Seguimiento tridimensional de múltiples objetos en espacios inteligentes mediante un único filtro de partículas

A. Marcos, M. Marrón, D. Pizarro, R. Jalvo, M. Mazo

Departamento de Electrónica, Universidad de Alcalá, 28001, Alcalá de Henares, Madrid, España.

***Resumen* — Este trabajo propone un nuevo método de seguimiento de múltiples objetos en un espacio inteligente. Se parte de una serie de cámaras sincronizadas y situadas estratégicamente en el entorno para extraer toda la información referente al mismo e interaccionar con él. A partir esta información, el trabajo aquí descrito permite, por un lado, obtener una reconstrucción tridimensional de las entidades presentes en el espacio inteligente, y por otro, realizar el seguimiento tridimensional de esta reconstrucción mediante técnicas probabilísticas. A partir del seguimiento, se define de forma individual cada una de las entidades existentes en el espacio inteligente en cada instante. En este documento se demuestra, además, la fiabilidad y robustez del algoritmo mediante diversas pruebas.**

I. Introducción

El seguimiento de entidades como personas y robots ha sido un campo de gran interés en los últimos tiempos. Así lo demuestra la extensa literatura publicada, en la que se definen diversas maneras de abordar la problemática que ello conlleva.

Asimismo, el concepto de espacio inteligente, caracterizado por contar con inteligencia distribuida en un entorno compuesto por cámaras, robots y otros agentes, es objeto de continuas propuestas ([1], [2]), y se muestra en la Fig 1.

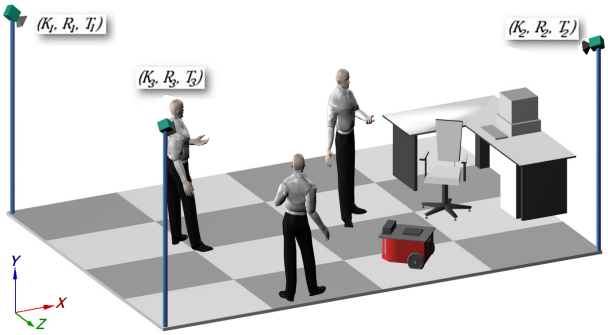

Fig. 1. Renderizado tridimensional del espacio inteligente. Cada cámara está definida por sus parámetros intrínsecos y extrínsecos: *K, R* y *T*.

En este trabajo se propone un seguimiento tridimensional de múltiples entidades en un espacio inteligente mediante un único PF (Particle Filter o Filtro de Partículas), consiguiéndose así un importante ahorro de carga computacional. Además, el uso de un algoritmo probabilístico robustece y aporta fiabilidad a la funcionalidad del seguidor frente a seguidores determinísticos como [3]. Asimismo, la resolución del problema de estimación con un solo algoritmo multimodal constituye una importante mejora frente a otras propuestas como [4], [5], donde se utiliza un filtro por cada elemento a seguir.

El sistema de observación se basa en un anillo de cámaras calibradas, sincronizadas y comunicadas con un conjunto de servidores (uno por cámara), que procesan cada imagen y envían la información elaborada a un equipo cliente mediante red local. La situación de las cámaras es tal que se evitan en la medida de lo posible las oclusiones de cualquiera de los objetos a seguir. A partir de las imágenes obtenidas por las cámaras se realiza una segmentación basada en resta de fondo. Los elementos presentes en la escena quedan diferenciados en una imagen binaria. Posteriormente, en el procesador cliente, mediante la técnica conocida como Visual Hull (descrita en [6]), se obtienen una serie de proyecciones a planos paralelos al del suelo de los elementos binarizados. Estas proyecciones se utilizan como observaciones de entrada del seguidor. En este caso se trata de un PF, con el objetivo, ya mencionado, de realizar el proceso con un solo seguidor multimodal.

El uso de un único filtro para seguir múltiples hipótesis no es novedoso, ya que ha sido anteriormente propuesto y aplicado con éxito al problema de seguimiento de múltiples objetos en [7], [8], [9] y [10]. La multimodalidad del PF se explota para el seguimiento robusto de múltiples objetos con un solo filtro gracias al uso de un clasificador. En el caso tridimensional se requiere una clasificación más elaborada para que sea adaptable a cualquier modelo deformable. En este trabajo se proponen otros mecanismos para robustecer la multimodalidad del PF.

El resto del documento se organiza de la siguiente forma: en el siguiente apartado se describe el proceso de obtención del conjunto de observaciones de entrada al PF. En el apartado III se describe sobre el algoritmo de seguimiento multi-hipótesis propuesto basado en un solo PF. En el apartado IV se constata la validez de dicho algoritmo mediante resultados obtenidos con datos reales. Finalmente, en el apartado V se exponen las conclusiones extraídas, y se abre la puerta a trabajos futuros.

II. Algoritmo De Observación

Para hacer posible el seguimiento tridimensional de uno o varios objetos, como se persigue en este trabajo,

es necesario contar con un sistema de observación que extraiga información tridimensional del entorno. Se propone la utilización de Visual Hull.

El Visual Hull se define como el mayor volumen encerrado por la intersección de los conos visuales de las siluetas obtenidas a partir de diferentes cámaras. El volumen que se obtiene no se corresponde con el volumen real que ocupan los objetos en la escena. Sin embargo, puede asegurarse que lo contiene. Puede verse gráficamente en la Fig. 2.

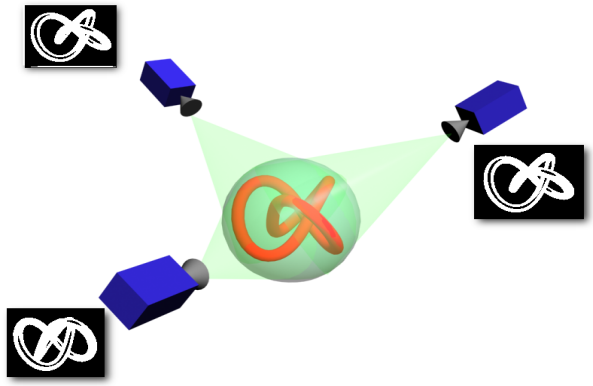

Fig. 2. Representación gráfica del concepto de Visual Hull.

Este volumen se obtiene siguiendo el proceso descrito a continuación.

Se parte de un anillo de cámaras calibradas. Se asume un modelo pin-hole de las mismas. El modelo pin-hole está definido por una matriz K que contiene características intrínsecas a la cámara, como la distancia focal o distorsiones, y dos matrices R y T que definen las propiedades extrínsecas de la cámara: rotación y traslación, respectivamente.

Mediante una resta de fondo por distancia de Mahalanobis se segmentan las imágenes recibidas por cada cámara, consiguiendo separar así los objetos presentes en el entorno del fondo del mismo. Los elementos de la escena quedan binarizados.

Una vez detectados los objetos a seguir en cada cámara, es necesario referenciarlos a un plano común, denotado por Π. Esto se realiza mediante la matriz H de homografía:

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \lambda^{-1} H^{-1} \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} \tag{1}$$

Transformando (1) es posible obtener una relación entre K, R y T con la matriz H, como puede verse en la siguiente expresión, equivalente a (1):

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \lambda K \left(R_c \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} + T_c \right) \tag{2}$$

Multiplicando las imágenes binarizadas de cada cámara una vez referidas al mismo plano, se obtiene la intersección de los objetos de la escena con ese plano.

El plano Π es discretizado mediante una rejilla de ocupación. Cada celda de la misma, de lado o, tiene dos posibles estados: lleno, si forma parte del volumen definido por Visual Hull, o vacío, si no forma parte.

La transformación originada al aplicar H crea una escala en milímetros. Se hace necesario el uso de una matriz M que re-escale la imagen a pixeles, para obtener la imagen final.

Repitiendo este proceso para una serie de planos Π paralelos, es posible reconstruir el volumen ocupado por los objetos presentes en el entorno, como se ilustra en Fig. 3.

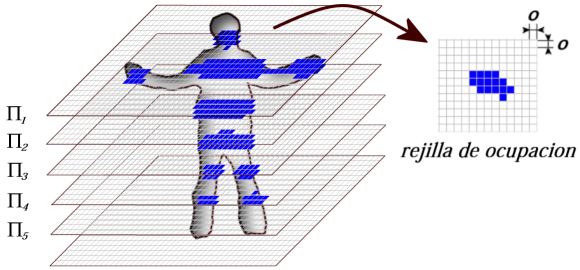

Fig. 3. Esquema del concepto de Visual Hull. Se utilizan distintos planos Π a diferentes alturas sobre las que se proyectan las siluetas del volumen a esas alturas. Como puede verse en el detalle del plano superior, el espacio de cada plano está discretizado, constituyendo lo que se conoce como rejilla de ocupación. El tamaño del lado de cada celda es o.

III. Algoritmo De Seguimiento

Existen dos tendencias principales en los trabajos expuestos en la literatura del área para realizar el seguimiento de uno o varios objetos, definiendo éste como la estimación del vector de posición x_t de cada uno de ellos en un espacio tridimensional. Por un lado se encuentra el Filtro de Kalman (KF), que permite obtener la solución óptima de estimación, si el modelo probabilístico del proceso es lineal y gausiano. La alternativa a éste, de aparición relativamente reciente, está en los Filtros de Partículas (PF), que no aportan una solución exacta pero no son tan estrictos en cuanto las características del sistema.

Este trabajo propone la utilización de un PF con el modelo de observación tridimensional anteriormente descrito. El PF utilizado ofrece, como se explica más adelante, mayor fiabilidad en la estimación con un modelo probabilístico no definido con total exactitud, y mayor robustez ante modelos de observación o problemas de kidnapping, como ha sido anteriormente demostrado [9]. Además, el hecho de utilizar un único filtro multimodal para realizar un seguimiento de múltiples objetos permite mantener el tiempo de ejecución del seguidor estable y bajo,

independientemente del número de hipótesis a seguir. Esto es así debido a que la carga computacional del PF depende del número de muestras que se tomen de la función de creencia $p(\overrightarrow{x_t}/\overrightarrow{y_{1:t}})$ (es decir, del número de partículas, [10]). Al utilizar un único PF de número constante de partículas este tiempo se mantiene, por tanto, constante.

Alternativas al uso de un solo PF para el seguimiento de múltiples objetos consisten en utilizar un filtro por cada objeto a seguir ([12], [5]), o ampliar el vector de estado para incluir todas las hipótesis a seguir ([8]), incrementándose en ambos casos el tiempo de ejecución del seguidor de forma exponencial con el número de objetos a seguir.

El uso de un único filtro para seguir múltiples entidades plantea, por otro lado, una serie de problemas que deben ser abordados. El más importante es la necesidad de mantener siempre la representación de todas las hipótesis sobre las que se realiza el seguimiento en la creencia multimodal representada por el set de partículas. Si se crea una distribución desigual de las partículas entre las distintas hipótesis se hace vulnerable el sistema frente a perturbaciones, por ejemplo, en el modelo de observación.

Para evitar este problema se propone el algoritmo que se muestra a continuación (Fig. 4) y se explica en los párrafos siguientes.

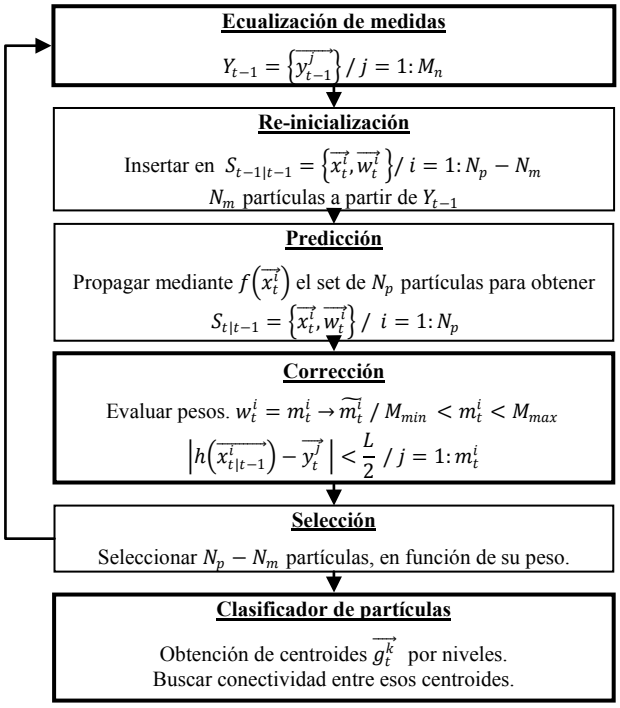

Fig. 4. Flujograma del PF propuesto para seguimiento tridimensional.

La estructura del algoritmo de la Fig. 4 está basada en la del Filtro de Partículas Extendido (XPF) descrito en [11]. La aportación más importante en el XPF frente al PF básico (algoritmo Bootstrap, [13]) consiste en la inserción de un primer paso de re-inicialización en el que se modifica la función de creencia a priori, sobre la que se ejecutan después los pasos clásicos de predicción, corrección y selección del estimador. En este paso se convierten en partículas N_m observaciones del instante anterior tomadas de forma aleatoria en partículas, obteniéndose así un set que incluya todas las hipótesis observadas en $t-1$. De esta forma, los objetos recientemente aparecidos en la escena forman parte de la función de creencia.

En el algoritmo descrito en [9] y [10] se añade además otra modificación al XPF. Consiste en incluir un proceso de clasificación de medidas que, utilizado en los pasos de re-inicialización y corrección, robustezca más la representación de todas las hipótesis en la creencia discreta $p(\overrightarrow{x_t}/\overrightarrow{y_{1:t}}) = S_{t|t-1}$. En este artículo se propone prescindir de este clasificador, con objeto de evitar un proceso de seguimiento demasiado lento y poco flexible para ser utilizado en aplicaciones de tiempo real, como se expone en la introducción. Para mantener en este caso el seguimiento multimodal es necesario actuar sobre el paso de corrección. Además, para obtener la salida final determinística del algoritmo se ha añadido a la salida del PF tridimensional un clasificador de partículas.

A. ECUALIZACIÓN DE MEDIDAS

El objetivo es eliminar las medidas y_{t-1} que no sean útiles para el proceso de seguimiento. Es decir, reducir la densidad de medidas en zonas en las que ya haya suficientes como para obtener una salida del algoritmo fiable. Esto se realiza umbralizando el histograma de medidas en las distintas dimensiones, y aporta dos ventajas. Por un lado se reduce el tiempo de ejecución del proceso de corrección (el cual influye en gran medida en el tiempo de ejecución total), ya que depende del número de medidas y del de partículas. Por otro lado se consigue una densidad de medidas más uniforme a lo largo de la escena. Debido a la naturaleza del algoritmo de asignación de pesos, a su vez se obtiene una mejor distribución de partículas. Ésta será más uniforme en torno a los distintos objetos. La Fig. 5 muestra el principio de funcionamiento.

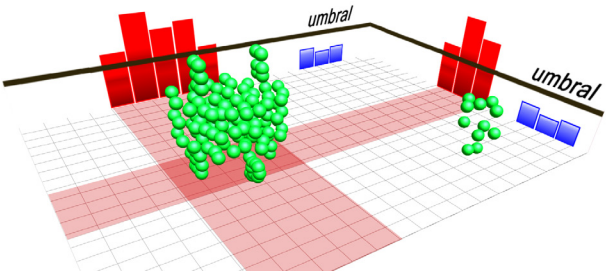

Fig. 5. Esquema de la ecualización de medidas. Se han representado los histogramas de medidas en dos dimensiones. Se ha dibujado en rojo la zona en la que hay medidas redundantes (se supera el umbral establecido).

B. CORRECCIÓN

En el algoritmo propuesto el paso de corrección o ponderación de cada partícula se realiza asignando a cada partícula un peso x_t^i proporcional al número de observaciones cercanas m_t^i a las hipótesis de posición que ésta representa. En la propuesta implementada se entiende como observaciones cercanas a aquellas que caen dentro de un cubo de lado L con centro en la partícula correspondiente.

Para reducir el impacto de las distintas densidades de medidas asociadas a la naturaleza de cada objeto (los objetos más finos estarán más pobremente sensados), se realiza una saturación del conjunto de pesos al rango

$$\frac{M_{min}}{\sum_{i=1:N_p} w_t^i} < \widetilde{m_t^i} < \frac{M_{max}}{\sum_{i=1:N_p} w_t^i} \quad (3)$$

Donde $\widetilde{m_t^i}$ es el número normalizado de medidas asociado a cada partícula M_{partic} y:

$$M_{max} = \alpha\, M_{\max\ cubo} \quad (4)$$
$$M_{min} = \beta\, M_{\max\ cubo} \quad (5)$$

Con $\alpha > \beta$, $\alpha, \beta \in [0,1]$, y:

$$M_{\max\ cubo} = \frac{L^3}{o^3} \quad (6)$$

Esta saturación asegura que no se asigne a la partícula i un número de medidas mayor que M_{max} ni menor que M_{min}. Se consigue así establecer un límite a partir del cual la densidad de medidas que rodea a cada partícula no modifica el peso de la partícula a la que se asocia. De esta forma, se disminuye la diferencia de pesos entre las partículas que representan a objetos pobremente sensados y las que representan a los mejor sensados.

C. CLASIFICACIÓN DE PARTÍCULAS

La salida del PF es una función de creencia discretizada mediante partículas de las hipótesis de posición de los múltiples objetos seguidos. La salida del algoritmo debe ser, sin embargo, una estimación determinística de la posición de cada objeto representado como un modo de esta creencia. Es por lo tanto necesario añadir un proceso de agrupación de partículas a la salida del PF tridimensional propuesto. El proceso de clasificación se realiza en dos pasos.

1. *Primer paso:* Se divide el volumen del entorno con P planos paralelos al del suelo, denominados Ω, con una distancia h entre dos planos contiguos. Las partículas contenidas entre dos planos de altura contigua son proyectadas en un plano intermedio con altura:

$$h_{proy,i} = \frac{h_{i+1}+h_i}{2} \quad (7)$$

De este modo se obtienen P planos paralelos al del suelo sobre los que se encuentra la proyección de las partículas de un rango de altura determinado. A continuación se aplica un algoritmo de clasificación K-medias extendido con soporte para un número de clases variables. De este modo se obtiene una serie de centroides $\overrightarrow{g_t^k} = [x_i\ z_i\ y_i]$ en el espacio *XZY* en alturas predeterminadas. Estos puntos se corresponden con las posiciones estimadas de las secciones de los objetos del entorno en ese rango de altura.

2. *Segundo paso:* Consiste en aplicar un filtro de conectividad para unir los centroides anteriormente obtenidos. El algoritmo de conectividad busca, para cada punto p_i, otros puntos $(p_j \dots p_n)$ a una distancia euclidea menor que *D,* que de este modo se definen como conexos a p_i. El proceso se repite para cada punto, desechando los ya analizados, y originando una serie de estructuras. Cada elemento de una estructura $\overrightarrow{g_t^k}$ es conexo a cada elemento de la misma estructura, y no conexo a cualquier elemento de otra. El esquema de funcionamiento se muestra en la Fig. 6.

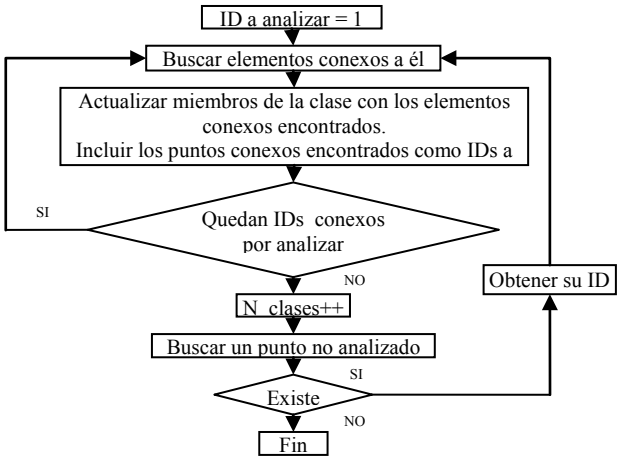

Fig. 6. Flujograma del segundo paso del algoritmo de clasificación.

Es necesario tener en cuenta que la coordenada *Y* en la que pueden estar presentes los centroides se ha discretizado. Sin embargo, el clasificador proporciona unas coordenadas *XZ* continuas. Para mantener la fiabilidad en el algoritmo de conectividad, es necesaria la inclusión de dos parámetros, γ y δ, que escalen las coordenadas tridimensionales para calcular la distancia de conectividad *d* según la siguiente expresión:

$$d_{i,j|t} = \sqrt{\gamma(x_i - x_j)^2 + \gamma(z_i - z_j)^2 + \delta\left(y_i - y_j\right)^2} \quad (8)$$

Se considera que cada estructura es una clase de salida. Se propone asignar unequívocamente una clase de salida a cada elemento presente en el entorno. En la

Fig. 7 se ilustra el funcionamiento del clasificador de partículas.

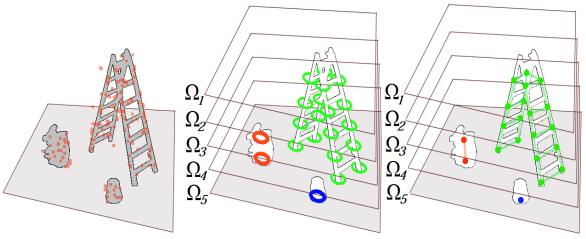

Fig. 7. Clasificación de partículas. Izquierda: partículas (en rojo) distribuidas sobre los objetos de la escena. Centro: una vez que se han proyectado en distintos planos, se aplica un k-medias por cada plano Ω, obteniendo una serie de centroides, que representan los distintos grupos de partículas en el plano correspondiente. Derecha: se aplica el algoritmo de conectividad propuesto sobre los centroides, obteniendo las clases finales.

IV. Resultados

Las propuestas de este trabajo se han puesto a prueba en el entorno de simulación Matlab con datos adquiridos de un entorno real, a partir de 4 cámaras color calibradas, tomando imágenes a una resolución de 640x480 píxeles. Las cámaras están conectadas a un elemento central o cliente por red. El cliente lleva a cabo la tarea de sincronización entre las cámaras.

El espacio de trabajo útil consiste en una habitación de 3,5 m x 3,5 m, sobre la que se orientan las cámaras.

La rejilla de ocupación está definida por el volumen definido por cada medida. Es decir, por el lado del cubo utilizado para discretizar el volumen. Para la realización de las pruebas se utiliza 40 mm de lado, y 1500 partículas.

En la Fig. 7 se muestra el efecto de las modificaciones propuestas en el paso de corrección.

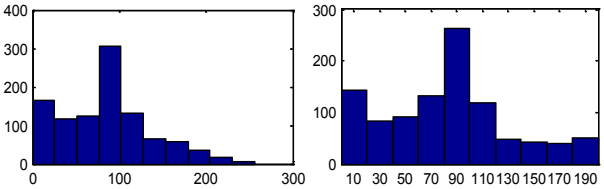

Fig. 7. Efectos del saturado de pesos: histogramas de pesos antes de la saturación (izquierda) y después (derecha).

A la izquierda se observa el histograma de pesos w_t^i de las partículas sin utilizar el sistema de saturación propuesto; en él se observa que existe un gran número de partículas con pocas medidas a su alrededor (peso bajo). Éstas serán eliminadas en gran medida, suprimiendo la hipótesis de seguimiento que representa y beneficiando a la escasa cantidad de partículas que están rodeadas por un gran número de medidas. En la zona derecha se muestra la misma distribución de pesos, una vez se ha aplicado el método de saturación, observándose que la importancia entre partículas está mejor distribuida. Se ha utilizado $\alpha = 0.01, \beta = 0.35$.

La Fig. 8 muestra los efectos de la ecualización de medidas en un experimento con hasta 5 objetos. La gráfica de la parte superior muestra la comparación entre el número de clases reales del número de clases presentes en la escena, establecido manualmente (en verde) y el número de clases que se obtiene mediante el algoritmo propuesto, sin realizar el paso de ecualización de medidas. Abajo, se muestra la misma prueba, añadiendo al proceso de seguimiento la ecualización de medidas propuesta.

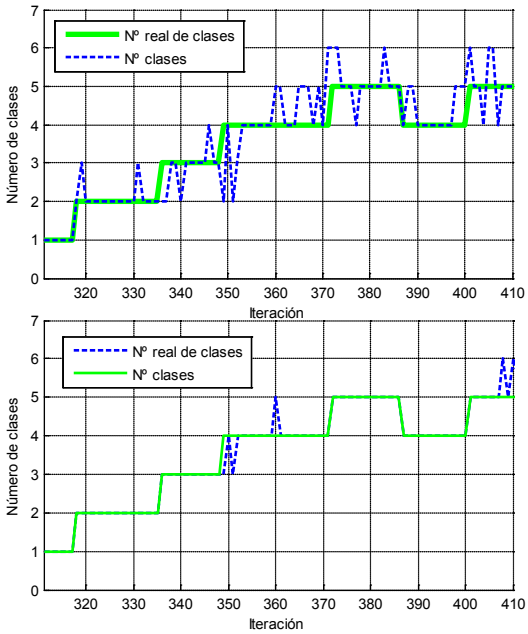

Fig. 8. Arriba: comparación entre el número de clases reales y número de clases de salida obtenidos, sin ecualizar las medidas. Abajo: misma prueba ecualizando las medidas.

En el caso de no utilizar la ecualización de medidas, se aprecian errores en u n 55% de las iteraciones. El 70% de estos errores son debido a la no detección de un objeto, mientras que el 30% restante es debido a la unión de dos clases. Al utilizar la ecualización, el sistema final de seguimiento arroja una tasa de error de sólo el 5. El 80% de estos errores se corresponde con la unión de dos clases, y el 20% restante con la no detección de objetos.

Para probar el sistema se ha utilizado una secuencia en la que se parte de un objeto en la escena. Una persona entra y coloca otros dos objetos, para después salir y volver a entrar. Los resultados pueden verse en la Fig. 9.

V. Conclusiones y Trabajos Futuros

En este trabajo se resuelve la problemática del seguimiento de múltiples objetos en tres dimensiones.

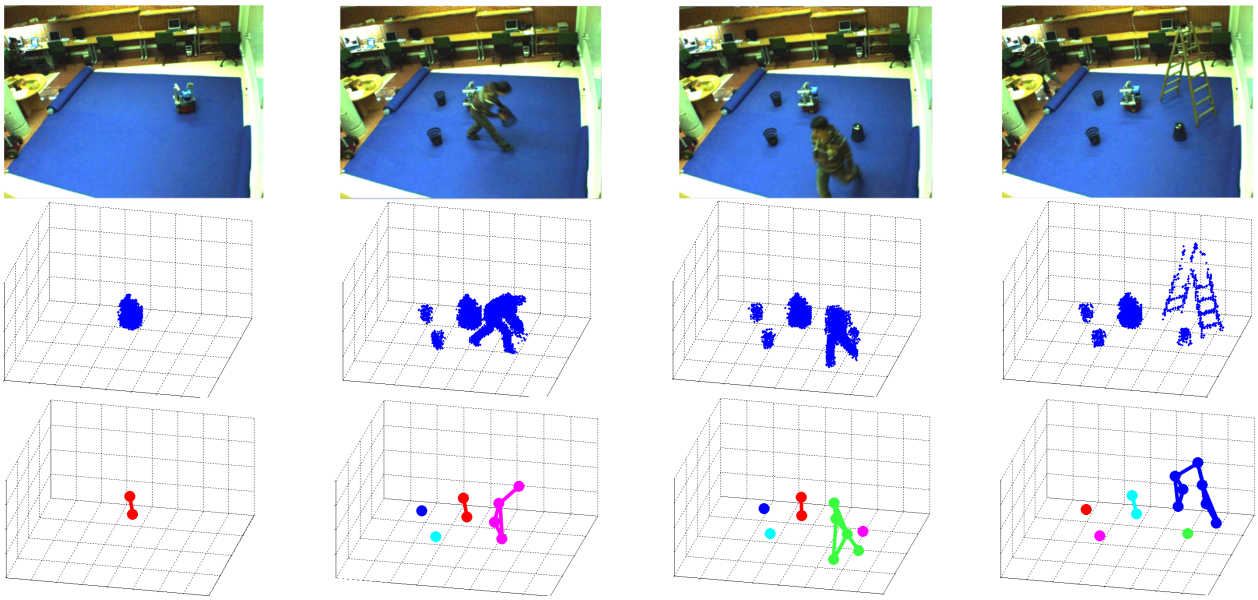

Fig 9. Secuencia de uno de los videos de test. Arriba: imágenes reales capturadas de una de las cámaras. Centro: reconstrucción volumétrica mediante Visual Hull. Abajo: Clases de salida del filtro.

Para ello se utiliza un sistema de observación basado en Visual Hull, que genera la entrada a un Filtro de Partículas Extendido modificado. Los resultados muestran que el sistema responde de manera satisfactoria en entornos reales.

Como trabajos futuros se propone implementar en tiempo real el sistema aquí propuesto, modificar el algoritmo de observación de modo que sea capaz de trabajar en condiciones adversas de iluminación (actualmente se coloca un tapiz de cualquier color para evitar los reflejos de luz interior en el suelo), e incluir un sistema de identificación de robots como el propuesto por los autores en [9] al sistema aquí descrito.

REFERENCIAS

[1] J. Lee, N. Ando and H. Hashimoto, "Intelligent space for human and mobile robot", *Proceedings of the 1999 IEEE/ASME, International Conference on Advanced Intelligent Mechatronics*, 1999.

[2] L. Jeni and Z. Istenes, "Mobile agent control in intelligent space using reinforcement learning", *7th International Symposium of Hungarian Researchers on Computational Intelligence*, 2006.

[3] M. Atyabi, M. S. Kharjeh Hosseini and M. Mokhtari. "The webcam mouse: visual 3D tracking of body features to provide computer access for people with severe disabilities", *Islamic Azad University science & research Branch,* 2006.

[4] J. Krumm, S. Harris, B. Meyers, B. Brumitt, M. Hale and Steve Shafer, "Multi-camera multi-person tracking for EasyLiving", *Third IEEE International Workshop on Visual Surveillance*, 2001.

[5] A. López, C. Canton-Ferrer and J. R. Casas, "Multi-person 3D tracking with particle filters on voxels", *Image Processing Group, Technical University of Catalonia*, 2007.

[6] D. B. Yang, H. H. González-Baños and L.J. Guibas, "Counting people in crowds with a real-time network of simple image sensors", *Proceedings of the Ninth IEEE International Conference on Computer Vision*, 2003.

[7] N. Checka, K. Wilson, V. Rangarajan and T. Darrell, "A probabilistic Framework for multi-modal multi-person tracking", *Artificial Intelligence Laboratory, Massachusetts Institute of Technology*, 2003.

[8] J. Kang, I. Cohen and G. Medioni, "Continuous Tracking Within and Across Camera Streams", *IRIS, Computer Vision Group, University of California*, 2003.

[9] D. Pizarro, M. Marrón, D. Peón, M. Mazo, J. C. García, M. A. Sotelo and Enrique Santiso, "Robot and obstacles localization and tracking with an external camera ring", *2008 IEEE International Conference on Robotics and Automation,* 2008.

[10] M. Marrón, M. A. Sotelo, J. C. García, D. Fernández, D. Pizarro, "XPFCP: An extended particle filter for tracking multiple and dynamic objects in complex environments", *Proceedings of the IEEE International Symosium on Industrial Electronics,* 2005.

[11] E. B. Koller-Meier, F. Ade, "Tracking Multiple Objets Using a Condensation Algorithm", *Journal of Robotics and Autonomous Systems, Vol. 34, pp. 93-105,* 2001.

[12] D. Focken and R. Stiefelhagen, "Towards visión-based 3-D people tracking in a smart room", *Interactive System Laboratories, Universität Karlsruhe (TH), Germany,* 2003.

[13] N.J. Gordon, D. J. Salmond, A. F. M. Smith. "Novel approach to nonlinear/non-gaussian bayesian state estimation", *IEEE Proceedings in Radar and Signal Processing, Vol. 140, nº2, pp. 107-113,* 1993.