

# Segmentación de movimiento a partir de la información de cámaras de tiempo de vuelo

Cristina Losada, Manuel Mazo, Sira Palazuelos, David Jiménez  
Departamento de Electrónica. Universidad de Alcalá.  
Alcalá de Henares. Madrid.  
losada@depeca.uah.es

**Abstract**— En este trabajo se plantea una solución para la segmentación en función del movimiento, de los usuarios presentes en un espacio inteligente, a partir de la información de intensidad y profundidad proporcionada por una cámara de tiempo de vuelo (ToF). Para la segmentación se propone una solución basada en la minimización de una función objetivo, definida en función de variables relacionadas con la posición y velocidad de los elementos móviles. Dicha función incorpora la información proporcionada por la cámara de tiempo de vuelo. De forma previa a la minimización, es necesario eliminar el fondo de la escena. Esto se lleva a cabo mediante el análisis de componentes principales generalizado (GPCA). Debido a la baja calidad de las imágenes de intensidad proporcionadas por la cámara ToF, se ha incorporado la información de profundidad para mejorar el resultado de la eliminación de fondo. Las diferentes propuestas planteadas en este trabajo han sido validadas mediante pruebas experimentales realizadas con secuencias de imágenes reales, adquiridas mediante una cámara de tiempo de vuelo.

**Keywords:** *interacción natural, cámaras de tiempo de vuelo, espacio inteligente.*

## I. INTRODUCCIÓN

En los últimos años, la evolución de la electrónica y las comunicaciones ha permitido el desarrollo de los denominados “espacios inteligentes” [1], [2], [3]. En ellos los usuarios pueden interactuar de manera natural con los diferentes sistemas y servicios computacionales que les facilitan la realización de sus tareas diarias.

En este trabajo se considera un espacio inteligente al área física, dotada de sensores que pueden ser de diferentes tipos (cámaras, micrófonos, sensores de ultrasonidos, etc), ubicados en posiciones fijas y conocidas del entorno. El espacio inteligente dispone, además, de un conjunto de actuadores (entre los que se encuentran los robots móviles) que permiten su interacción con el entorno. Tanto los sensores, como los actuadores son controlados por un sistema de supervisión, dotado de capacidad de análisis y toma de decisiones. Dentro del espacio inteligente pueden encontrarse además otros elementos como, posibles obstáculos y usuarios potenciales. Los diferentes elementos mencionados pueden observarse en el esquema de la Figura 1.

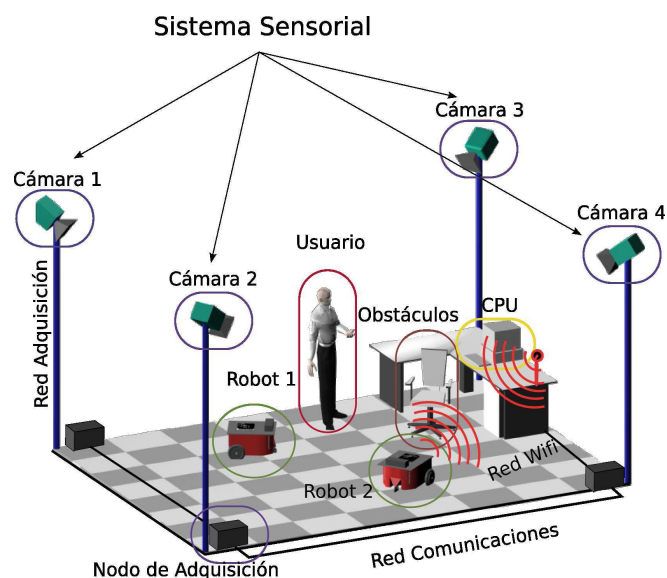


Figura 1. Estructura general del espacio inteligente

En la mayor parte de las aplicaciones de los espacios inteligentes es indispensable que los usuarios puedan interactuar, de forma intuitiva, con el entorno, para generar órdenes o solicitar servicios del mismo. La interacción entre los usuarios y el espacio inteligente puede llevarse a cabo mediante comandos de voz [4], o a través de gestos sencillos. En esta línea, son numerosas las alternativas existentes en la literatura relacionadas con el proceso de interacción, en ellos, los elementos utilizados para la adquisición de información son sensores tanto pasivos [5] (cámaras de vídeo), como activos [6] (sensores de flexión, acelerómetros, etc.) e incluso, fusión entre ellos, como en el trabajo de Brashear et al. [7] en el que se presenta una fusión de sensores que incluye cámara situada convenientemente y unos brazaletes de distintos colores para cada mano, que incorporan acelerómetros con tres grados de libertad, para obtener información de las rotaciones y movimientos hacia la línea visual de la cámara

En los últimos años, se han desarrollado diferentes sensores que permiten obtener, además de la intensidad o color, una medida de profundidad (entendiendo como tal la distancia de cada punto de la escena al sensor). El uso de este tipo de

sensores permite obtener información en tres dimensiones de la escena utilizando una única cámara, sin necesidad de incorporar elementos externos que puedan ser invasivos para los usuarios. Sin embargo, este tipo de cámaras también tiene limitaciones, como la baja resolución de las imágenes adquiridas, o los errores debidos al multicamino.

En este trabajo se emplea una cámara de tiempo de vuelo (cámara ToF) [8] que proporciona una imagen de intensidad, además de la medida de profundidad. Son numerosos los trabajos que emplean esta tecnología para facilitar y mejorar la detección y obtención de la pose de personas [9], [10], así como para la detección de gestos a partir de la información de profundidad [11] o mediante la fusión de imágenes de intensidad y profundidad [12].

En este trabajo se presenta una alternativa para la segmentación en función del movimiento, a partir de la información adquirida por una cámara ToF. La propuesta presentada consta de dos etapas una de eliminación de fondo, y otra de segmentación que se describen en las secciones II y III respectivamente. En la sección IV se muestran los resultados experimentales y, por último, en la V las conclusiones y los trabajos futuros.

## II. ELIMINACIÓN DE FONDO

El primer paso para lograr el objetivo planteado es la eliminación del fondo de la escena. Esta etapa es de gran importancia debido a que permite procesar únicamente las regiones de interés dentro de la imagen.

Debido a que en el espacio inteligente los sensores (en este caso la cámara de tiempo de vuelo) se encuentran ubicados en posiciones fijas y conocidas del entorno, es posible realizar la eliminación de fondo mediante la comparación de la información adquirida en el instante actual, con un modelo del fondo de la escena. Tanto para el modelado del fondo, como para la comparación de la información de entrada con dicho modelo, se utilizará el análisis de componentes principales generalizado [13] (GPCA).

Como ya se ha comentado en la introducción, en este trabajo se utiliza una cámara ToF que, además de una imagen de intensidad ( $I_i$ ), proporciona una imagen de profundidad ( $I_p$ ) que contiene, por cada píxel, la distancia del punto 3D correspondiente al origen de coordenadas del sistema de referencia de la cámara. Este tipo de cámaras consta de un array de leds infrarrojos, que emiten un haz de luz modulada. El tiempo de vuelo del haz recibido (medido en función de la diferencia de fase entre la señal emitida y la recibida) permite estimar la distancia a cada punto mientras que, la amplitud de la señal recibida proporciona la información de intensidad. El principio de funcionamiento de las cámaras ToF se muestra en el esquema de la Figura 2, basado en el presentado en [14].

Debido a las características de la cámara de tiempo de vuelo, a pesar de que el uso de GPCA para la eliminación de fondo en imágenes de intensidad (escala de grises) ha sido validada en diversos experimentos [15], en este trabajo se ha modificado el algoritmo de obtención de contornos de los

objetos presentes en la imagen mediante GPCA, incorporando la información de profundidad proporcionada por la cámara ToF. Esto permite la reducción de los errores cometidos en la segmentación debidos a los cambios en la iluminación, así como a la presencia de sombras proyectadas por los objetos en la imagen, ya que estos no afectan a las imágenes de profundidad.

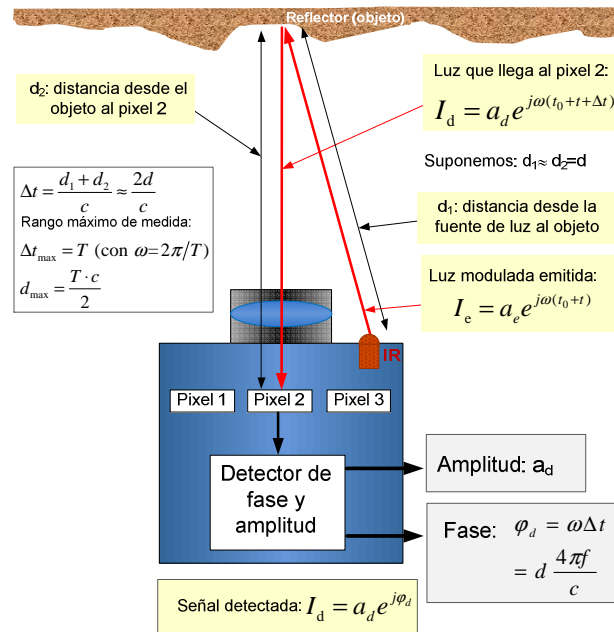


Figura 2. Principio de funcionamiento de una cámara de tiempo de vuelo

Para llevar a cabo la segmentación, en primer lugar es necesario modelar el fondo de la escena. Este modelado se hace a partir de un conjunto de imágenes en las que no aparece ningún elemento móvil (robot o usuario potencial). Dado que la cámara se encuentra ubicada en una posición fija, sólo es necesario modelar el fondo una vez. Por cada tipo de imagen (intensidad y profundidad) se utilizan 10 imágenes y, mediante GPCA se obtienen dos matrices de transformación  $\mathbf{L}$  y  $\mathbf{R}$  que contienen las componentes principales de las imágenes de fondo. Así, tras la etapa de modelado de fondo se obtienen cuatro matrices de transformación: dos que modelan el fondo para las imágenes de intensidad ( $\mathbf{L}_i, \mathbf{R}_i$ ) y otras dos que para las imágenes de profundidad ( $\mathbf{L}_p, \mathbf{R}_p$ ).

Estas matrices se utilizan para determinar qué píxeles de la imagen pertenecen al fondo de la escena y cuáles corresponden a elementos móviles. Para ello, en primer lugar se proyecta la imagen de intensidad al espacio transformado GPCA mediante (1) donde  $\mathbf{M}_i$  es la imagen media de las imágenes de intensidad del fondo. A continuación se obtiene la imagen recuperada mediante la transformación inversa (2) y se estima el error de recuperación como la diferencia entre la imagen original y la recuperada.

De igual forma, se calcula el error de recuperación de la imagen de profundidad, utilizando las matrices  $\mathbf{L}_p$  y  $\mathbf{R}_p$ , y la matriz media de imágenes de fondo  $\mathbf{M}_p$  para transformar y recuperar la imagen de profundidad.

$$\tilde{\mathbf{I}}_{iT} = \mathbf{L}_i^T (\mathbf{I}_i - \mathbf{M}_i) \mathbf{R}_i \quad (1)$$

$$\tilde{\mathbf{I}}_{iR} = \mathbf{L}_i \tilde{\mathbf{I}}_{iT} \mathbf{R}_i^T + \mathbf{M}_i \quad (2)$$

El siguiente paso es umbralizar el error de recuperación para determinar qué píxeles pertenecen al fondo de la escena, y cuáles a elementos móviles (usuarios o robots), ya que dicho error será mayor en aquellos píxeles que pertenezcan a elementos no incluidos en el modelo de fondo. En el caso de la imagen de intensidad, se emplea un umbral adaptativo [16] que permite reducir el efecto de los cambios de iluminación. Mientras que para el caso del error de recuperación de la imagen de profundidad, se emplea un umbral fijo, cuyo valor se obtiene de forma experimental en función de la distancia a la que se encuentran los diferentes elementos pertenecientes al fondo de la escena. En este caso no es necesario utilizar un umbral adaptativo debido a que los cambios en la iluminación, así como las sombras, no afectan a la imagen de profundidad. De esta forma, se considerará que un píxel pertenece a un elemento móvil (usuario, robot, etc.) si el error de recuperación de la imagen de intensidad y de la imagen de profundidad para ese píxel supera los umbrales establecidos. Las diferentes etapas explicadas anteriormente se muestran de forma esquemática en el diagrama de bloques de la Figura 3

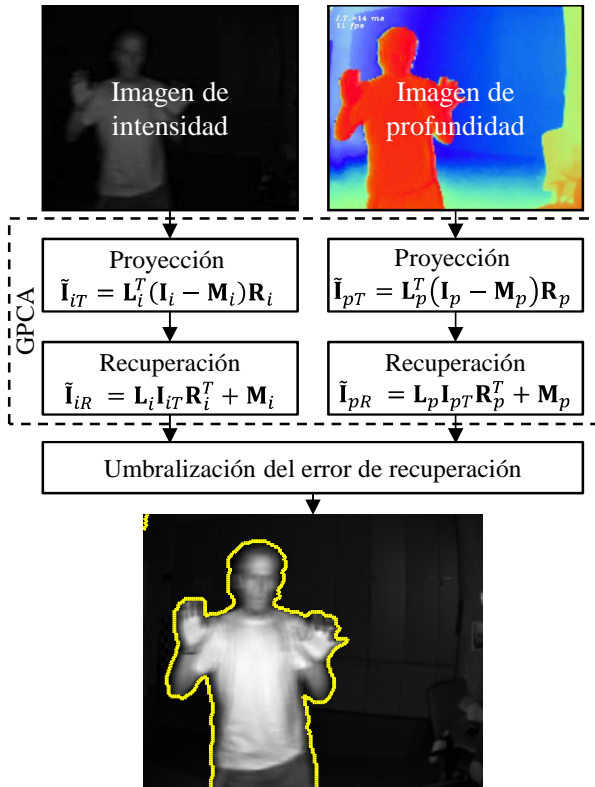


Figura 3. Diagrama de bloques del proceso de eliminación de fondo de la escena, mediante GPCA, a partir de la información de intensidad y profundidad proporcionada por una cámara ToF

Como resultado del proceso de eliminación del fondo de la escena se obtiene una nueva matriz  $\mathbf{R}$  de las mismas dimensiones que las imágenes de entrada en la que los píxeles pertenecientes a elementos móviles tienen un valor 1, mientras

que los pertenecientes al fondo de la escena toman el valor -1, según la expresión mostrada en (3) donde  $\varepsilon_{ri}(i, j)$ ,  $\varepsilon_{rp}(i, j)$  representan el error de recuperación en el píxel  $(i, j)$  para la imagen de intensidad y profundidad respectivamente y  $U_i$ ,  $U_p$  corresponden al valor del umbral.

$$\mathbf{R}(i, j) = \begin{cases} 1 & \text{si } \varepsilon_{ri}(i, j) > U_i \text{ y } \varepsilon_{rp}(i, j) > U_p \\ -1 & \text{resto de casos} \end{cases} \quad (3)$$

$$i = 1, 2, \dots, n_{\text{filas}}; j = 1, 2, \dots, n_{\text{columnas}}$$

La incorporación de la información de profundidad en el proceso de segmentación del fondo permite eliminar regiones de la imagen que se han detectado como parte de elementos móviles debido a brillos, cambios en la iluminación o sombras presentes en la escena. Esto puede observarse en la imagen de la Figura 4, en la que se muestra una imagen de ejemplo, sobre la que se ha representado en color rojo el resultado de la segmentación al utilizar únicamente la información de intensidad, y en verde la segmentación obtenida tras incorporar la información de profundidad. Se puede observar cómo la curva verde se encuentra más próxima al contorno real, habiendo eliminado los efectos del ruido y cambios de iluminación presentes en la imagen de intensidad.



Figura 4. Resultado de la segmentación de fondo en una imagen de ejemplo perteneciente a una de las secuencias de test. En rojo se representa el resultado obtenido al utilizar únicamente la imagen de intensidad y en verde el resultado tras incorporar la información de profundidad

### III. SEGMENTACIÓN DE MOVIMIENTO

Tras la eliminación del fondo de la escena, es necesario realizar la división de las diferentes partes del cuerpo para posteriormente, obtener la posición de las manos y la cabeza. En este trabajo, la segmentación se realiza mediante la minimización de una función objetivo, bajo el supuesto de que cada parte del cuerpo (tronco, cabeza, y brazos) se comporta como un sólido rígido en movimiento entre una imagen y la siguiente en la secuencia.

El punto de partida es la función objetivo (4) definida en el trabajo de Sekkati y Mitiche [17], y obtenida a partir de la restricción de brillo para objetos rígidos. Mediante la minimización de (4) se obtiene, además de la segmentación, el valor de las componentes de velocidad lineal y angular de cada objeto rígido móvil, y una estimación de la profundidad (distancia a la cámara) de cada punto.

En (4)  $\lambda$  y  $\mu$  son constantes reales y positivas que ponderan la contribución de cada uno de los términos de la ecuación y  $\nabla = (\partial_x, \partial_y)$  es el operador gradiente espacial.

$$E[\{\gamma_k\}_{k=1}^{N-1}, \{\mathbf{v}_k\}_{k=1}^N, \{\omega_k\}_{k=1}^N, Z] = \sum_{k=1}^N \left[ \int_{\Omega_k} \Psi_k^2(\mathbf{x}) d\mathbf{x} + \mu \int_{\Omega_k} g(\|\nabla Z\|) d\mathbf{x} \right] + \sum_{k=1}^{N-1} \lambda \oint_{\gamma_k} ds \quad (4)$$

$\lambda, \mu \in \mathfrak{R}$  y  $\lambda, \mu > 0$

De forma previa a la minimización, es necesario conocer el número de grupos ( $N$ ) en que se segmentará la imagen, así como inicializar un conjunto de curvas que evolucionarán, hasta ajustarse a los bordes de los elementos a segmentar. En la propuesta de [17] es necesario inicializar también el valor de la profundidad para cada píxel, sin embargo, debido a que en este trabajo tenemos una medida de profundidad proporcionada por la cámara ToF, consideraremos que el valor de  $Z$  es conocido, y no cambiará durante el proceso de minimización.

A continuación, se presenta el procedimiento de inicialización de las curvas, así como de minimización de la función objetivo.

#### A. Inicialización

El proceso de inicialización permite realizar una primera división del usuario detectado en cabeza, tronco, brazo izquierdo y brazo derecho, siendo, el resultado de esta etapa, las curvas iniciales utilizadas en la fase de minimización de la función objetivo.

De cara a la inicialización de las curvas, pueden encontrarse dos situaciones diferentes: el procesamiento de la primera imagen en que aparece un usuario, en la que no se tiene ningún tipo de información previa, y el caso en que se ha detectado ya la pose del usuario, y la información obtenida se realimenta para la inicialización de la siguiente imagen de la secuencia.

En el primer caso, para facilitar la inicialización, en la primera imagen de la secuencia se solicita al usuario que se coloque en una posición en la que los brazos no se encuentren delante del cuerpo. En esta situación, se comienza ajustando un círculo a la cabeza del usuario. El radio de dicho círculo se obtiene en función de la distancia del usuario a la cámara. A continuación, se ajusta un rectángulo al área del tronco. Igual que en el caso de la cabeza, la anchura del rectángulo utilizado se calcula en función de la distancia a la cámara. Los píxeles que no pertenezcan a la cabeza ni al tronco se consideran parte de los brazos, asignándolos al brazo izquierdo o al brazo derecho en función de su posición respecto al centroide del tronco.

Por otro lado, en caso de tener disponible la pose del usuario en la imagen anterior, esta información se utiliza para la inicialización de las curvas de forma que, bajo el supuesto de que los cambios entre una imagen y la siguiente son pequeños, los píxeles correspondientes a la imagen actual, que pertenecían a una parte del cuerpo de la imagen anterior, se asignan a esa misma parte del cuerpo en la imagen actual. Respecto a los píxeles que queden sin asignar, se incorporan a la región más cercana. En este caso, la información de profundidad permite resolver los casos en los que existan ambigüedades debidas a que los brazos se encuentran ocultando zonas del tronco o la cabeza.

#### B. Minimización de la función objetivo

Tras la inicialización, se realiza un ajuste en función de la velocidad a la que se desplazan las diferentes partes del cuerpo, mediante la minimización de (4).

Tras la inicialización de las curvas y la profundidad, la minimización de la función objetivo (4) se realiza en dos pasos que se repiten hasta alcanzar la convergencia. Estas etapas se describen brevemente a continuación:

En la primera etapa, se suponen conocidas las curvas, de forma que la función de energía a minimizar se reduce a (5):

$$E(\{\mathbf{v}_k\}_{k=1}^N, \{\omega_k\}_{k=1}^N) = \sum_{k=1}^N \int_{\Omega_k} \Psi_k^2(\mathbf{x}) d\mathbf{x} \quad (5)$$

donde  $\Psi_k = \mathbf{I}_t + \mathbf{s} \cdot \mathbf{v}_k / Z + \mathbf{q} \cdot \omega_k$ ,  $\{\mathbf{v}_k\}_{k=1}^N$ ,  $\{\omega_k\}_{k=1}^N$  son las componentes de velocidad lineal y angular respectivamente y  $\mathbf{s}$  y  $\mathbf{q}$  son los siguientes vectores columna, que dependen de la distancia focal de la cámara  $f$  y las derivadas parciales ( $\mathbf{I}_x$ ,  $\mathbf{I}_y$ )

$$\mathbf{s} = \begin{pmatrix} f \mathbf{I}_x \\ f \mathbf{I}_y \\ -x \mathbf{I}_x - y \mathbf{I}_y \end{pmatrix} \quad \mathbf{q} = \begin{pmatrix} -f \mathbf{I}_y - \frac{y}{f} (x \mathbf{I}_x + y \mathbf{I}_y) \\ -f \mathbf{I}_x - \frac{x}{f} (x \mathbf{I}_x + y \mathbf{I}_y) \\ -y \mathbf{I}_x + x \mathbf{I}_y \end{pmatrix}$$

Teniendo en cuenta que  $\Psi_k$  depende linealmente de  $\mathbf{v}_k$  y  $\omega_k$ , la minimización se realiza resolviendo un sistema de ecuaciones lineales mediante mínimos cuadrados.

En la segunda etapa, dada la profundidad y los parámetros del movimiento rígido 3D, la función de energía a minimizar con respecto a las curvas es:

$$E[\{\gamma_k\}_{k=1}^{N-1}] = \sum_{k=1}^N \int_{\Omega_k} \xi_k(\mathbf{x}) d\mathbf{x} + \lambda \sum_{k=1}^{N-1} \oint_{\gamma_k} ds \quad (6)$$

donde  $\xi_k(\mathbf{x}) = \Psi_k^2(\mathbf{x}) + \mu g(\|\nabla Z\|)$ . A partir de esta ecuación es posible actualizar las curvas mediante una iteración del descenso por curvas de nivel.

El proceso de minimización finaliza cuando se alcanza la convergencia. Se considera que el algoritmo converge cuando, entre una iteración y la siguiente no hay cambio en las curvas que definen los contornos de la segmentación.

#### IV. RESULTADOS EXPERIMENTALES

Con objeto de validar la solución propuesta en este trabajo se han realizado múltiples pruebas experimentales. A continuación se presentan los resultados obtenidos para las imágenes pertenecientes a dos secuencias adquiridas usando una cámara ToF modelo SR4000 de Mesa Imaging [18]. Cada una de las capturas de esta cámara proporciona una imagen de intensidad y otra de profundidad, ambas de dimensiones 176x144 píxeles. En la Figura 5 se muestran, a modo de ejemplo, dos imágenes pertenecientes a las secuencias de prueba. En la columna izquierda se presenta la imagen de intensidad mientras que, en la columna derecha se muestra una representación de la imagen de profundidad en la que cada color representa un valor de distancia.

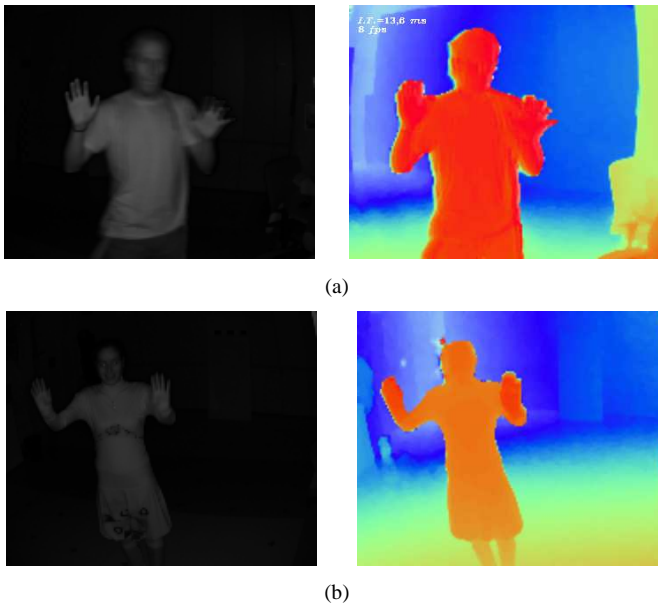


Figura 5. Imágenes pertenecientes a las secuencias de prueba adquiridas con la cámara ToF SR4000. A la columna izquierda se muestra la imagen de intensidad y a la derecha la de profundidad. (a) Secuencia 1 (b) Secuencia 2.

Para comenzar se presentan los resultados obtenidos en la etapa de eliminación del fondo para varias imágenes pertenecientes a las 2 secuencias de prueba. En la Figura 6(a) se muestra el resultado de esta etapa para dos imágenes pertenecientes a la secuencia 1 (Figura 5(a)). La línea roja muestra el resultado obtenido al aplicar GPCA a las imágenes de intensidad mientras que, en verde se ha representado el resultado final tras incorporar la información de profundidad. De igual forma, en la Figura 6(b) se muestra de nuevo el resultado de la eliminación de fondo para dos imágenes pertenecientes a la secuencia 2 (Figura 5(b)).

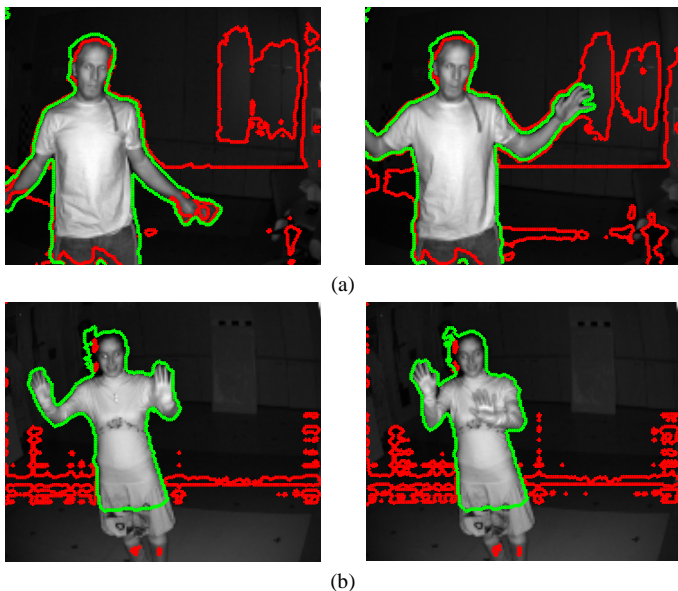


Figura 6. Resultado de la etapa de eliminación de fondo. En rojo se muestra el resultado tras aplicar GPCA a las imágenes de intensidad y en verde el resultado final tras incorporar la información de profundidad. (a) Secuencia 1 (Figura 5(a)). (b) Secuencia 2 (Figura 5(b))

En estas imágenes puede observar como, debido a la baja calidad de las imágenes de intensidad adquiridas por la cámara ToF, al emplear únicamente la información de intensidad, un gran número de píxeles pertenecientes al fondo de la escena se han considerado como parte de los elementos móviles. Sin embargo, al incorporar la información de profundidad, el resultado de la eliminación de fondo mejora notablemente, aproximándose al contorno real de la persona.

A continuación en la Figura 7 se muestran las curvas iniciales obtenidas para la primera imagen de cada una de las secuencias de test.

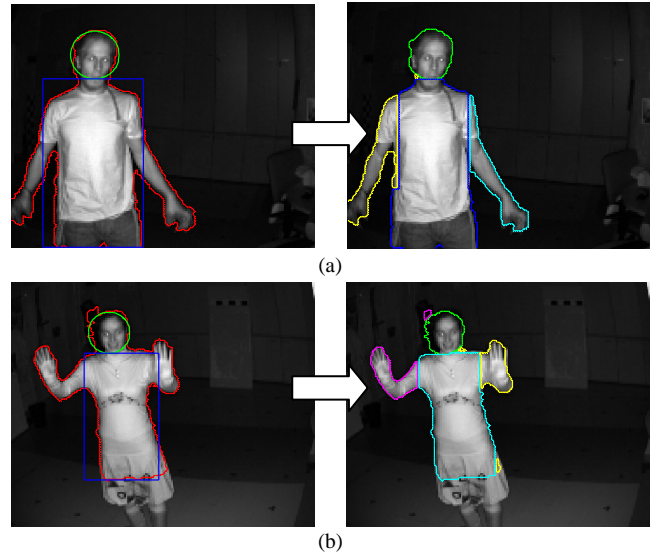


Figura 7. Inicialización de las curvas para la primera imagen de cada una de las secuencias de test. (a) Secuencia 1. (b) Secuencia 2.

Para finalizar, en la Figura 8 se presenta el resultado de la segmentación.

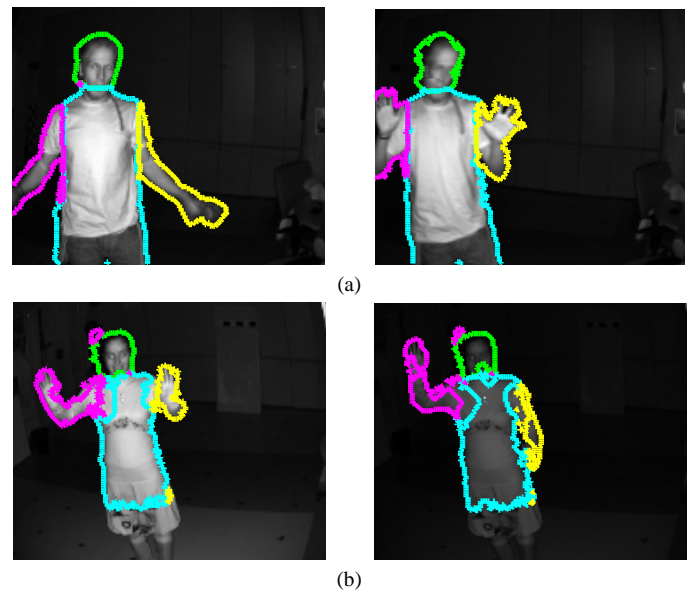


Figura 8. Resultado de la segmentación en función del movimiento. (a) Imágenes pertenecientes a la secuencia 1. (b) Imágenes pertenecientes a la secuencia 2.

En la Figura 8, cada una de las regiones en que se ha dividido el cuerpo del usuario (cabeza, tronco, brazo izquierdo y brazo derecho) se rodea con una curva de un color diferente. En estas figuras se puede observar que la segmentación es correcta, aunque, en otras aparecen errores debidos a que la suposición de que los brazos se comportan como un sólido rígido no se cumple.

## V. CONCLUSIONES Y TRABAJOS FUTUROS

Este trabajo se enmarca en el ámbito de la interacción entre los usuarios y entorno en el espacio inteligente a partir de la información adquirida por cámaras ToF, de gran relevancia dado el enorme auge de las cámaras ToF en la actualidad, y sus múltiples aplicaciones.

En este artículo, se ha presentado una solución para la segmentación en función del movimiento, a partir de la información de intensidad y profundidad proporcionada por una cámara de tiempo de vuelo (ToF), de los usuarios presentes en un espacio inteligente.

Para la segmentación se ha propuesto una solución basada en la minimización de una función objetivo, definida en función de variables relacionadas con la posición y velocidad de los elementos móviles, que incorpora la información proporcionada por la cámara de tiempo de vuelo.

De forma previa a la minimización, es necesario eliminar el fondo de la escena. Esto se lleva a cabo mediante el análisis de componentes principales generalizado (GPCA). Debido a la baja calidad de las imágenes de intensidad proporcionadas por la cámara de tiempo de vuelo, se ha incorporado la información de profundidad, habiéndose verificado en las pruebas experimentales que la inclusión de esta información mejora notablemente el resultado de la eliminación de fondo.

Dado que este trabajo se encuentra en su etapa inicial, son numerosas las líneas de trabajo futuro que se presentan. La más inmediata es la incorporación de modelos articulados a la función objetivo para evitar los problemas derivados de suponer sólidos rígidos en movimientos, y la obtención de la pose 3D de los usuarios para, posteriormente, trabajar en la detección de signos sencillos.

## AGRADECIMIENTOS

Este trabajo ha sido posible gracias a la financiación del Ministerio de Ciencia e Innovación mediante el proyecto VISNÚ (ref. TIN2009-08984).

## REFERENCIAS

- [1] M. Weiser, "The computer for the 21st century," *Scientific American*, vol. 265, no. 3, pp. 94-104, 1991.
- [2] H. Hashimoto, "Intelligent Interactive Spaces - Integration of IT and Robotics," *Advanced Robotics*, pp. 85-90, 2005.
- [3] R. a. Brooks, "The Intelligent Room project," *Proceedings Second International Conference on Cognitive Technology Humanizing the Information Age*, pp. 271-278, 1997.
- [4] F. Zhou, G. Tian, Y. Yang, H. Xiao, and J. Chen, "Research and implementation of voice interaction system based on PC in intelligent space," in *2010 IEEE International Conference on Automation and Logistics (ICAL)*, 2010, pp. 679-684.
- [5] K. Imagawa, S. Lu, and S. Igi, "Color-based hands tracking system for sign language recognition," in *Third IEEE International Conference on Automatic Face and Gesture Recognition, 1998. Proceedings.*, 1998, pp. 462-467.
- [6] L. Dipietro, A. M. Sabatini, and P. Dario, "A survey of glove-based systems and their applications," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 38, no. 4, pp. 461-482, 2008.
- [7] H. Brashear, T. Starner, P. Lukowicz, and H. Junker, "Using multiple sensors for mobile sign language recognition," in *Seventh IEEE International Symposium on Wearable Computers, 2003. Proceedings.*, 2003, pp. 45-52.
- [8] R. Lange, "Solid-state time-of-flight range camera," *IEEE Journal of Quantum Electronics*, vol. 37, no. 3, pp. 390-397, 2001.
- [9] C. Plagemann and V. Ganapathi, "Real-time identification and localization of body parts from depth images," *2010 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3108-3113, 2010.
- [10] J. Shotton et al., "Real-time human pose recognition in parts from single depth images," in *CVPR*, 2011, vol. 2, p. 3.
- [11] P. Breuer and C. Eckes, "Hand gesture recognition with a novel IR time-of-flight range camera: a pilot study," in *Proceedings of the 3rd international conference on Computer vision/computer graphics collaboration techniques (MIRAGE'07)*, 2007, pp. 247-260.
- [12] M. Holte, T. Moeslund, and P. Fihl, "Fusion of range and intensity information for view invariant gesture recognition," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2008. CVPRW '08.*, pp. 1-7, 2008.
- [13] J. Ye and R. Janardan, "GPCA: an efficient dimension reduction scheme for image compression and retrieval," *Proceedings of the tenth ACM SIGKDD*, 2004.
- [14] D. Falie and V. Buzuloiu, "Distance errors correction for the Time of Flight ( ToF ) Cameras . Dragos Falie Laboratorul de Analiza si Prelucrarea Vasile Buzuloiu Laboratorul de Analiza si Prelucrarea," *IEEE International Workshop on Imaging Systems and Techniques, 2008. IST 2008.*, pp. 123-126, 2008.
- [15] C. Losada, M. Mazo, S. Palazuelos, D. Pizarro, and M. Marrón, "Multi-Camera Sensor System for 3D Segmentation and Localization of Multiple Mobile Robots," *Sensors*, vol. 10, no. 4, pp. 3261-3279, Apr. 2010.
- [16] C. Losada, M. Mazo, and S. Palazuelos, "Adaptive threshold for robust segmentation of mobile robots from visual information of their own movement," in *IEEE International Symposium on Intelligent Signal Processing, 2009. WISP 2009.*, 2009, pp. 293-298.
- [17] H. Sekkati and A. Mitiche, "Concurrent 3-D motion segmentation and 3-D interpretation of temporal sequences of monocular images.," *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 15, no. 3, pp. 641-53, Mar. 2006.
- [18] <http://www.mesa-imaging.ch/prodview4k.php> (accesible el 03/05/2012)