Multi-agent 3D Tracking in Intelligent Spaces with a Single Extended Particle Filter

M. Marrón, D. Pizarro, J. C. García, A. Marcos, R. Jalvo, M. Mazo
Departamento de Electrónica, Universidad de Alcalá, 28871, Alcalá de Henares, Madrid, España.

***Abstract* — A new method for tracking multiple objects in an intelligent space is proposed in this paper. The observation model is based on a camera ring statically mounted at the ceiling of the environment in order to obtain all relevant information related to the different objects that wonder (get into and go out) in the space of interest. In the paper, the two subsystems used to track all static and dynamic entities wondering in the intelligent space: a three-dimensional reconstruction of these entities; and, an individual track of all these entities in their movement along the environment with probabilistic techniques. The reliability, and robustness of the proposal presented is finally also demonstrated in this paper with different tests.**

***Keywords* – position estimation, probabilistic algorithms, three-dimensional reconstruction, intelligent spaces.**

I. Introduction

Tracking a variable number of different entities (dynamic and static, controllable and uncontrollable) in a standard 3D environment has become a very interesting researching topic in the last years. This interest can be measured through the increasing amount of publications related to this question that proposes different algorithms, sensors, and techniques to solve the problems that appear in these kinds of situations.

Intelligent spaces (IS [1], [2]) are one of the most recent areas in which multi-tracking has experimented a huge increase of interest. IS term define environments, generally indoor, in which the sensing and processing capability is distributed among different elements already installed on them; those elements allow to achieve a global and instantaneous knowledge and control of the different agents (persons, robots, obstacles) located at each moment inside the IS.

A general description of an IS based on visual sensors, as the one used in this work, is shown in Fig. 1.

On the other hand, many processing algorithms suitable to be used in the tracking task are based on particle filters (PF). Nevertheless, this kind of solutions frequently has got some problems to manage the multiple objects tracking issue.

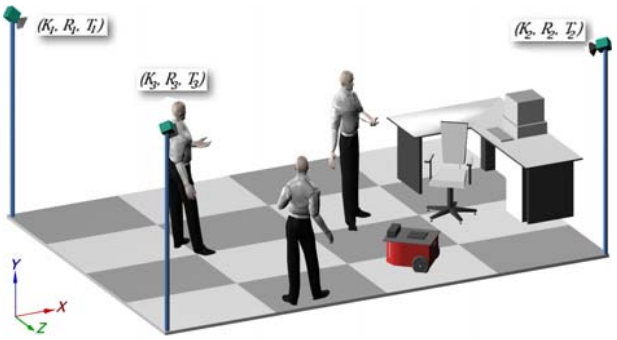

Fig. 1. 3D render of the IS. Each camera is defined by its intrinsic and extrinsic parameters: *Ka*, *Ra* and *Ta* (a=*1..3* in the figure).

The work presented in this paper proposes a new method of multi agent tracking in intelligent spaces using a single PF to develop the 3D estimation process involved in the multi-tracking task. This proposal has two important advantages:

- Execution time of this task is independent of the number of elements being tracked at each time. This fact becomes essential if the multi-tracking task is to be used in real-time applications and crowded rooms, in which unimodal algorithms (one estimator for each object to be tracked) have demonstrated to be unsuccessful [3] [4].
- A probabilistic character is given to the estimation problem, increasing the robustness and reliability of the processing system if compared to the one of some other deterministic solutions [5].

Using a single PF in the tracking task has already been successfully proposed in previous works [6] [7] [8] [9]. In some of them (i.e. [8]) a measurements' clustering process has been added to the filter in order to exploit the PF multimodality. In the work presented in this paper, a more complex solution is needed, in order to obtain a solution flexible enough to be used with non rigid 3D objects.

The observation system used in this work is based on a ring of calibrated cameras, synchronized and interconnected in server-client architecture through a

978-1-4244-5059-6/09/$26.00 ©2009 IEEE

LAN. Cameras are distributed in the IS in places that try to avoid occlusions in the tracking task.

The multi-camera data is processed as follows:

- At every server, a simple background subtraction technique is used to obtain from the image captured with the associated camera a binary representation with the information related to tracked objects.
- The client application is in charge of fusing all cameras' binary maps in a global 3D grid in a projection process based on a technique called Visual-Hull [10].

This 3D discrete representation is then used as input of the multimodal PF.

The rest of the document is organized as follows: in section II a detailed description of the mentioned observation system is included; in section III the multi-hypothesis tracker based on a PF is described; the global functionality of the proposal is analyzed in section IV; finally section V is used to highlight conclusions and future works observed within the development of the proposal described.

II. Observation Process

In order to develop the 3D tracking proposed, a three-dimensional representation of the environment analyzed is needed. Visual-Hull has appeared as a robust, fast and powerful technique in order to obtain such a representation of a scene observed by a ring of static cameras strategically located [10].

The Visual-Hull is defined as the largest volume enclosed by the intersection of the different visual cones extracted from a set of cameras looking to a common set of silhouettes (see Fig. 2). The obtained volume is not the real one occupied by the objects that generate the set of silhouettes, but it can be ensured that is enclosed.

The 3D Visual-Hull of a scene is generated with the following steps:

- In order to obtain it, a calibrated set of static cameras strategically located in the environment of interest (to avoid objects' occlusions) is needed. Therefore, and supposing a pin-hole model for all cameras, intrinsic parameters (given by K matrix) and extrinsic ones (given by rotation –R–and translation –T– ones) are known for all cameras in the set.

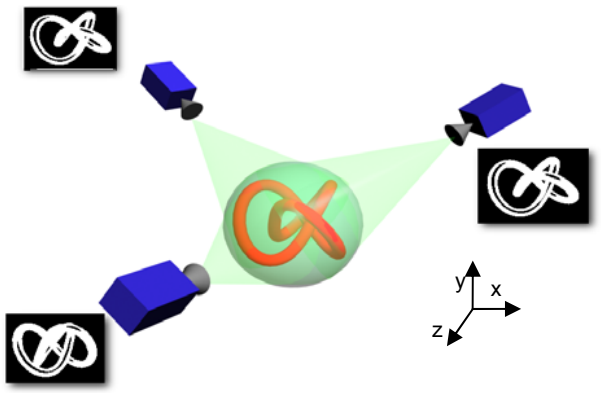

Fig. 2. Graphical representation of the Visual-Hull concept.

- A standard background subtraction process, based on the Mahalanobis distance, is used to segment all silhouettes in the images extracted with every camera in the array. A set of binary representations of all objects in the environment, from different perspectives, is hence obtained.
- All these binary images have to be re-projected in a common space. In the proposed observation system, the 3D space used to reconstruct the global environment is discretized in height (Y coordinate each o meters), in order to have a faster execution of the Visual-Hull process. Therefore, the common re-projection space is a set of Π_b planes in the XZ space, one for each b discrete value of Y. As a result, the re-projection process is characterized by the transformation expressed by eq. 1:

$$\begin{bmatrix} x_{a,b} \\ z_{a,b} \\ 1 \end{bmatrix} = \lambda^{-1} H_{a,b}^{-1} \begin{bmatrix} u_a \\ v_a \\ 1 \end{bmatrix}, \qquad (1)$$

where the pair $\begin{bmatrix} x_{a,b} & z_{a,b} \end{bmatrix}^T$ is the representation in the corresponding Π_b plane of the point given by $\begin{bmatrix} u_a & v_a \end{bmatrix}^T$ in the binary image; and $H_{a,b}$ is the related homography matrix, that can be obtained from the related K_a, R_a and T_a ones as shown:

$$\begin{bmatrix} x_{a,b} \\ z_{a,b} \\ 1 \end{bmatrix} = \lambda \cdot K_a \cdot (R_a \begin{bmatrix} u_a \\ v_a \\ 1 \end{bmatrix} + T_a) \qquad (2)$$

- Once the homography related to each a binary image is obtained for each Π_b plane, they are multiplied to obtain the intersection of all objects in the observed scene in Π_b plane.
- In order to decrease even more the homography process computational load Π_b plane is also discretized in squares (with o meters long per side). Thus, the final 3D representation of all objects in the observed scene is an occupancy grid.

The global process, described in previous paragraphs, is represented in Fig. 3.

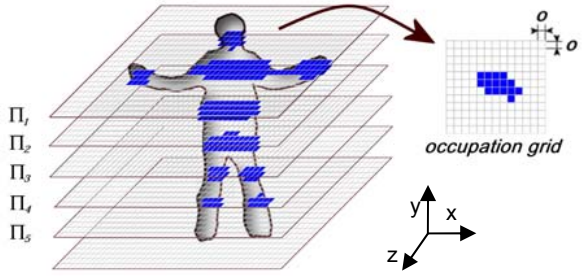

Fig. 3. Functional representation of the Visual-Hull. Several Π_b planes (b=5..1) in different heights are used. The silhouettes resulting of the intersections between the volume and these planes are projected onto a common Π_b plane for all cameras in the 3D world. In the detailed image of the top plane it can be noticed that, each plane surface has been discretized in squares of o meters long per side.

III. TRACKING PROCESS

Two main algorithms have traditionally been proposed in the related literature to solve the state $\vec{x}_t = [x_t \quad y_t \quad z_t]^T$ estimation problem involved in the multi-tracking application of interest: the Kalman Filter (KF), that provides the optimal estimation solution $\hat{\vec{x}}_t = \vec{x}_{t|t}$ given the linear and known probabilistic (necessary Gaussian) model of each object dynamics and the observation system; the PF, that provides a sampled density function of this estimation $p(\vec{x}_t^{(i)} | \vec{y}_{1..t}^{(i)})$ (generally called belief), without restrictions of linearity, unimodality and exactitude in the models description.

In this work, a single PF is used to track the different entities detected and presented in the 3D occupancy grid with the already described observation system. The PF based tracking solution ensures bigger reliability and robustness, and more constant execution time that the one based on the KF (as demonstrated in [8] and [9]).

There are some interesting works in the related literature that propose the use of a unimodal PF for each object to be tracked (i.e. [11], [4]), or a single unimodal one with an extended state vector $\vec{x}_t$ that includes the one of every object to be tracked (i.e. [7]). Both solutions are neither suitable for real-time applications nor for crowded scenes, as the PF execution time increases linearly with the number of samples (also called particles) in the generated belief, and exponentially with their dimension [12].

Besides, using the multimodality of the PF to characterize the different objects state is not straight-forward: the multimodality has to be ensured within all conditions, as each mode in the discretized belief $p(\vec{x}_t^{(i)} | \vec{y}_{1..t}^{(i)})$ generated by the PF will describe the state $\vec{x}_t$ of each object being tracked. An unequal distribution of the samples $\boldsymbol{S}_t \cong p(\vec{x}_t^{(i)} | \vec{y}_{1..t}^{(i)})$ among all elements being tracked will make the estimation solution weak to disturbances and inaccuracies in the observation and dynamic models, making the particle set degenerate.

In order to overcome this problem, a specific version of the standard PF has been developed. Fig. 4 shows the flowchart of the proposed PF for 3D tracking.

The PF proposed is based on the Extended Particle Filter (XPF) firstly described in [13]. The main difference between the standard PF (the Bootstrap version described in [12]) and the XPF consists on inserting a re-initialization step, before the prediction one. This new step modifies the prior belief $p(\vec{x}_{t-1} | \vec{y}_{1..t-1})$ in order to include information related to all objects detected by the observation process in the previous step (given by the probability density function $p(\vec{y}_{t-1})$), through the inclusion of n_m new samples, of the total of n that conform $p(\vec{x}_{t-1} | \vec{y}_{1..t-1})$.

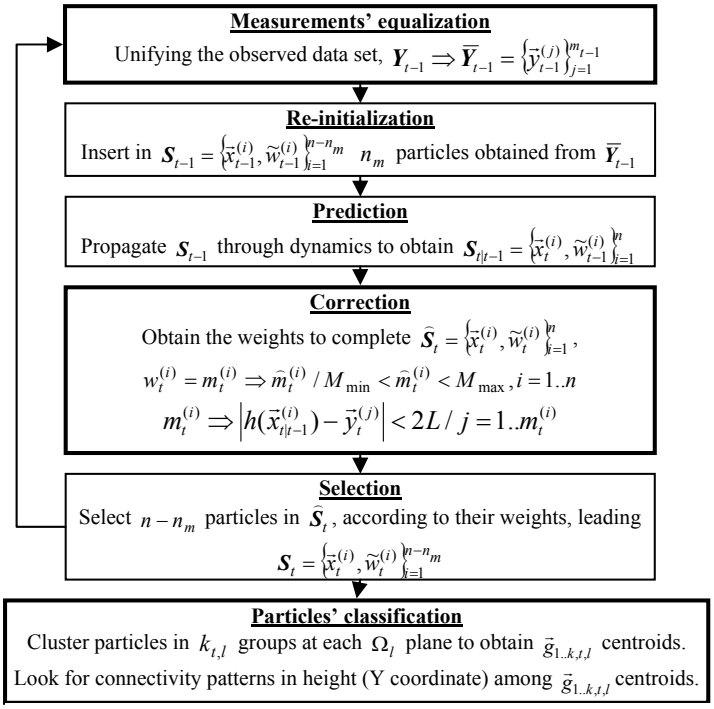

Fig. 4. Flowchart of the proposed PF for 3D tracking.

The XPF is modified in the proposal presented in [8] and [10] in order to increase the robustness of the algorithm multimodality. This is achieved thanks to a measurements' clustering process that is used in the new re-initialization step and in the correction one. The resulting algorithm is called Extended Particle Filter with Clustering Process (XPFCP).

The algorithm proposed in this paper achieves the same robustness replacing the clustering process in the XPFCP by some others (see Fig. 4) with the same objective but faster and more flexible than that one, according to the real-time and three-dimensional character of the pursuit application.

The different functional details of the PF proposed and shallowly described and remarked with thick lines in Fig. 4 are explained in the following subsections.

A. MEASUREMENTS' EQUALIZATION

A measurements' equalization process is developed previously to the re-initialization step, with the objective of filtering useless information from the data set $\boldsymbol{Y}_{t-1}$ got from the observation system. The equalized set of measurements $\overline{\boldsymbol{Y}}_{t-1}$ will be therefore used in the re-initialization step in a similar way as it is in the XPF.

The equalization procedure consists on decreasing the 3D grid density in order to make it similar all over the space of interest. To achieve this equalization a histogram thresholding is developed at each Π_b plane in the global 3D grid obtained from the observation step, as shown in Fig. 5.

This equalization process improves the robustness in the two steps that uses the generated data set (see Fig. 4): the re-initialization and the correction one.

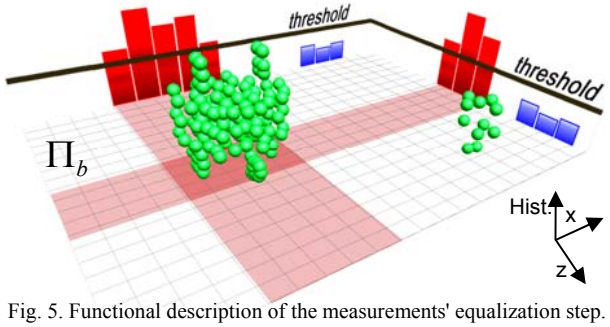

Fig. 5. Functional description of the measurements' equalization step. Zones with redundant measurements, and are therefore erased with the normalization, have been colored in red.

This technique therefore implies a better distribution of the particles in the belief among all hypotheses to be tracked, contributing in reducing the harmful particle set degeneration.

Besides, the execution time of these two mentioned steps is also decreased as it is linearly related to the number of measurements in the observation data set.

B. PARTICLES' WEIGHTING: CORRECTION STEP

The algorithm proposed at the correction step consists on giving a weight $w_t^{(i)}$ to each particle $\vec{x}_t^{(i)}$ according to the number $m_t^{(i)}$ of equalized observations ($\overline{\boldsymbol{Y}}_t$) in the 3D grid that are near (within a Manhattan distance of *2L* meters) a particle when it ($\vec{x}_{t|t-1}^{(i)} / i = 1..n$) is projected to the observation space, through the observation model, $h(\vec{x}_{t|t-1}^{(i)})$.

Once again, to decrease the harmful effect of unequally sensed objects in the scene, $m_t^{(i)}$ is saturated to $\bar{m}_t^{(i)}$ as shown:

$$M_{\min} < \bar{m}_t^{(i)} < M_{\max}, i = 1..n, \qquad (3)$$

where $M_{\min}$ and $M_{\max}$ are respectively obtained weighting the number of measurements that may be inside a block defined by *L* (that is L^3/o^3) by $\alpha 1$ and $\beta 1$. It is, therefore, taken into account the discretization made in the height coordinate Y during the observation process.

C. PARTICLES' CLASSIFICATION

The output generated by the 3D PF described in previous subsections is a discretized version of the belief $\boldsymbol{S}_t \cong p(\vec{x}_t | \vec{y}_{1..t})$, representing distribution of hypotheses or objects being tracked in the state space.

Consequently, a deterministic output has to be obtained from this probabilistic one, informing about the number of objects finally detected and their track.

In order to obtain the deterministic solution from the 3D PF output, a particles' clustering process is developed in this work in three steps:

1. The proposed classifier is again performed in parallel XZ planes. Therefore the first step is to discretize the state space along the Y coordinate, projecting all particles in $\boldsymbol{S}_t$ to the nearest XZ plane Ω_l.
2. An extended K-means clustering algorithm is then applied at each Ω_l to obtain a set of particle's clusters characterized by the group of $k_{t,l}$ centroids $\vec{g}_{1..k,t,l} = [x_{1..k,t,l} \quad y_{1..k,t,l} \quad z_{1..k,t,l}]^T$, where height values are discrete. These 3D points are, in fact, the estimated position of the mass center of whole objects or parts of them at each height section.
3. Finally, a boolean and bidirectional connectivity process is included in order to join each $\vec{g}_{c,t,l1} / c = 1..k_{t,l1}$ at plane Ω_{l1} with any (zero, one or more) $\vec{g}_{c,t,l2} / c = 1..k_{t,l2}$ at a contiguous plane Ω_{l2} (defined for a discrete height next to the one of Ω_{l1}). An euclidean distance is used as connectivity variable, parameterized by threshold *D* and including a scaling factor (through $\alpha 2$ for Z and $\beta 2$ for Y).

Each connected set of centroids resulting at the end of this process is considered to be a specific object in the environment of interest. Thanks to the connectivity process non rigid objects with different joints can be properly distinguished and track in real-time and all over the IS analyzed by the observation system.

Fig. 6 shows a functionality example of the connectivity process described in previous paragraphs.

IV. RESULTS

The global 3D tracking process proposed in this paper has been deeply tested over real observation situations.

All tests have been done in Matlab, obtaining the observation data set from a real IS conformed by 4 color cameras acquiring images at 15fps with 640x480 pixels of resolution. The size of the observed space in the IS is about 3.5x3.5m. In all examples shown a constant speed model define the dynamics and $h()$, $n = 1500$ and $n_m = 15\%$. Besides $o = 0.04m$, the threshold is set to 40 measurements and $L = 0.1m$ in the observation, equalization and correction processes, respectively.

Fig. 7 shows the effect of the proposed measurements' equalization process, in a piece of an experiment of up to 5 agents (robots and persons) that get into and out of the IS.

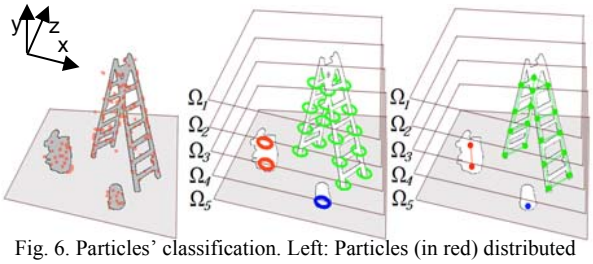

Fig. 6. Particles' classification. Left: Particles (in red) distributed among the scene objects. Middle: Clusters generated by the particles' K-means classification, once they are projected onto planes Ω_l (*l*=5..1) of different height. Right: Segmenting different entities (in colors) after applying the connectivity algorithm to cluster centroids.

The 2 plots in Fig. 7 present the comparison between the real number (#) of objects extracted from the experiment (ground truth, in whole green line) versus the # of correctly tracked by the proposed algorithm (in dash blue line). The equalization process was applied in the image at the bottom and not in the one at the top.

Comparing the 2 plots in Fig. 7, it can be easily noticed that the reliability of the global 3D tracking improves when the measurements' equalization process is used: it rises from a 55% of iterations without error to a 95%.

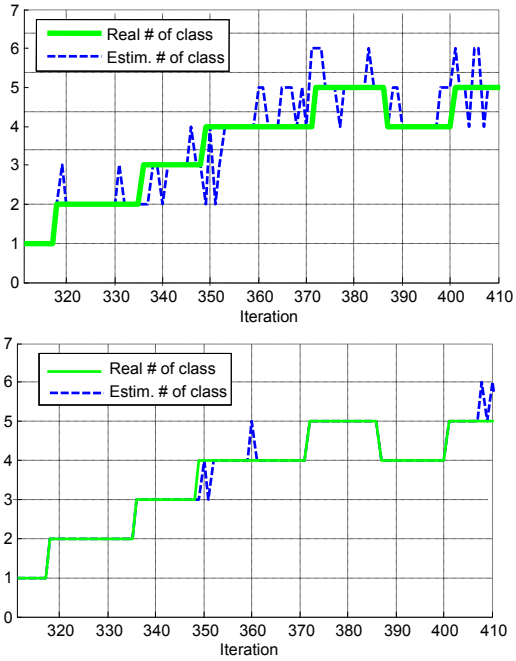

Fig. 7. System reliability. Top: comparison between the real number (#) of classes and the # of the output classes, without measurements' equalization. Bottom: same test, using measurements' equalization.

Fig. 8 shows the effect of the saturation process at the correction step. The left image in the figure shows the set of n particles' weights histogram with $\alpha 1 = 0.1$, $\beta 1 = Inf$, and the right one the same histogram with $\beta 1 = 3.5$. It can be noticed that particles show a more uniform distribution of weights $w_t^{(1..n)}$ in the histogram at the right, that is, if the saturation process is included.

As a global result, Fig. 9 shows the 3D tracker proposed functionality in different instants extracted from another real experiment (see the top row of images), that starts with a robot getting into the IS, continues with a person putting three dustbins down in the floor and getting out after, and finishes with the same person leaving a ladder among the dustbins and leaving the room again.

In order to obtain these results the final classification process parameters' have been tuned as follows: 2D K-means has been run in XZ planes each 40cm and with a gating distance of 30cm; $D = 0.5m$, $\alpha 2 = 1$ and $\beta 2 = 1.4$.

Analyzing Fig. 9, it can be concluded that the global proposal successfully tracks all objects in the scene at every moment of the experiment:

- The 3D occupancy grid generated by the proposed observation system (see the middle row of images) represents with a high level of exactitude the different situations that happen in the IS.

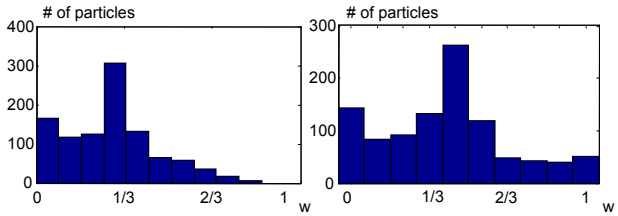

Fig. 8. Effect of the weight saturation step: weight (w) histograms before (left) and after (right) the saturation.

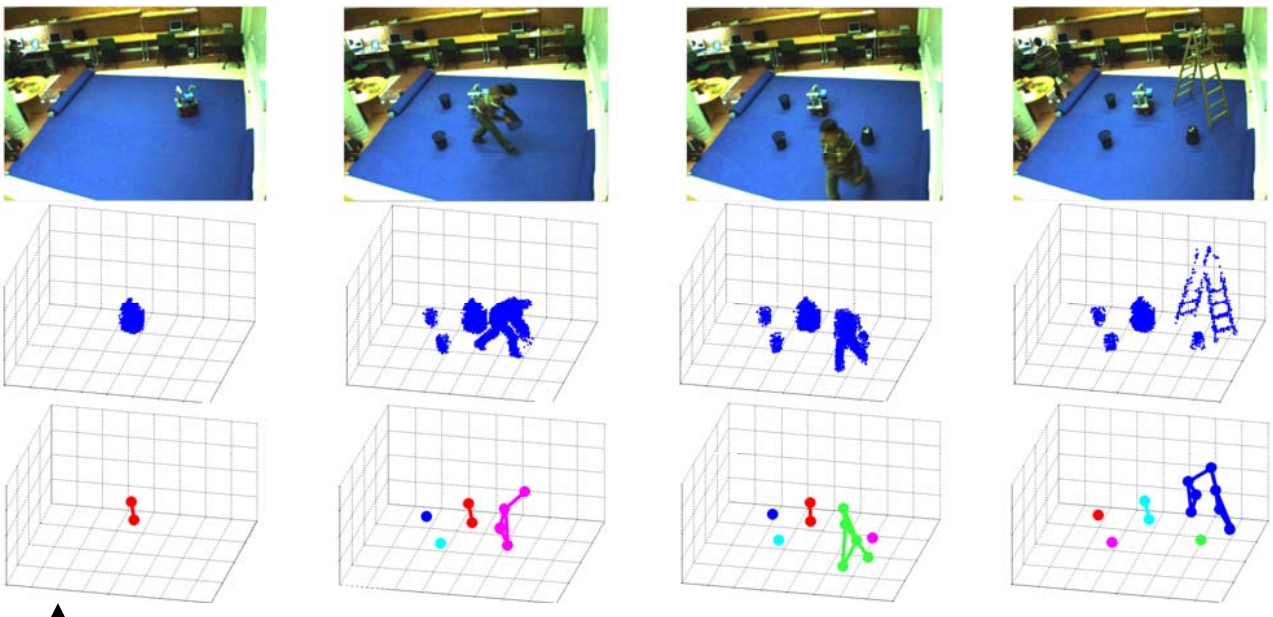

Fig 9. Real sequence obtained from one of the test videos. Top: real images captured from the cameras. Center: Visual Hull volumetric reconstruction. Bottom: 3D Tracker output classes; each tracked class is identified with a different color.

- The deterministic output generated by the 3D PF (see the lower row of images) tracks individually (with different colours) and reliably the different objects detected by the observation system at every moment in the scene.

These results demonstrate the correct functionality of the 3D tracking proposal in a dynamic environment occupied by flexible objects of different sizes and movement patterns.

V. Conclusions and Future Works

In this paper, an algorithm to solve the 3D tracking problem in an intelligent space is proposed and successfully tested. In order to obtain the global functionality, a Visual-Hull technique is used as observation system, and a modification of the XPF is used in the estimation step. Real results demonstrate both the effect and the correct functionality of the different processes included in the global solution.

The authors are already working in the real-time implementation of the proposal in order to complete the demonstration of its online applicability. Besides, a most developed background subtraction technique has also to be included in the observation system proposed, in order to make it robust to big lighting changes.

Finally, an identification process, similar to the one already proposed by the authors in [8] (for a 2D tracker in an IS), is also to be incorporated in the system presented in this paper.

Acknowledgment

This work was supported jointly by the Ministry of Science and Technology under the projects RESELAI (reference TIN2006-14896-C02-01) SD-TEAM (reference TIN2008-06856-C05-01).

References

[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", *Proceedings of the 7th International Symposium of Hungarian Researchers on Computational Intelligence,* 2006.

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

[4] 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.

[5] 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.

[6] 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.

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

[8] 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", *Proceedings of the 2008 IEEE International Conference on Robotics and Automation,* 2008.

[9] 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 Symposium on Industrial Electronics,* 2005.

[10] 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.

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

[12] 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.

[13] E. B. Koller-Meier, F. Ade, "Tracking multiple objects using a condensation algorithm", *Journal of Robotics and Autonomous Systems, Vol. 34, pp. 93-105,* 2001.