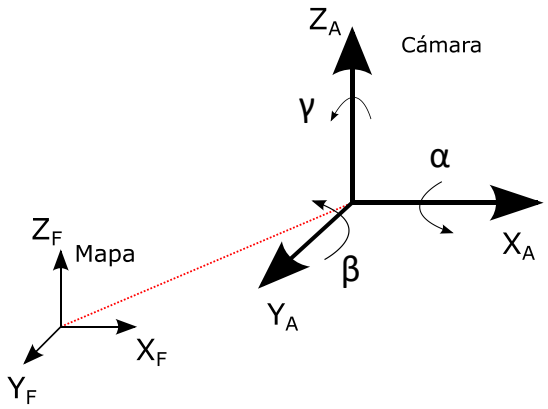


Figura 4: Ejes y ángulos de referencia para posicionar la cámara respecto al mapa

Con motivo de ubicar las coordenadas X, Y y γ, se ha propuesto el uso del algoritmo de localización AMCL [9], equiparando la nube de puntos tridimensionales que ofrece la cámara TOF con los puntos bidimensionales proporcionados por el LIDAR de la base móvil autónoma. Para reducir dimensionalidad, se escoge un plano de la misma altura de la nube de puntos . Simulando la integración SR4000+IMU como un robot móvil y utilizando el mapa creado en la etapa anterior, es posible realizar una autolocalización del sistema (figura 5), con lo que se permite una calibración prácticamente automática. Al igual que la autolocalización con AMCL en un robot, es necesario indicar manualmente un punto inicial cercano al posible para agilizar el proceso de filtrado de partículas, e igualmente, después permite una corrección automática realizando movimientos de la propia cámara (en el ángulo γ).

La coordenada Z y última incógnita, es altura respecto al suelo en el sistema cartesiano y queda determinada mediante diseño/medición del soporte de la cámara, permitiendo autolocalización con tolerancias de dicha medida de ± 5mm.

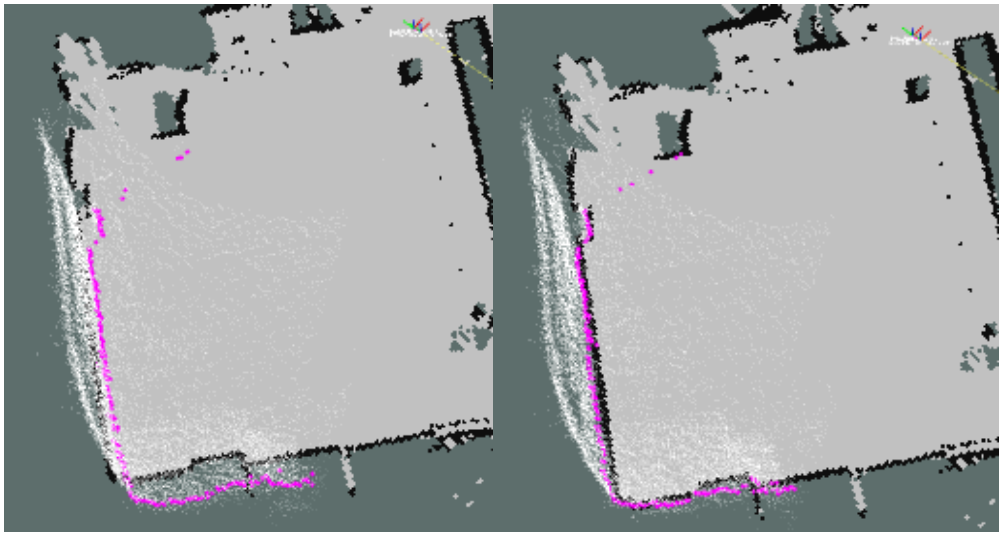


Figura 5: Primer posicionamiento (izquierda) y autoposicionamiento final (derecha). Nube de puntos (blanco), posición posible de la cámara (esquina superior derecha) y elección de puntos láser en un plano (rosa)

2.2.2 Detección de personas

Tras referenciar la cámara 3D correctamente en el entorno, debe cumplir la función de la detección de personas.

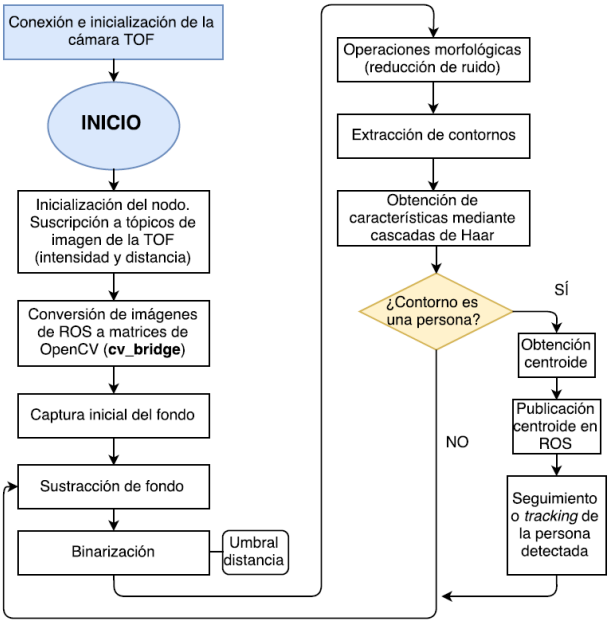


Figura 6: Algoritmo de detección de personas

La metodología propuesta (figura 6) y validada consiste en la de obtener una imagen del entorno y sustraerla, para posteriormente binarizar mediante un umbral global predefinido de distancia. Tras la binarización, se obtiene cualquier objeto no presente anteriormente, incluyendo ruido, por lo que se filtra mediante operación morfológica de cierre (Ec. 1). El algoritmo de decisión que discrimina entre personas u otros objetos se detalla en el siguiente apartado.

$$A \cdot B = (A \oplus B) \ominus B \tag{1}$$

Para posicionar a la persona, se obtuvo un punto de referencia arbitrario de la silueta bidimensional detectada por la cámara TOF, en este caso, se eligió el centroide de la misma. Considerando área "A" y el peso distribuido uniformemente, el centroide se calcula con la Ec. 2:

$$c_x = \frac{1}{6A}\sum_{i=1}^{n}(x_i + x_{i+1})(x_i y_{i+1} - x_{i+1} y_i)$$
$$c_y = \frac{1}{6A}\sum_{i=1}^{n}(y_i + y_{i+1})(x_i y_{i+1} - x_{i+1} y_i) \tag{2}$$

Después, se calcularon las coordenadas de dicho punto respecto a la cámara, transformándose en coordenadas para la posterior utilización de los mismos por el robot móvil PMB2. Para que el robot sea capaz de navegar hasta la persona, debe conocerse la ubicación de la cámara de tiempo de vuelo en el mapa, tarea cumplida por la etapa previa de calibración.

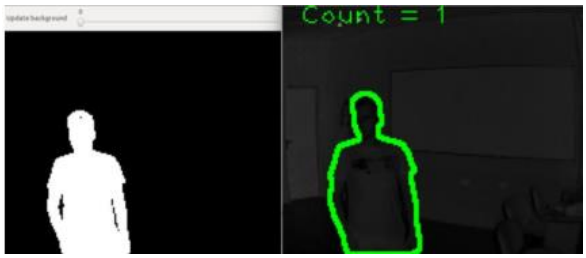


Figura 7: Binarización y filtrado de contorno (izquierda), validación de persona y numeración (derecha)

2.2.3 Algoritmo de decisión

En el procesamiento de imágenes anterior se utiliza un umbral de distancia predefinido para distancias de trabajo indicadas, pero configurables (los resultados utilizan r = 0.8 para detecciones de 1 a 8 metros de distancia).

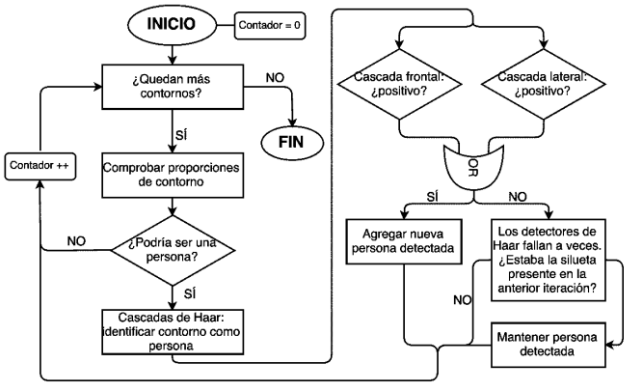


Figura 8: Algoritmo de decisión que discrimina el contorno es una persona

Los contornos extraídos pasan previamente por un filtrado de proporciones, principalmente para evitar falsos positivos detectando la plataforma robótica como un humano. En este caso, una persona tiene un ratio altura/anchura mayor que el robot, pero dicho ratio debe variar con la distancia tridimensional de la cámara al contorno, con lo que se propone la Ec. 3:

$$\boldsymbol{R = d * r} \qquad (3)$$

donde "r" es un ratio de aspecto fijo, establecido mediante mediciones previas; "d" es la distancia de la cámara al centroide y se obtiene el ratio final "R", con un valor mayor para potenciales humanos.

Finalmente, se utilizan paralelamente dos clasificadores supervisados en cascada Haar usualmente utilizados para el reconocimiento de caras [12], entrenados cada uno en posiciones frontales y laterales de personas, como se detalla en la figura 8. Dicho algoritmo se aplica a tantos contornos como se detecten, descrito en el subapartado anterior, por lo que permite distinguir personas que no se solapen entre sí, comparando con la imagen de la iteración anterior.

Al permitir un rastreo de la persona en sucesivas iteraciones, permite calcular la velocidad lineal entre dichas posiciones, conociendo así, si una persona está en movimiento o se encuentra parada.

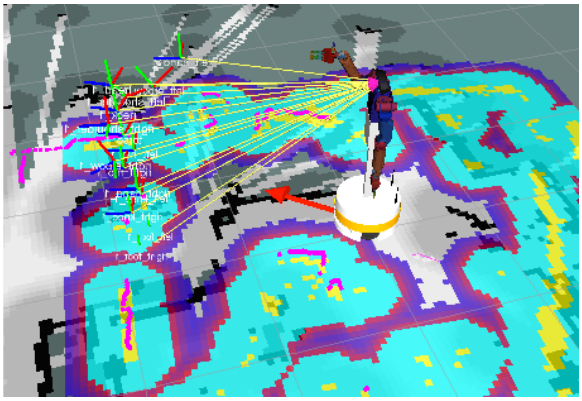
2.3 IDENTIFICACIÓN DE PERSONAS

La cámara situada en el entorno tiene baja resolución para la identificación de personas, aunque cuenta con mayor alcance de funcionamiento útil para la detección de personas en un entorno, por lo que se integra otra cámara RGB-D en el robot móvil. Para el presente documento, se utiliza la cámara de bajo coste Microsoft Kinect, que proporciona imágenes de alta resolución en color y profundidad. Gran cantidad de algoritmos para hacer seguimiento de objetos y reconocimiento de actividades humanas han sido desarrollados [14]. Se ha utilizado OpenNI (*Open Natural Interaction*) junto con el software NITE para realizar el seguimiento del esqueleto (estructura morfológica, figura 3) de personas en tiempo real, también en uso, por ejemplo, para el reconocimiento de gestos en [15] o incluso en aplicaciones médicas [16]. En un conjunto de datos limitado, dicha estructura es característica y diferente entre personas, lo que permite identificarlas independientemente de una posible visión reducida del rostro.

En la tarea de identificación de personas, se utiliza un clasificador supervisado FNN (red neuronal prealimentada) [17], donde se aplican las distancias características de la fisionomía como datos para las neuronas de entrada, resultando como salida la identificación de la persona. El software con el que se ha implementado la FNN es PyBrain, incluido en Scikit-learn [18]. Si en la toma de datos se inscribe el nombre, facilita la interacción humano-máquina en gran medida, ya que permite al sistema dirigirse a la persona por su nombre, una vez identificada.

2.4 INTERACCIÓN HOMBRE-MÁQUINA

Gracias la localización, identificación y el seguimiento de las extremidades de personas, es posible realizar una interacción hombre-máquina haciendo uso de la movilidad y flexibilidad que la plataforma robótica posee. Múltiples robots han utilizado conversiones de texto a voz artificial para crear una interacción más natural con humanos [19]. Como prueba de contacto para testear funcionalidades, se probó que el robot imitara los gestos de la persona con su brazo, haciendo mímica de un posible saludo gracias al seguimiento de las articulaciones proporcionadas por la cámara RGB-D, como se puede observar en la figura 9.



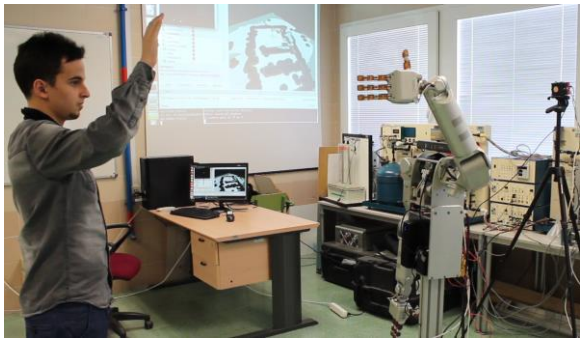


Figura 9: Imagen superior (representación) e inferior (real) del ejemplo de interacción imitando movimiento de la mano tras la navegación (humano, robot y TOF de izquierda a derecha en la inferior)

3 RESULTADOS DE LA INTEGRACIÓN DEL SISTEMA

El objetivo principal engloba muchas tareas de alto nivel de programación, como son la navegación autónoma, detección, posicionamiento e identificación de personas, así como el movimiento de un brazo robótico. Para gestionar procesos complejos que involucren el accionamiento de muchos sistemas en conjunción que sigan un plan determinado se definen máquinas de estados finitos. Al ejecutar un proceso, se pueden definir todos los estados posibles y sus transiciones explícitamente, desgranando el proceso en tareas sencillas y ofreciendo modularidad, con lo que se puede predecir todos los estados posibles y programar cómo debe reaccionar el sistema.

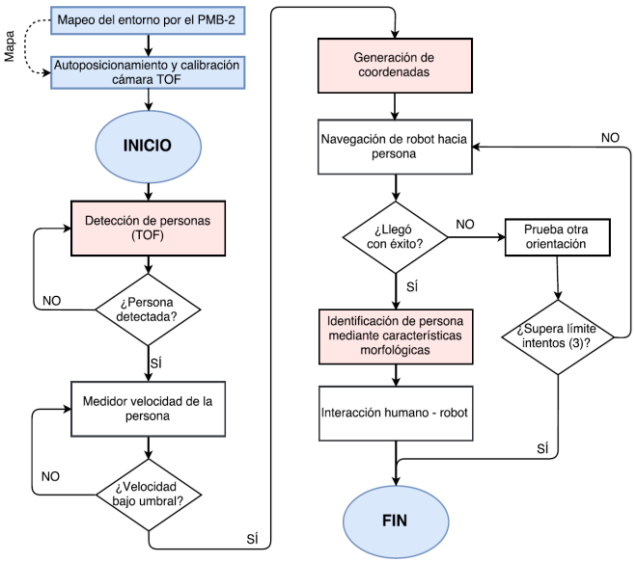


Figura 10: Diagrama de flujo de una posible combinación de los subsistemas y sus funciones

El caso de ejemplo incluye desde la toma de contacto con el entorno, su mapeo, autocalibración de la cámara TOF, aprendizaje para la identificación entre personas y puesta en marcha.

Por lo tanto, detectará a una persona mediante la cámara RS4000, la posiciona en el entorno y espera a que se detenga. Una vez detenido el individuo, envía posiciones plausibles alrededor de él, con lo que el robot avanzará mediante la navegación autónoma, para finalizar con una interacción sencilla entre el humano y el robot, guiado mediante la voz artificial (figura 10).

Los resultados de fases intermedias se pueden verificar (figura 11), como la detección de personas y su posicionamiento. Como se mencionó, se hace un muestreo a lo largo del tiempo, con lo que posibilita la medición de la velocidad de cada persona de forma independiente.

Figura 11: Detección de dos personas y su posicionamiento en el espacio

Por otro lado, la autocalibración de la cámara TOF (figura 5) respecto del entorno haciendo uso del mapa es correcta, pero tiene el defecto de necesitar una esquina o geometría característica, por lo que no es posible en todo tipo de entornos, especialmente en entornos abiertos y sin referencias.

4 CONCLUSIONES

El desarrollo de la robótica para la interacción con personas es de vital importancia para el avance de la automatización. El estudio de técnicas de movilidad autónoma permite la adaptación a entornos parcialmente estructurados. El caso de estudio se ha podido llevar a cabo de forma satisfactoria, minimizando los posibles errores de detección y posicionamiento de humanos en las fases de desarrollo. El proceso de autolocalización de la cámara se ha producido con éxito gracias a la fusión sensorial y especialmente a la capacidad de construcción de mapas que provee SLAM.

La identificación de personas ofrece buenos resultados en una población controlada. Una posterior interacción humana se ha llevado a cabo dotando de una mayor inteligencia gracias a la identificación, con lo que en el futuro se podrían cambiar patrones de comportamiento entre diferentes usuarios.

Agradecimientos

Este trabajo ha sido parcialmente respaldado por el proyecto DPI2016-78290-R.

English summary

IDENTIFICATION, NAVIGATION AND HUMAN-ROBOT INTERACTION IN PARTIALLY STRUCTURED ENVIRONMENTS USING SENSOR FUSION

Abstract

Creating and setting up smart systems aware of a partially or totally unknown environment has been studied in a large amount of research. In addition, sensor fusion is used in a large number of robotic tasks, such as location, map construction, interaction with the environment, quality inspection, etc… In this paper, a case study has been proposed, where different sensors have been integrated along robots in order to carry out tasks of detection and identification of people in conjunction with tools of mapping and human-robot interaction, whose systems are interconnected with each other using the open source middleware ROS. To achieve this, different techniques have been used for positioning and detection of humans, autonomous navigation as well as their identification at close range.

Keywords: Autonomous navigation, sensor fusion, human-robot interaction, TOF, LIDAR, Haar cascade classifier, ROS, FNN, Kinect, IMU

Referencias

[1] P. Althaus, H. Ishiguro, T. Kanda, T. Miyashita, and H. I. Christensen, "Navigation for human-robot interaction tasks," in *IEEE International Conference on Robotics and Automation, 2004. Proceedings. ICRA '04. 2004*, 2004, vol. 2, p. 1894–1900 Vol.2.

[2] J. Guivant, E. Nebot, and S. Baiker, "Autonomous navigation and map building using laser range sensors in outdoor applications," *J. Robot. Syst.*, vol. 17, no. 10, pp. 565–583, 2000.

[3] K. Konolige, M. Agrawal, R. C. Bolles, C. Cowan, M. Fischler, and B. Gerkey, "Outdoor Mapping and Navigation Using Stereo Vision," in *Experimental Robotics*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 179–190.

[4] W. Gao, D. Hsu, W. S. Lee, S. Shen, and K. Subramanian, "Intention-Net: Integrating Planning and Deep Learning for Goal-Directed Autonomous Navigation," no. Figure 1, pp. 1–10, Oct. 2017.

[5] "ROS," 2018. [Online]. Available: http://www.ros.org/.

[6] T. Foote, "tf: The Transform Library," *Open Source Robot. Found.*

[7] Y. Zhao and S. L. BeMent, "Kinematics, dynamics and control of wheeled mobile robots," *Proc. 1992 IEEE Int. Conf. Robot. Autom.*, pp. 91–96, 1992.

[8] M. W. M. G. Dissanayake, P. Newman, S. Clark, H. F. Durrant-Whyte, and M. Csorba, "A solution to the simultaneous localization and map building (SLAM) problem," *IEEE Trans. Robot. Autom.*, vol. 17, no. 3, pp. 229–241, Jun. 2001.

[9] D. Fox, W. Burgard, F. Dellaert, and S. Thrun, "Monte Carlo Localization: Efficient Position Estimation for Mobile Robots," *Proc. Natl. Conf. Artif. Intell.*, no. Handschin 1970, pp. 343–349, 1999.

[10] E. Marder-Eppstein, E. Berger, T. Foote, B. Gerkey, and K. Konolige, "The Office Marathon: Robust navigation in an indoor office environment," in *2010 IEEE International Conference on Robotics and Automation*, 2010, vol. 11, no. 1, pp. 300–307.

[11] D. Fox, W. Burgard, and S. Thrun, "The dynamic window approach to collision avoidance," *IEEE Robot. Autom. Mag.*, vol. 4, no. 1, pp. 23–33, Mar. 1997.

[12] L. Li, "Time of Flight Camera: An Introduction," *Texas Instruments - Tech. White Pap.*, no. January, p. 10, 2014.

[13] R. Chatila and J. Laumond, "Position referencing and consistent world modeling for mobile robots," in *Proceedings. 1985 IEEE International Conference on Robotics and Automation*, 1985, vol. 2, pp. 138–145.

[14] Jungong Han, Ling Shao, Dong Xu, and J. Shotton, "Enhanced Computer Vision With Microsoft Kinect Sensor: A Review," *IEEE Trans. Cybern.*, vol. 43, no. 5, pp. 1318–1334, Oct. 2013.

[15] B. Wang, Z. Chen, and J. Chen, "Gesture Recognition by Using Kinect Skeleton Tracking System," in *2013 5th International Conference on Intelligent Human-Machine Systems and Cybernetics*, 2013, pp. 418–422.

[16] S. Motiian, P. Pergami, K. Guffey, C. A. Mancinelli, and G. Doretto, "Automated extraction and validation of children's gait parameters with the Kinect," *Biomed. Eng. Online*, vol. 14, no. 1, p. 112, Dec. 2015.

[17] D. Svozil, V. Kvasnicka, and J. Pospichal, “Introduction to multi-layer feed-forward neural networks,” *Chemom. Intell. Lab. Syst.*, vol. 39, no. 1, pp. 43–62, Nov. 1997.
[18] F. Pedregosa *et al.*, “Scikit-learn: Machine Learning in Python,” *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2012.
[19] M. Salichs *et al.*, “Maggie: A Robotic Platform for Human-Robot Social Interaction,” in *2006 IEEE Conference on Robotics, Automation and Mechatronics*, 2006, pp. 1–7.