UNIVERSIDAD DE ALCALÁ
ESCUELA POLITÉCNICA SUPERIOR
DEPARTAMENTO DE ELECTRÓNICA

MOTION SEGMENTATION AND 3D POSITIONING OF MOBILE ROBOTS IN AN INTELLIGENT SPACE USING AN ARRAY OF STATIC CAMERAS

Cristina Losada Gutiérrez

Ph.D. Thesis Extended Abstract

Abstract

This document presents a summary of the thesis "*Motion segmentation and 3D positioning of mobile robots in an intelligent space, using an array of static cameras*" developed by the PhD candidate Cristina Losada. The main objective of the thesis is to obtain the motion segmentation and 3D localization of multiple mobile robots in an intelligent space using a multi-camera sensor system. The set of calibrated and synchronized cameras are placed in fixed positions within the environment (intelligent space). The proposed algorithm for motion segmentation and 3D localization does not rely on previous knowledge or invasive landmarks on board the robots. It is based on the minimization of an objective function that combines information from all the cameras. The proposed objective function depends on three groups of variables: the segmentation boundaries, the motion parameters and the depth. For the objective function minimization, we use a greedy iterative algorithm with tree steps that, after initialization of segmentation boundaries and depth, are repeated until convergence.

1. Introduction and thesis objectives

A common problem in the field of autonomous robots is to obtain the position and orientation of the robots within the environment with sufficient accuracy. Several methods have been developed to carry out this task. The localization methods can be classified into two groups: those that require sensors on board the robots and those that incorporate sensors within the work environment [Pizarro et al. 2009]

The use of sensors within the environment presents several advantages: it allows reducing the complexity of the electronic onboard the robots and facilitates simultaneous navigation of multiple mobile robots within the same environment without increasing the complexity of the infrastructure. Moreover, the information obtained from the robots movement is more complete, thereby it is possible to obtain information about the position of all of the robots, facilitating the cooperation between them. This alternative includes intelligent environments [Lee et al. 2001], [Steinhaus et al. 2004] characterized by the use of an array of sensors located in fixed positions and distributed strategically to cover the entire field ofmovement of the robots. The information

provided by the sensors should allow the localization of the robots and other mobile objects accurately.

The sensor system in this thesis is based on an array of calibrated and synchronized cameras. There are several methods to locate mobile robots using an external camera array. The most significant approaches can be divided into two groups, depending on the previous knowledge about the robots that is required by the method. The first group includes those works that make use of strong prior knowledge by using artificial landmarks attached to the robots [Sogo et al. 1999], [Fernandez et al. 2007]. The second group includes the works that use the natural appearance of the robots and the camera geometry to obtain the positions [Pizarro et al. 2009].

The proposal presented in this thesis is included in the second group. It uses a set of calibrated cameras, placed in fixed positions within the environment to obtain the position of the robots and their orientation. This proposal does not rely on previous knowledge or invasive landmarks. Robots segmentation and position are obtained through the minimization of an objective function. There are many works that use an objective function [Sekkati & Mitiche 2006a], [Sekkati and Mitiche 2006b]. However, these works present several disadvantages such as high computational cost or reliance on the initial values of the variables. Moreover, these methods are not robust because they use information from a single camera.

In an intelligent space, two types of agents can be found: controlled agents (mobile robots) and uncontrolled ones (potential users and obstacles). The proposed solution allows the location of mobile elements, even if they are not robots, however, this thesis is focused on the mobile agents that are controlled by the intelligent space. For this reason, in this document we will only refer to mobile robots.

The solution proposed in this thesis allows the detection, segmentation and 3D positioning of a variable number of mobile robots, in an intelligent space. It does not rely on previous knowledge or invasive landmarks on board the robots.

The mobile robots we want to segment are equipped with odometric sensors onboard. The measurements of these sensors are transmitted via radio to the intelligent space. Thus, if it is necessary, the odometric sensor information can be included in the algorithm for motion segmentation and 3D positioning, in order to improve its performance.

The sensor system used in this work is based on a set of calibrated and synchronized cameras placed in fixed positions within the environment (Intelligent Space of University of Alcalá, ISPACE-UAH). These cameras are distributed strategically to cover the entire field of movement of the robots. The software architecture chosen is a client-server system using common TCP/IP connections, where some servers receive commands and requests from a client. Figure 1 shows a general diagram of the proposed hardware/software architecture.

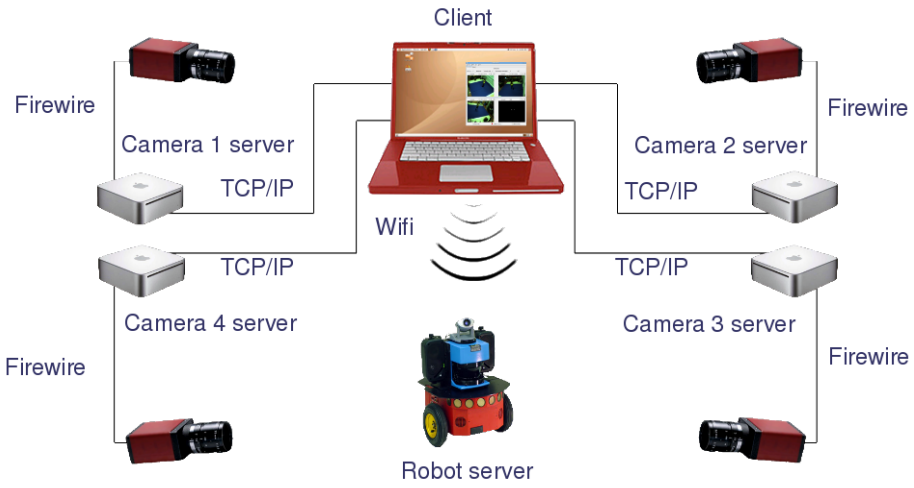


Figure 1. General diagram of the hardware/software architecture in the ISPACE-UAH.

2. Proposed system for motion segmentation and 3D positioning

In order to achieve the objectives proposed in this thesis, under the conditions described above, it is necessary to carry out a set of tasks. These tasks allow us to solve the different problems that arise within the line of research in this thesis. The ultimate goal of all tasks is the development of a comprehensive system that allows obtaining the motion segmentation and 3D positioning of multiple mobile robots, in an intelligent space, with enough accuracy. The proposed solution must be robust against lighting changes, shadows or occlusions in the images acquired by the different cameras in the intelligent space. Figure 2 shows a general block diagram that includes the different stages involved in the process.

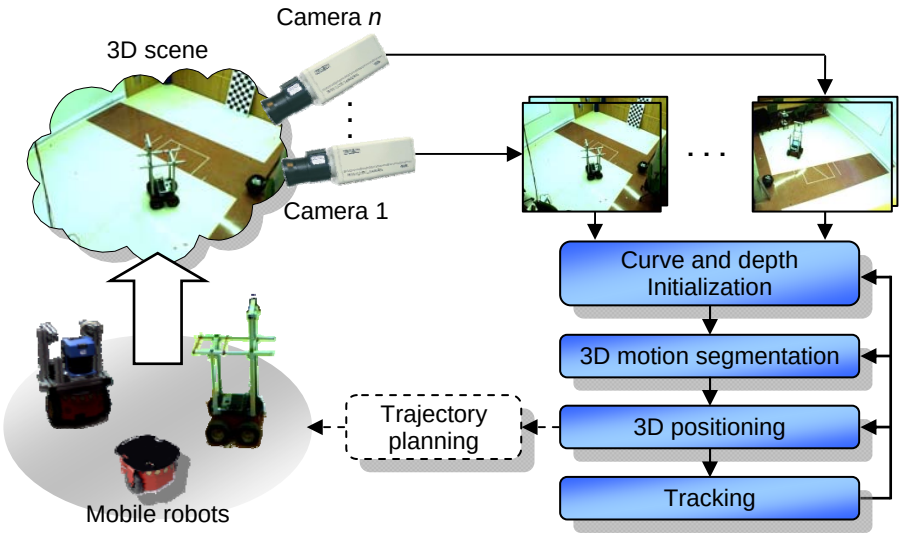


Figure 2. General block diagram of the proposed method for motion segmentation and 3D positioning of multiple mobile robots in an intelligent space.

Using the work of Sekkati and Mitiche [Sekkati & Mitiche 2006b] as a starting point, in this work motion segmentation and 3D localization are obtained through the minimization of an objective function. The objective function proposed in [Sekkati & Mitiche 2006b] (shown in equation (1)) depends on three groups of variables: a set of curves that defines the mobile robot segmentation boundaries in the image plane $\{\gamma_k\}_{k=1}^{N-1}$, the components of linear and angular velocity of each robot $\{\mathbf{v}_{ck}\}_{k=1}^{N}$, $\{\boldsymbol{\omega}_{ck}\}_{k=1}^{N}$ and the depth. In equation (1) λ y μ are positive, real constants. These constants weight the contribution of each term to the objective function.

$$E\left[\{\gamma_k\}_{k=1}^{N-1}, \{\mathbf{v}_{ck}\}_{k=1}^{N}, \{\boldsymbol{\omega}_{ck}\}_{k=1}^{N}, Z\right] = \sum_{k=1}^{N}\left[\int_{\Omega_k}\psi_k^2(\mathbf{x})d\mathbf{x} + \mu\int_{\Omega_k} g\left(\|\nabla Z\|\right)d\mathbf{x}\right] + \sum_{k=1}^{N-1}\lambda\oint_{\gamma_k} ds\,, \qquad (1)$$
$$\lambda, \mu \in \Re \quad \lambda, \mu > 0$$

As can be observed in equation (1), the objective function proposed by Sekkati and Mitiche in [Sekkati & Mitiche 2006b] contains three different terms. The first term measures the conformity of the 3D interpretation within each region of segmentation to the image sequence spatiotemporal variations. This measure is given by the three-dimensional brightness constraint for rigid objects proposed in [Sekkati & Mitiche 2006b] and shown in equation (2). The remaining two terms in equation (1) are regularization terms, one for depth via a boundary preserving function (g(a)) and the other one for segmentation boundaries.

$$I_t + \mathbf{s}\frac{\mathbf{v}_c}{Z_c} + \mathbf{q}\boldsymbol{\omega}_c = 0 \qquad (2)$$

In equation (2), **s** and **q** are two vectors that depend on the image spatiotemporal derivatives [I_x, I_y, I_t], the coordinates of each point in the image plane (x, y) and the focal lengths f_x, f_y.

$$\mathbf{s} = \begin{pmatrix} f_x I_x \\ f_y I_y \\ -(x-s_1)I_x - (y-s_2)I_y \end{pmatrix}^T \quad \mathbf{q} = \begin{pmatrix} -f_y I_y - \frac{y-s_2}{f_y}\left((x-s_1)I_x + (y-s_2)I_y\right) \\ -f_x I_x - \frac{x-s_1}{f_x}\left((x-s_1)I_x + (y-s_2)I_y\right) \\ -\frac{f_x}{f_y}(y-s_2)I_x + \frac{f_y}{f_x}(x-s_1)I_y \end{pmatrix}^T$$

In [Sekkati & Mitiche 2006b], the minimization of the objective function (1) is carried out using a greedy algorithm that consists of three iterated steps. After the initialization of the segmentation boundaries and depth, the three steps are repeated until the convergence of the algorithm. In each step, two of the three groups of variables are fixed, and the equation is solved for the remaining one. After minimization, motion segmentation of the mobile robots is obtained. However, proposal of [Sekkati & Mitiche 2006b] presents several disadvantages such as high computational cost, or reliance on the initial values of the variables (segmentation boundaries and depth). Moreover, this method is not robust, and it does not allow obtaining 3D position of the mobile robots because it uses information from a single camera.

Since there are multiple cameras available in the intelligent space, we have proposed a new objective function that includes information of all the cameras. The minimization of the proposed function allows us to obtain both motion segmentation and 3D position of multiple mobile robots in an intelligent space. The use of multiple cameras increases notably the robustness of the system. It also improves the accuracy of the results (segmentation and 3D positioning).

2.1. Objective function for a Multi-camera Sensor System

The coordinates of a point $\mathbf{P} = (X, Y, Z)^T$ can be expressed in the different coordinate systems in ISPACE-UAH: world, robot and camera coordinate systems. It is important to define the different coordinate systems and its relationships to obtain the objective function for the multi-camera sensor system. There is a global reference system named “world coordinate system” and represented by Γ_w. There is also a local reference system associated with each camera (Γ_{ci}, i=1,…,n_c) whose origin is located in the center of projection. These coordinate systems are represented in Figure 3, where

the world coordinate system (Γ_w) has been represented in red color and the coordinate systems associated to the cameras (Γ_{ci}) have been represented in blue.

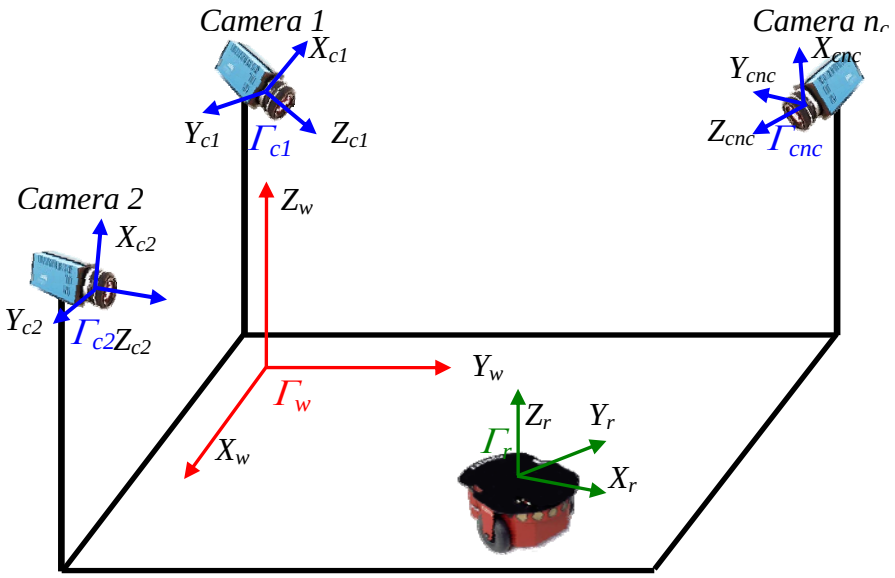


Figure 3. Coordinate systems in the intelligent space (ISPACE-UAH): World coordinate system (Γ_w) in red color. Camera coordinate system (Γ_{ci} i = 1,2,…,n_c) in blue color. Robot coordinate system (Γ_r) in green color.

Cameras are modelled as pinhole cameras. This is a simple model that describes the mathematical relationship between the coordinates of a 3D point in the camera coordinate system (Γ_c) and its projection onto the image plane in an ideal camera without lenses through the expressions in equation (3) where f_x and f_y are the camera focal lengths along x and y axis.

$$x = f_x \frac{X_c}{Z_c}, \quad y = f_y \frac{Y_c}{Z_c} \tag{3}$$

If the origin of the image coordinate system is not in the center of the image plane, the displacement (s_1,s_2) from the origin to the center of the image plane is included in the projection equations, obtaining the perspective projection equation (4).

$$x = f_x \frac{X_c}{Z_c} + s_1, \quad y = f_y \frac{Y_c}{Z_c} + s_2 \tag{4}$$

These equations can be expressed using homogeneous coordinates, as shown in equation (5).

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} f_x & 0 & s_1 \\ 0 & f_y & s_2 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X_c \\ Y_c \\ Z_c \end{pmatrix} \tag{5}$$

Before presenting the objective function for multiple cameras, it is necessary to describe the 3D brightness constraint for multiple cameras, that is a generalization of the 3D brightness constraint for a single camera presented in [Sekkati & Mitiche 2006b].

Let $\mathbf{P}_w=(X_w, Y_w, Z_w)^T$ be the 3D coordinates of point **P** on a mobile robot related to the world coordinate system Γ_w. Let $\mathbf{v}_w = \left(v_w^x\ v_w^y\ v_w^z\right)^T$ and $\boldsymbol{\omega}_w = \left(\omega_w^x\ \omega_w^y\ \omega_w^z\right)^T$ be, respectively, the components of the linear and angular velocity of the robot motion in Γ_w. Then, the velocity of **P**, relative to Γ_w, is given by equation (6).

$$\dot{\mathbf{P}}_w = \left(\dot{X}_w \quad \dot{Y}_w \quad \dot{Z}_w\right)^T = \mathbf{v}_w + \boldsymbol{\omega}_w \times \mathbf{P}_w \tag{6}$$

In the same way, if $\mathbf{P}_c=(X_c, Y_c, Z_c)^T$ are the coordinates of **P** relative to Γ_c and $\mathbf{v}_c = \left(v_c^x\ v_c^y\ v_c^z\right)^T$ and $\boldsymbol{\omega}_c = \left(\omega_c^x\ \omega_c^y\ \omega_c^z\right)^T$ are the components of the linear and angular velocity of the robot motion in Γ_c. The velocity of **P** relative to Γ_c is given by equation (7):

$$\dot{\mathbf{P}}_c = \left(\dot{X}_c \quad \dot{Y}_c \quad \dot{Z}_c\right)^T = \mathbf{v}_c + \boldsymbol{\omega}_c \times \mathbf{P}_c \tag{7}$$

Let $\mathbf{R}_{wc}$ be the (3x3) rotation matrix and $\mathbf{T}_{wc}$ the (1x3) translation vector that represent the coordinate transformation from the world coordinate system (Γ_w) to the camera coordinate system (Γ_c). The coordinate transformation is carried out using the expression in equation (8).

$$\mathbf{P}_c = \mathbf{R}_{wc}\mathbf{P}_w + \boldsymbol{T}_{wc} \tag{8}$$

Deriving the equation (8) with respect to time, and substituting the expressions of the velocities in Γ_w (equation (6)) and Γ_c (equation (7)), equation (9) is obtained.

$$\mathbf{v}_c + \boldsymbol{\omega}_c \times \mathbf{P}_c = \mathbf{R}_{wc}\left(\mathbf{v}_w + \boldsymbol{\omega}_w \times \mathbf{P}_w\right) \tag{9}$$

Taking into account that cross product $\boldsymbol{\omega} \times \mathbf{P}$ can be expressed as a scalar product $\hat{\boldsymbol{\omega}} \cdot \mathbf{P}$, where $\hat{\boldsymbol{\omega}}$ is the following antisymmetric matrix:

$$\hat{\boldsymbol{\omega}} = \begin{pmatrix} 0 & -\omega^z & \omega^y \\ \omega^z & 0 & -\omega^x \\ -\omega^y & \omega^x & 0 \end{pmatrix}$$

equation (9) can be rewritten to obtain equation (10), where the components of linear and angular velocities in Γ_c ($\mathbf{v}_c$, $\boldsymbol{\omega}_c$) are expressed as a function of the components of velocity in Γ_w ($\mathbf{v}_w$, $\boldsymbol{\omega}_w$) and the transformation matrices ($\mathbf{R}_{wc}$, $\mathbf{T}_{wc}$).

$$\begin{aligned}\mathbf{v}_c &= \mathbf{R}_{wc}\mathbf{v}_w - \mathbf{R}_{wc}\boldsymbol{\omega}_w\mathbf{R}_{wc}^T\mathbf{T}_{wc} \\ \boldsymbol{\omega}_c &= adj(\mathbf{R}_{wc})\boldsymbol{\omega}_w\end{aligned} \tag{10}$$

Let (x, y) be the coordinates of the projection of a point **P** on the image plane, the derivative of the perspective projection equations (equation (4)) with respect to time, and the subsequent substitution of the expression of the velocity components of **P** in Γ_c allows us to obtain the following equations for motion components in the image plane $(\dot{x}, \dot{y})$:

$$\dot{x} = \tfrac{1}{Z_c}\left(f_x R_{wc}^1 - xR_{wc}^3\right)\mathbf{v}_w + \mathbf{q}_u adj(\mathbf{R}_{wc})\boldsymbol{\omega}_w \tag{11}$$

$$\dot{y} = \tfrac{1}{Z_c}\left(f_y R_{wc}^2 - yR_{wc}^3\right)\mathbf{v}_w + \mathbf{q}_v adj(\mathbf{R}_{wc})\boldsymbol{\omega}_w \tag{12}$$

where R_{wc}^i is the *i*-th row in the rotation matrix from Γ_w to Γ_c ($\mathbf{R}_{wc}$) and $\mathbf{q}_u$, $\mathbf{q}_v$ are the following vectors:

$$\mathbf{q}_u = \left[x\left(\frac{t_{wc}^y}{Z_c} - \frac{y}{f_y}\right) \quad \left(f_x + \frac{x^2}{f_x} - \frac{1}{Z_c}\left(f_x t_{wc}^z + x t_{wc}^x\right)\right) \quad f_x\left(\frac{t_{wc}^y}{Z_c} - \frac{y}{f_y}\right)\right]$$

$$\mathbf{q}_v = -\left[\left(f_y + \frac{y^2}{f_y} - \frac{1}{Z_c}\left(f_y t_{wc}^z + y t_{wc}^y\right)\right) \quad y\left(\frac{t_{wc}^x}{Z_c} - \frac{x}{f_x}\right) \quad f_y\left(\frac{t_{wc}^x}{Z_c} - \frac{x}{f_x}\right)\right]$$

The substitution of the velocity components in the image plane $(\dot{x}, \dot{y})$ in the well known brightness constraint ($I_x\dot{x} + I_y\dot{y} + I_t = 0$) allows to obtain a 3D brightness constraint for rigid objects in terms of the linear and angular velocity components in Γ_w ($\mathbf{v}_w$ and $\boldsymbol{\omega}_w$). This constraint is shown in equation (13).

$$\Psi_k(\mathbf{x}) = I_t + \mathbf{s}\cdot\mathbf{R}_{wc}\frac{\mathbf{v}_w}{Z_c} + \mathbf{q}\cdot adj(\mathbf{R}_{wc})\boldsymbol{\omega}_w + T_{wc}^T\mathbf{r}\cdot adj(\mathbf{R}_{wc})\frac{\boldsymbol{\omega}_w}{Z_c} = 0 \tag{13}$$

where the matrices **s, q** and **r** in equation (13) are given, respectively, by equations (14), (15) and (16):

$$\mathbf{s} = \begin{pmatrix} f_x I_x & f_y I_y & -(xI_x + yI_y)\end{pmatrix} \tag{14}$$

$$\mathbf{q} = \begin{pmatrix} -f_v I_y - \frac{y}{f_v}(x\cdot I_x + y\cdot I_y) & f_u I_x + \frac{x}{f_u}(x\cdot I_x + y\cdot I_y) & -\frac{f_u}{f_v}y\cdot I_x + \frac{f_v}{f_u}x\cdot I_y\end{pmatrix} \tag{15}$$

$$\mathbf{r} = \begin{pmatrix} 0 & -(xI_x + yI_y) & -f_v I_y \\ (xI_x + yI_y) & 0 & f_u I_x \\ f_v I_y & -f_u I_x & 0\end{pmatrix} \tag{16}$$

3D brightness constraint in equation (13) must be satisfied in all of the n_c cameras. Knowing it, we define a new 3D brightness constraint for rigid objects which includes all the information provided by the n_c cameras available in the intelligent space (equation (17)).

$$\psi_{ki}(\mathbf{x}) = I_{ti} + \mathbf{s}_i \cdot \mathbf{R}_{wci} \cdot \frac{\mathbf{v}_{wk}}{Z_{ci}} + \mathbf{q}_i \cdot adj(\mathbf{R}_{wci})\boldsymbol{\omega}_{wk} + \mathbf{t}_{wci}^T \mathbf{r}_i \cdot adj(\mathbf{R}_{wci})\frac{\boldsymbol{\omega}_{wk}}{Z_{ci}} \qquad (17)$$
$$k = 1,2,\ldots,N; \quad i = 1,2,\ldots,n_c$$

Constraint in equation (17) is defined for each region, in each camera. If there are *N*-1 robots in a scene, the scene is divided into *N* regions (region *N* corresponds to the background). We have added two subscripts to denote a region: subscript *k* (*k*=1,2,…,*N*), which indicates the region in each image, and subscript *i* (*i*=1,2,…,n_c) which indicates the camera. It is worth pointing out that the components of the linear and angular velocity in the world coordinate system do not include the subscript *i* to indicate the camera because these velocities are equal for the n_c cameras.

The objective function for the multi-camera sensor system proposed in this work, equation (18), depends on three groups of variables:

- A set of *N*-1 curves $\{\gamma_{ki}\}_{k=1,\ldots,N-1}^{i=1,\ldots,n_c}$ that divide each image in *N* regions. These curves define the boundaries of the segmentation in the images acquired by each camera.
- The components of the linear and angular velocities $\{\mathbf{v}_{wk}\}_{k=1}^{N}$, $\{\boldsymbol{\omega}_{wk}\}_{k=1}^{N}$ of the (*N*-1) mobile robots and background. These velocities are related to the world reference system Γ_w and are equal for the n_c cameras.
- The depth (distance from each 3D point **P** to each camera). The value of depth in each point coincides with the Z_{ci} coordinate of the point **P** related to the coordinate system of the camera *i* Γ_{ci}.

$$E\left[\{\gamma_{ki}\}_{k=1,\ldots,N-1}^{i=1,\ldots,n_c}, \{\mathbf{T}_{wk}\}_{k=1}^{N}, \{\boldsymbol{\omega}_{wk}\}_{k=1}^{N}, \{Z_{ci}\}_{i=1}^{n_c}\right] =$$
$$\sum_{k=1}^{N}\sum_{i=1}^{n_{ck}}\left[\int_{\Omega_{ki}} \psi_{ki}^2(\mathbf{x})d\mathbf{x} + \mu\int_{\Omega_{ki}} g\left(\left\|\nabla Z_{ci}\right\|\right)d\mathbf{x}\right] + \sum_{k=1}^{N-1}\sum_{i=1}^{n_{ck}}\lambda\oint_{\gamma_{ki}} ds \qquad (18)$$

In equation (18), ψ_{ki} is the 3D brightness constraint (defined in equation (17)) for the pixels inside the curve *k* in the image acquired by the camera *i*; λ y μ are positive

and real constants to weigh the contribution of the terms in the objective function (18) and $\nabla = (\partial_x, \partial_y)$ is the spatial gradient operator.

As in the objective function for one camera (equation (1)), the first term in (18) measures the conformity of 3D interpretation to the sequence spatiotemporal variations in each region through the 3D brightness constraint for a multi-camera sensor system. The second integral is a regularization term of smoothness of depth, and the third integral is a regularization term of the *N*-1 boundaries.

2.2. Objective function minimization

The objective function in equation (18) includes information of all the cameras in the intelligent space. In this work, objective function minimization is carried out using a greedy algorithm that, after the initialization of the variables, consists of three iterative steps. Before the minimization, it is necessary to initialize the curves that define the contours of the segmentation and depth in the images acquired by each camera. Both, the initialization process and the minimization algorithm are explained below.

a) Obtaining 3D motion parameters by least-squares method

In the first step the curves that define the contours of the *N*-1 mobile robots on the scene acquired by each camera are fixed. The value of the depth (Z_c) of each point on each camera image plane is also fixed. So, the energy to minimize, defined as a function of the components of the linear and angular velocities related to Γ_w, reduces to (19).

$$E\left(\{\mathbf{v}_{wk}\}_{k=1}^{N}, \{\boldsymbol{\omega}_{wk}\}_{k=1}^{N}\right) = \sum_{k=1}^{N}\sum_{i=1}^{n_c}\int_{\Omega_{ki}} \psi_{ki}^2(\mathbf{x})d\mathbf{x} \tag{19}$$

Since the 3D brightness constraint for n_c cameras (17) depends linearly on $\mathbf{v}_{wk}$ and $\boldsymbol{\omega}_{wk}$, 3D motion parameters may be obtained using the linear least squares method.

b) Estimating the depth using the gradient descent method

In the second step, the function to be minimized in order to recover the depth is shown in (20). In this function, χ_{ki} is the characteristic function of region *k* in the image acquired by the camera *i* (Ω_{ki}).

$$E(Z) = \sum_{k=1}^{N}\sum_{i=1}^{n_c}\int_{\Omega_{ki}}\left[\psi_{ki}^2(\mathbf{x}) + \mu g\left(\|\nabla Z_{ci}\|\right)\right]d\mathbf{x} = \int_{\Omega}\sum_{k=1}^{N}\sum_{i=1}^{n_c}\left[\chi_{ki}(\mathbf{x})\left(\psi_{ki}^2(\mathbf{x}) + \mu g\left(\|\nabla Z_{ci}\|\right)\right)\right]d\mathbf{x} \tag{20}$$

Given a set of curves $\{\gamma_{ki}\}_{k=1}^{N-1}$ that divides the image acquired by each of the n_c cameras into N regions $\{\Omega_{ki}\}_{k=1}^{N}$, the derivative of (20) with respect to the depth, for each region k in each of the available cameras is given by (21).

$$\frac{\partial E}{\partial Z_{ci}} = \sum_{k=1}^{N} \chi_{ki} \left[\frac{-2}{Z_{ci}^2} \left(\mathbf{S}_i \mathbf{v}_{wk} + \mathbf{R}_i \boldsymbol{\omega}_{wk} \right) \psi_{ki} - \mu div \left(\frac{g'\left(\left\| \nabla Z_{ci} \right\| \right)}{\left\| \nabla Z_{ci} \right\|} \nabla Z_{ci} \right) \right] \tag{21}$$

The Neumann boundary condition is added to these equations: $\partial Z / \partial \mathbf{n} = 0$, where $\mathbf{n}$ is the unit normal vector to the boundary of the region Ω_k. The descent equations for any region and for any camera are shown in (22). In these equations τ indicates the algorithm execution time and g' is the ordinary derivative of the boundary preserving function g.

$$\frac{\partial Z_{ci}}{\partial \tau} = \frac{2}{Z_{ci}^2} \left(\mathbf{S}_i \mathbf{v}_{wk} + \mathbf{R}_i \boldsymbol{\omega}_{wk} \right) \psi_{ki} + \mu div \left(\frac{g'\left(\left\| \nabla Z_{ci} \right\| \right)}{\left\| \nabla Z_{ci} \right\|} \nabla Z_{ci} \right) \quad \begin{matrix} i = 1, \ldots, n_c \\ k = 1, \ldots, N \end{matrix} \tag{22}$$

There are different boundary preserving functions ($g(a)$) in the literature. In this thesis some of these functions have been studied, concluding that all boundary preserving functions are equally valid, although they require a proper adjustment of the constant μ. For simplicity, in this thesis, the quadratic function ($g(a) = a^2$) has been used.

c) Curve evolution for 3D motion segmentation

In this step the depth Z_{ci} and 3D motion parameters $\{\mathbf{v}_{wk}, \boldsymbol{\omega}_{wk}\}_{k=1}^{N}$ in Γ_w are fixed. Then, the energy to be minimized with respect to the curves that define the mobile robot contours $\{\gamma_{ki}\}_{k=1,\ldots,N-1}^{i=1,\ldots,n_c}$ in each image, is shown in equation (23), where $\xi_{ki}(\mathbf{x}) = \psi_{ki}^2(\mathbf{x}) + \mu g\left(\left\| \nabla Z_{ci} \right\| \right)$.

$$E\left[\{\gamma_{ki}\}_{k=1,\ldots,N-1}^{i=1,\ldots,n_c} \right] = \sum_{k=1}^{N} \sum_{i=1}^{n_c} \int_{\Omega_{ki}} \xi_k(\mathbf{x}) d\mathbf{x} + \lambda \sum_{k=1}^{N-1} \sum_{i=1}^{n_c} \oint_{\gamma_{ki}} ds \tag{23}$$

For multiple region segmentation, the following derivative, for the images acquired for each of the n_c cameras, are obtained:

$$\frac{\partial E_{ki}}{\partial \gamma_{ki}} = \left(\xi_{ki} - \varphi_{ki} + \lambda \kappa_{\gamma_{ki}} \right) \mathbf{n}_{ki}, \quad \begin{matrix} i = 1, \ldots, n_c \\ k = 1, \ldots, N-1 \end{matrix} \tag{24}$$

And the corresponding Euler-Lagrange descent equations:

$$\frac{\partial \gamma_{ki}}{\partial \tau}(s,\tau) = -\left(\xi_{ki}(\gamma_{ki}(s)) - \varphi_{ki}(\gamma_{ki}(s)) + \lambda\kappa_{\gamma_{ki}}(\gamma_{ki}(s))\right) \times \mathbf{n}_{ki}(\gamma_{ki}(s)) \qquad i = 1,\ldots,n_c \quad k = 1,\ldots,N-1 \tag{25}$$

In (24) and (25) $\kappa_{\gamma_{ki}}$ is the mean curvature of contour defined by γ_{ki}, n_{ki} is the exterior unit normal function to the curve γ_{ki} and functions φ_{ki} are defined by equation (26).

$$\varphi_{ki}(\gamma_{ki}(s)) = \min_{j \neq k} \xi_{ji}(\gamma_{ki}(s)) \tag{26}$$

For the implementation of (25), each curve γ_{ki} is represented by the zero level set of a function Φ_{ki}, the interior of γ_{ki} corresponds to the set $\{\Phi_{ki} > 0\}$ and the exterior points of γ_{ki} correspond to $\{\Phi_{ki} < 0\}$. So, the following system of coupled partial differential equations are obtained:

$$\frac{\partial \Phi_{ki}}{\partial \tau}(\mathbf{x},\tau) = -\left(\xi_{ki}(\mathbf{x}) - \varphi_{ki}(\mathbf{x}) + \lambda\kappa_{\Phi_{ki}}(\mathbf{x})\right)\left\|\nabla\Phi_{ki}(\mathbf{x})\right\| \qquad i = 1,\ldots,n_c \quad k = 1,\ldots,N-1 \tag{27}$$

where the mean curvature $\kappa_{\Phi_{ki}}$ is given by: $div\left(\nabla\Phi_{ki}/\left\|\nabla\Phi_{ki}\right\|\right)$ and the function $\varphi_{ki}(x)$ is shown in (28).

$$\varphi_{ki}(\mathbf{x}) = \begin{cases} \min \xi_j(\mathbf{x}) & for\ \ j \neq k \in \{1,\ldots,N-1\},\ \Phi_{ji}(\mathbf{x}) > 0 \\ \xi_{Ni}(\mathbf{x}) & else \end{cases} \tag{28}$$

After initialization, the three stages described above are repeated until the algorithm convergence. The algorithm converges when the computed variables cease to evolve significantly.

3. Curve and depth initialization

The proposed method for motion segmentation and 3D positioning requires the initialization of several variables:

- A set of *N*-1 curves $\{\gamma_{ki}\}_{k=1,\ldots,N-1}^{i=1,\ldots,n_c}$ that divide each image in *N* regions. These curves define the boundaries of the segmentation in the images acquired by each camera.
- The depth (distance from each 3D point **P** to each camera). The value of depth in each point coincides with the Z_{ci} coordinate of the point **P** related to the coordinate system of the camera *i*, Γ_{ci}.

It is also necessary to estimate the number of mobile robots in the scene.

The initialization process is very important due to the high reliance of the results on the initial values of the variables. This process includes three different steps: in the first step, we obtain the initial curves. Since cameras are located in fixed position within the intelligent space, the *N*-1 initial curves are obtained using GPCA (Generalized Principal Components Analysis) [Jieping et al. 2004]. Then, the initial depth (relative to each camera coordinate system Γ_{ci}) is obtained using Visual Hull 3D [Laurentini 1991] which allows obtaining a 3D occupancy grid (composed by cubes with size Δh) in Γ_w from the initial segmentation boundaries, that have been computed previously using GPCA. Finally, an extended version of the k-means algorithm is used to estimate the number of mobile robots in the scene. The three steps are described below.

3.1. Curve Initialization

As previously mentioned, GPCA [Jieping et al. 2004] is used in this work to obtain a background model for each of the n_c cameras. Background modelling is carried out from a set of N_i background images $\{\mathbf{I}_j\}_{j=1}^{N_i}$ that do not contain any mobile robot. Using GPCA we obtain two transformation matrices, $\mathbf{L}_{ci}$ and $\mathbf{R}_{ci}$, for each camera. These matrices are calculated in each camera, and they represent the background model. Since the cameras are placed in fixed positions within the environment, the background modelling stage needs to be carried out only once, and it can be done off-line.

Both, the number of images (N_i) considered to create the background model and the number of principal components (d), determine the quality of the transformed image. In this sense, several experimental tests have been carried out. In these tests, different image sequences acquired in the ISPCE-UAH have been used. These experimental tests have allowed us to conclude that 10 images and 1 eigenvector are enough to build a suitable background model under the conditions raised in this thesis (interior environment with artificial lighting).

GPCA [Jieping et al. 2004] is also used to initialize the segmentation boundaries by comparing each image to the background model. In this stage, each image is projected (equation (29)) to the GPCA space using the matrices **L** and **R** (that have been obtained previously). After that, the image is reconstructed (equation (30)). In these two

equations **M** represents the mean of the N_i images that have been used to obtain the background model.

$$\mathbf{I}_T = \mathbf{L}^T(\mathbf{I} - \mathbf{M})\mathbf{R} \tag{29}$$

$$\mathbf{I}_R = \mathbf{L}\mathbf{I}_T\mathbf{R}^T + \mathbf{M} \tag{30}$$

Then, the reconstruction error is computed. This error is defined as the difference between the reconstructed ($\mathbf{I}_R$) and the original (**I**) image and can be calculated subtracting the images pixel-to-pixel, but this approach is not robust against noise. Therefore, we define a set of pixels (window) around each pixel (with dimensions *q*x*q*) called $\mathbf{\Phi}_{wi}$ in the original image an $\hat{\mathbf{\Phi}}_{wi}$ in the reconstructed image, and we obtain the reconstruction error for these windows, using equation (31).

In order to setting up the window size (*q*), several experimental tests with different images belonging to different image sequences acquired in the ISPACE-UAH have been carried out. As a result of the experimental tests, it has been decided to use a window size of 3x3 pixels for the calculation of the reconstruction error. This window size allows reducing the noise in the input images, without increasing excessively the time required for the calculation of the reconstruction error.

$$\varepsilon_{wi} = \left\| \mathbf{\Phi}_{wi} - \hat{\mathbf{\Phi}}_{wi} \right\| \tag{31}$$

Pixels whose reconstruction error (calculated using equation (31)) is higher than a threshold are candidates to belong to a mobile robot, because in those pixels there is an important difference between the current image and the background model. The value of the threshold is very important. In this line, an adaptive threshold has been defined as a function of the difference between the average intensity of the input image (*I*) and the average intensity of the background images (I_{media}^{fondo}) [Losada et al. 2009]. This adaptive threshold is defined in equation (32)

$$U = \left(\alpha + \beta \left| I - I_{media}^{fondo} \right|\right) \cdot \max\left(\left\{ \varepsilon_{r,\,fondo}(i,j) \right\}_{\substack{i=1,\ldots,nfilas \\ j=1,\ldots,ncols}} \right) \qquad K, \alpha, \beta \in \Re,\ K, \alpha, \beta \geq 0 \tag{32}$$

The adaptive threshold defined in (32) has been validated by means of a set of experimental tests. These experimental tests have been carried out using several image sequences, with different characteristics, acquired in the ISPACE-UAH. In order to

evaluate the effect of the lighting changes, the test image sequences include both, artificial and natural changes in the intensity of the images.

After setting the threshold, the proposed solution for curve initialization allows us to obtain the initial curves by using GPCA. A block diagram including all the stages is shown in Figure 4. All these stages have to be executed for each camera to obtain the set of initial curves $\{\gamma_{ki}\}_{k=1,..,N-1}^{i=1,..,n_c}$ for each one.

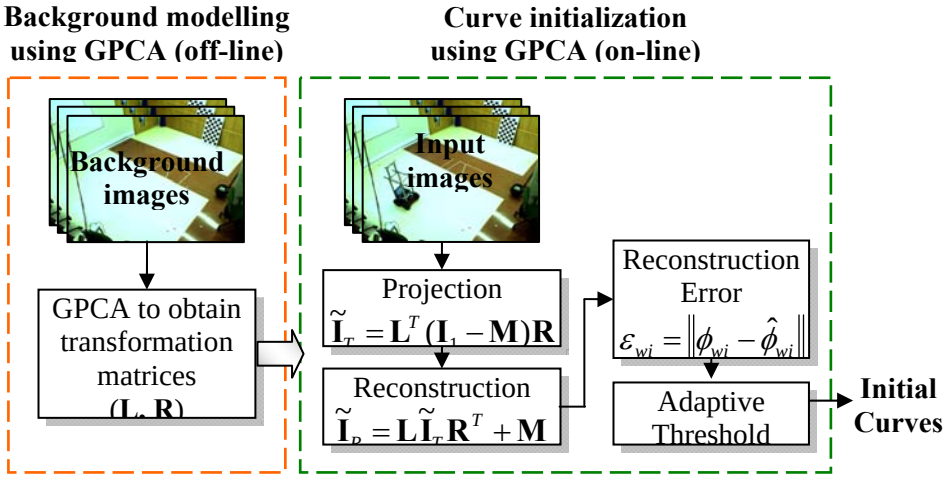


Figure 4. General block diagram of the proposed method for curve initialization using GPCA.

In order to reduce the effect of the shadows casted by different objects, we have added a final step to remove them. When the robots move in a plane, it is possible to remove shadows easily by removing the pixels whose height in Γ_w is zero. However, we have decided to consider the case of robots that are moving in a not flat surface. Under these conditions, shadow removal is done by projecting each input image to an illumination-invariant space [Finlayson et al. 2002], [Finlayson et al. 2006].

Before projecting the input images, it is necessary to determine the projection direction. This direction can be obtained through the color calibration of the cameras. [Finlayson et al. 2002]. In this thesis we have not calibrated the cameras, so, the invariant direction is set by entropy minimization [Finlayson et al. 2004].

It is necessary to highlight that the use of the proposed solution for the initialization of the curves improves the quality of the initial curves with respect to the use of circles (as proposed in the work [Sekkati & Mitiche 2006a] or [Sekkati & Mitiche 2006b]) because the initial curves are closer to the real contours of the robots. This involves a reduction in the number of iterations of the algorithm when minimizing the objective function, with the consequent reduction in the processing time.

3.2. Depth initialization

After curve initialization, Visual Hull 3D [Laurentini 1991] is used to obtain a 3D occupancy grid (composed of cubes of size Δh) in Γ_w from the initial segmentation boundaries computed previously. Figure 5 shows an occupancy grid obtained by using VH3D. This grid has been obtained from the curves shown in the images of the same figure.

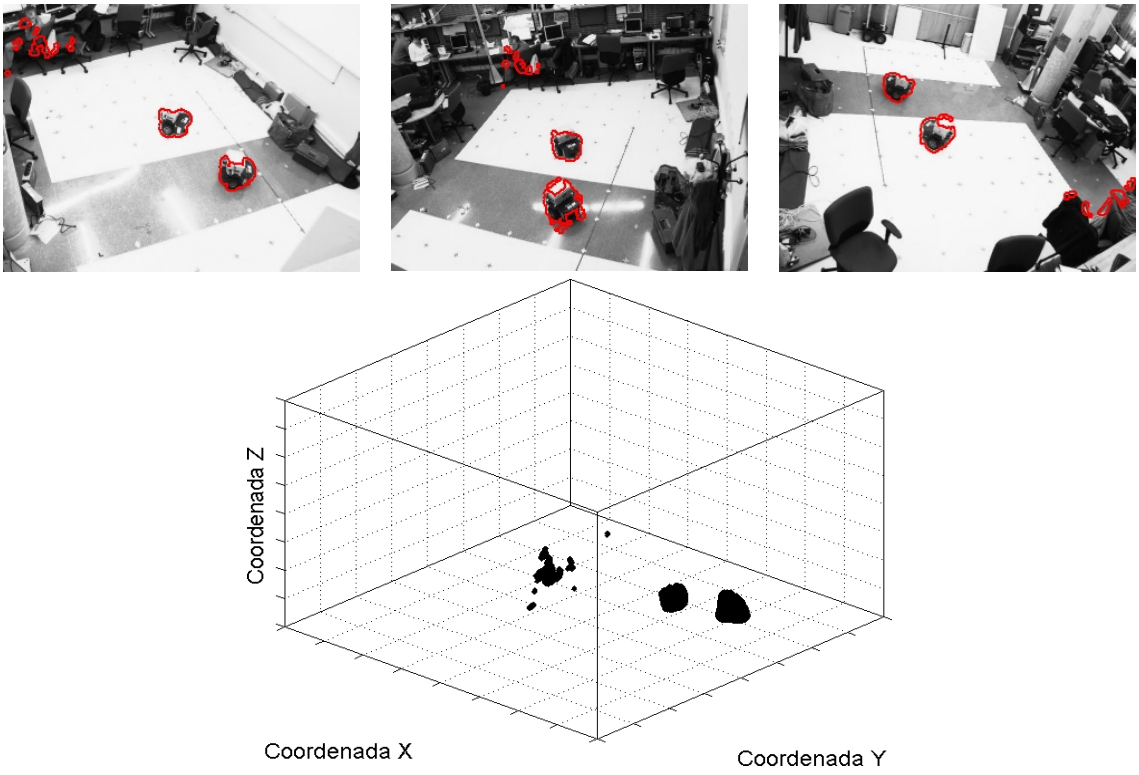


Figure 5. Input images belonging to a test image sequence, initial curves obtained using GPCA and 3D occupancy grid obtained from the initial curves by using VH3D.

The 3D coordinates of the occupied cell are projected from Γ_w to each camera coordinate system Γ_{ci} (i=1,…,n_c) through the transformation matrices ($\mathbf{R}_{wci}$ and $\mathbf{T}_{wci}$) to obtain a set of points on the mobile robots in Γ_{ci}. This process provides an effective method for depth initialization in each camera. Figure 6 presents a block diagram including the main steps in the depth initialization process.

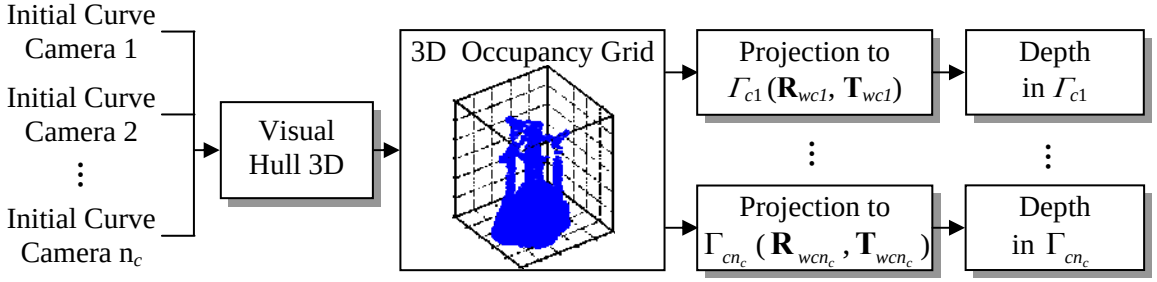
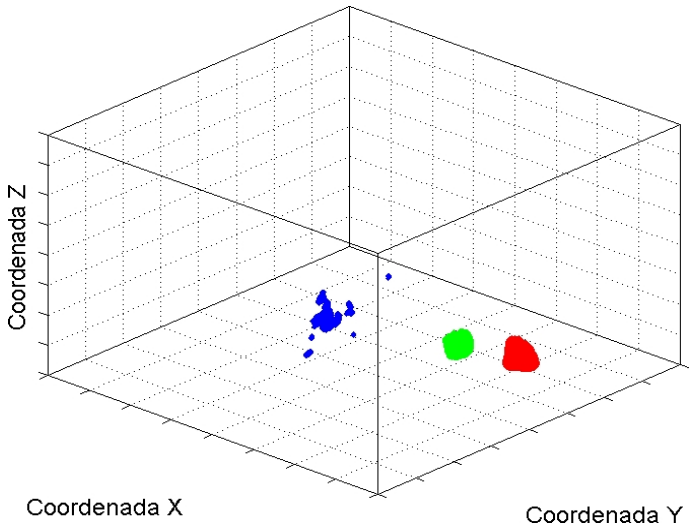


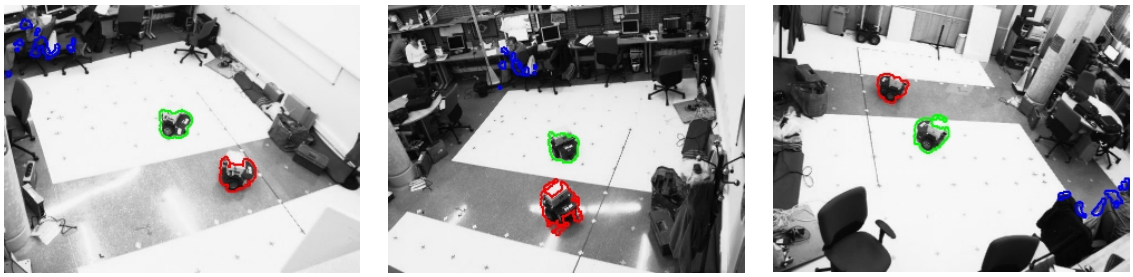
Figure 6. General block diagram of the proposed method for depth initialization using GPCA.

The algorithm used for motion segmentation and 3D positioning requires previous knowledge about the number of mobile robots. In order to estimate this value, we have included a clustering algorithm in the initialization process. In this stage, we project the coordinates of the occupied cell in the 3D occupancy grid obtained using Visual Hull 3D onto XY plane in Γ_w. Then, we cluster the 2D data using an extended version of k-means [Lloyd 1982], [Kanungo et al. 2002]. This clustering algorithm allows us to obtain a good estimation of the number of robots in the scene, and a division of the initial curves in each image.

After the classification algorithm, we obtain an estimation of the number of mobile robots. That information can be used in the stages of motion segmentation and 3D positioning. Moreover, 3D points corresponding to the occupied positions are classified according to the element to which they belong (Figure 7(a)). This classification allows dividing the initial curves obtained by using GPCA. As a result of that we obtain a set of curves, identified according to the robot (or mobile element) they correspond with, as shown in Figure 7(b).



(a)



(b)

Figure 7. (a) Results of the classification obtained by applying k-means clustering to the occupied positions in the occupancy grid shown in Figure 5.(a). (b) Initial curves obtained after estimating the number of robots, from the classified samples in Γ_w.

4. Identification and tracking of multiple mobile robots

As a result of the stages of curves and depth initialization and motion segmentation we obtain a set of samples belonging to each mobile robot in the scene. After the motion segmentation and 3D position estimation of all mobile objects in the scene, the identification of the robots is obtained by comparing the components of the linear velocity estimated by the motion segmentation algorithm and the components of the linear velocity measured by the odometric sensors onboard the robots. It is possible because the mobile robots are controlled by the intelligent space.

To that end, we have a set of n_r velocities $\left\{\mathbf{v}_m^i = \left(v_{mx}^i, v_{my}^i, v_{mz}^i\right)\right\}_{i=1}^{n_r}$ which have been measured by the odometric sensor onboard the robots. We also have a set of velocities $\left\{\mathbf{v}_e^j = \left(v_{ex}^j, v_{ey}^j, v_{ez}^j\right)\right\}_{j=1}^{N-1}$ for each of the N-1 clusters $\mathbf{G}_{j=1:N-1}$ in Γ_w, which have been estimated by the algorithm based on the minimization of an objective function. For each controlled agent, robot$_i$ (i=1,2,…,n_r), we obtain the set of points $\left\{\mathbf{G}_j\right\}_{j=1}^{N-1}$ that corresponds to each robot. The correspondence between $\left\{\mathbf{G}_j\right\}_{j=1}^{N-1}$ and robot$_i$ is given by (33).

$$G_j \in Robot_i \Leftrightarrow j = \underset{j=1,2,\ldots,N-1}{\arg\min}\left\{\sqrt{\left(v_{mx}^i - v_{ex}^j\right)^2 + \left(v_{my}^i - v_{ey}^j\right)^2}\right\} \quad i = 1,2,\ldots,n_r \tag{33}$$

Thus, it is possible to identify which set of points $\left\{\mathbf{G}_j\right\}_{j=1}^{N-1}$ corresponds to each mobile robot. This also allows the discrimination between the controlled agents and other mobile elements (users or obstacles) that may be present in the ISPACE-UAH.

In order to track the mobile robots an eXtended Particle Filter with Classification Process (XPFCP) [Marron 2008] is used. The choice of XPFCP against the basic particle filter (PF) or the extended particle filter (XPF) is due to its multimodal nature. The XPFCP allows tracking a variable number of robots with a single estimator, without increasing the state vector.

In order to characterize the evolution of the position and velocity of each mobile robot we want to follow, a generic discrete first-order model is used. This model is obtained from the 3D coordinates that define the position of the robots and their linear velocity components along X_w and Y_w axis in Γ_w. Thus, we define the state $\mathbf{x}_t$ and measurement $\mathbf{y}_t$ vectors in t through the 5x1 vectors shown in equations (34) and (35) respectively.

$$\mathbf{x}_t = \begin{bmatrix} X_{w,t} & Y_{w,t} & Z_{w,t} & v_{x,t}^w & v_{y,t}^w \end{bmatrix}^T \quad (34)$$

$$\mathbf{y}_t = \begin{bmatrix} X_{w,t} & Y_{w,t} & Z_{w,t} & v_{x,t}^w & v_{y,t}^w \end{bmatrix}^T \quad (35)$$

For each of the 3D measures obtained from the stage of motion segmentation and positioning of the mobile robots, it is obtained a measurement vector. This vector is composed by the 3D coordinates of the measurement and the linear velocity components along the X_w and Y_w axis in Γ_w of the set of measurements $\{\mathbf{G}_j\}_{j=1}^{N-1}$ the measurement belongs to. These vectors are added to the XPFCP in the re-initialization step.

The XPFCP provides a set of measurements that are clustered according to the robot they correspond to. This set of measurements provides a good estimation for the position of each robot in the next image that is incorporated into the initialization step. Specifically, the information provided by XPFCP is incorporated to the step where we obtain the occupancy grid by using VH3D. It is also included in the stage of estimation of the number of mobile robots by using k-means. This feedback allows reducing significantly the time spent in the initialization of variables.

5. Experimental results

In order to validate the proposed system, several experiments have been carried out in the ISPACE-UAH. In these experiments we have used four five-hundred image sequences. These sequences have been acquired using three of the four cameras in the ISPACE-UAH. Figure 8 shows one scene belonging to each sequence. As can be noticed in Figure 8, sequences 1 and 2 contain one robot. The main difference between the images belonging to sequences 1 and 2 is that the image sequence 2 includes real lighting changes. It is worth noting that the image sequence 1 has been used to set up the values of the different parameters involved in the proposed solution. Sequence 3 contains two mobile robots and sequence 4 includes a robot, a user, and some obstacles.

The proposed algorithm for motion segmentation and 3D localization using a multi-camera sensor system has been used to obtain motion segmentation and 3D position for each couple of images in each sequence. All the experiments shown in this work have been carried out on Intel® core 2, 6600 with 2.4 GHz using Matlab.

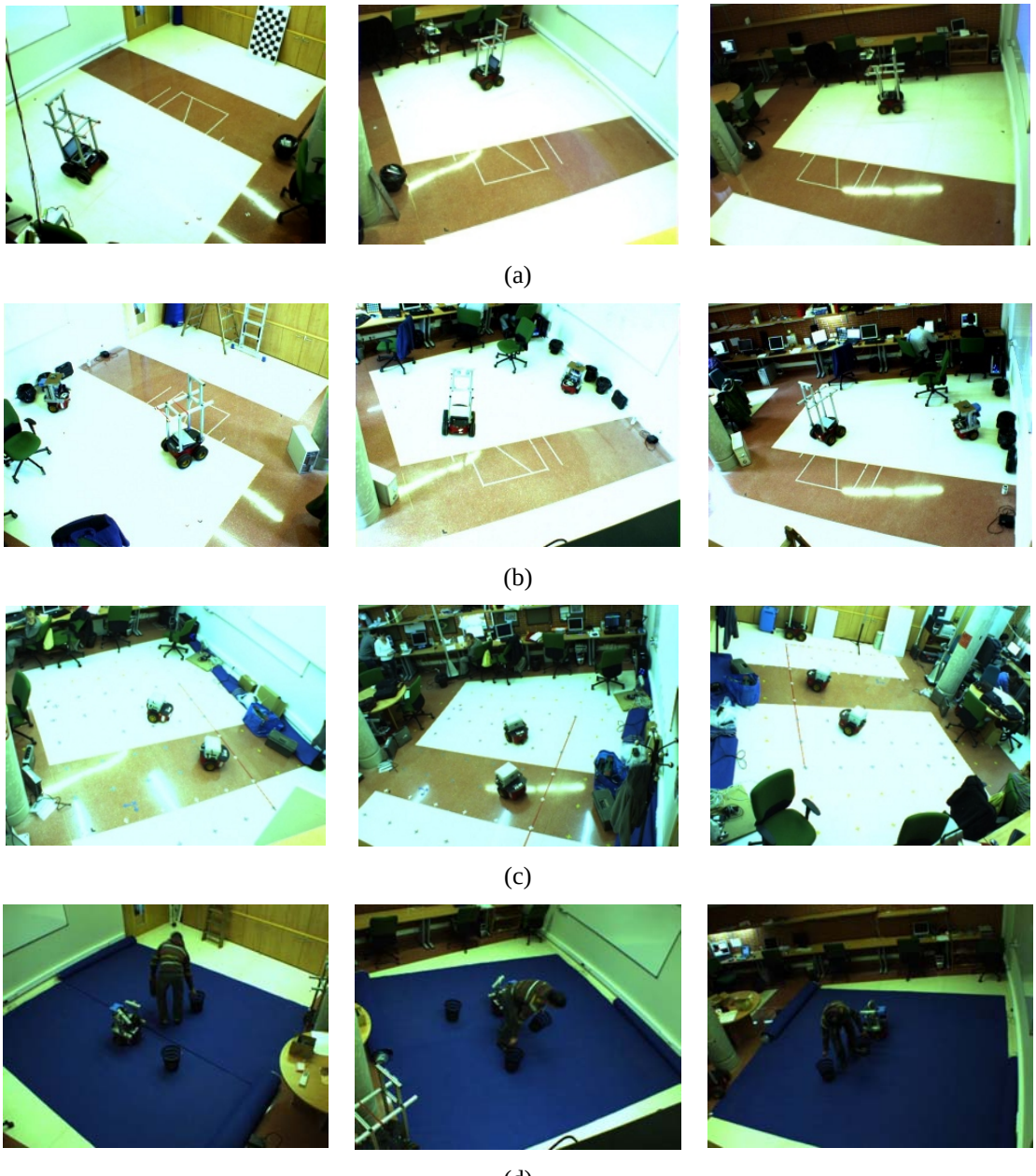
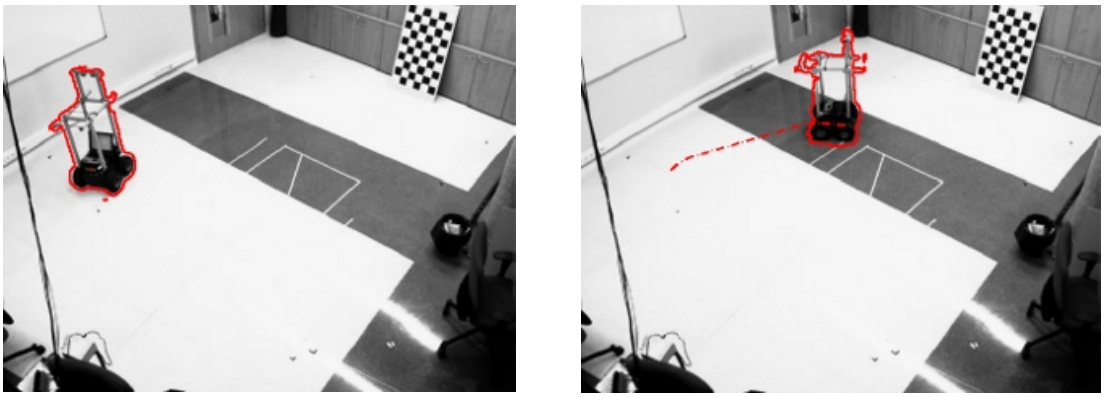


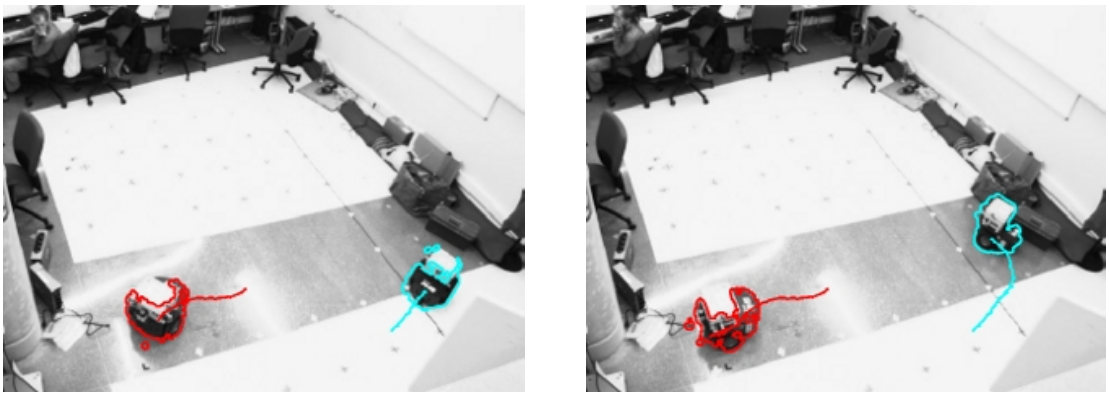
Figure 8. Images belonging to the test sequences, acquired by fixed cameras in the ISPACE-UAH. (a) Images belonging to the sequence 1 (b) Images belonging to the sequence 2 (c) Images belonging to the sequence 3. (d) Images belonging to the sequence 3.

To start with the results, the boundaries of the motion segmentation in two images belonging to the sequence 1 (Figure 8(a)), the sequence 2 (Figure 8 (b)), the sequence 3 (Figure 8(c)) and the sequence 4 (Figure 8(d)) respectively are shown in Figure 9. The estimated trajectory is also shown in these figures. In all the images, the segmentation boundary is close to the real contour of the mobile robot in the image plane.

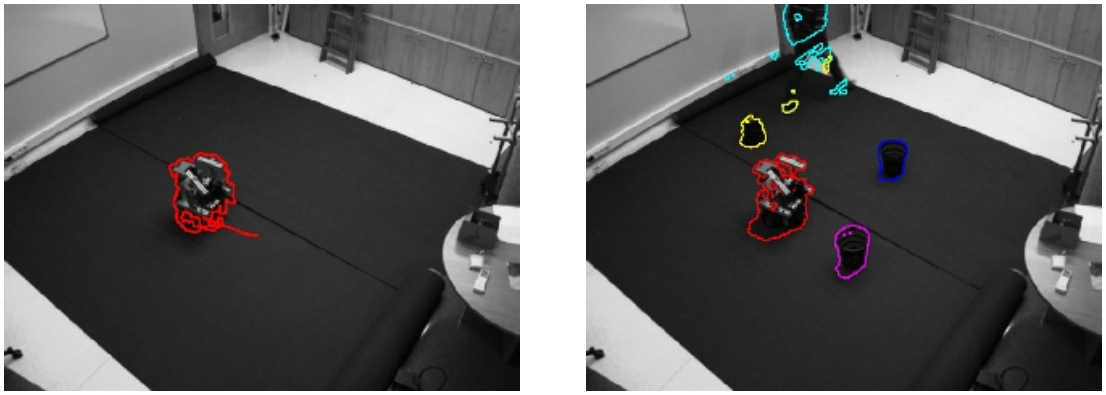


(a)

(b)



(c)



(d)

Figure 9. Boundaries of the segmentation obtained after the objective function minimization for two images belonging to each test sequences. (a) Sequence 1 (Figure 8(a)). (b) Sequence 2 (Figure 8(b)). (c) Sequence 3 (Figure 8(c)). (d) Sequence 4 (Figure 8(d)).

The proposed algorithm for multiple cameras also allows obtaining a 3D reconstruction of the mobile robots. Figure 10 shows the result of the 3D reconstruction obtained by using VH3D with the boundaries of the segmentation shown in Figure 9.

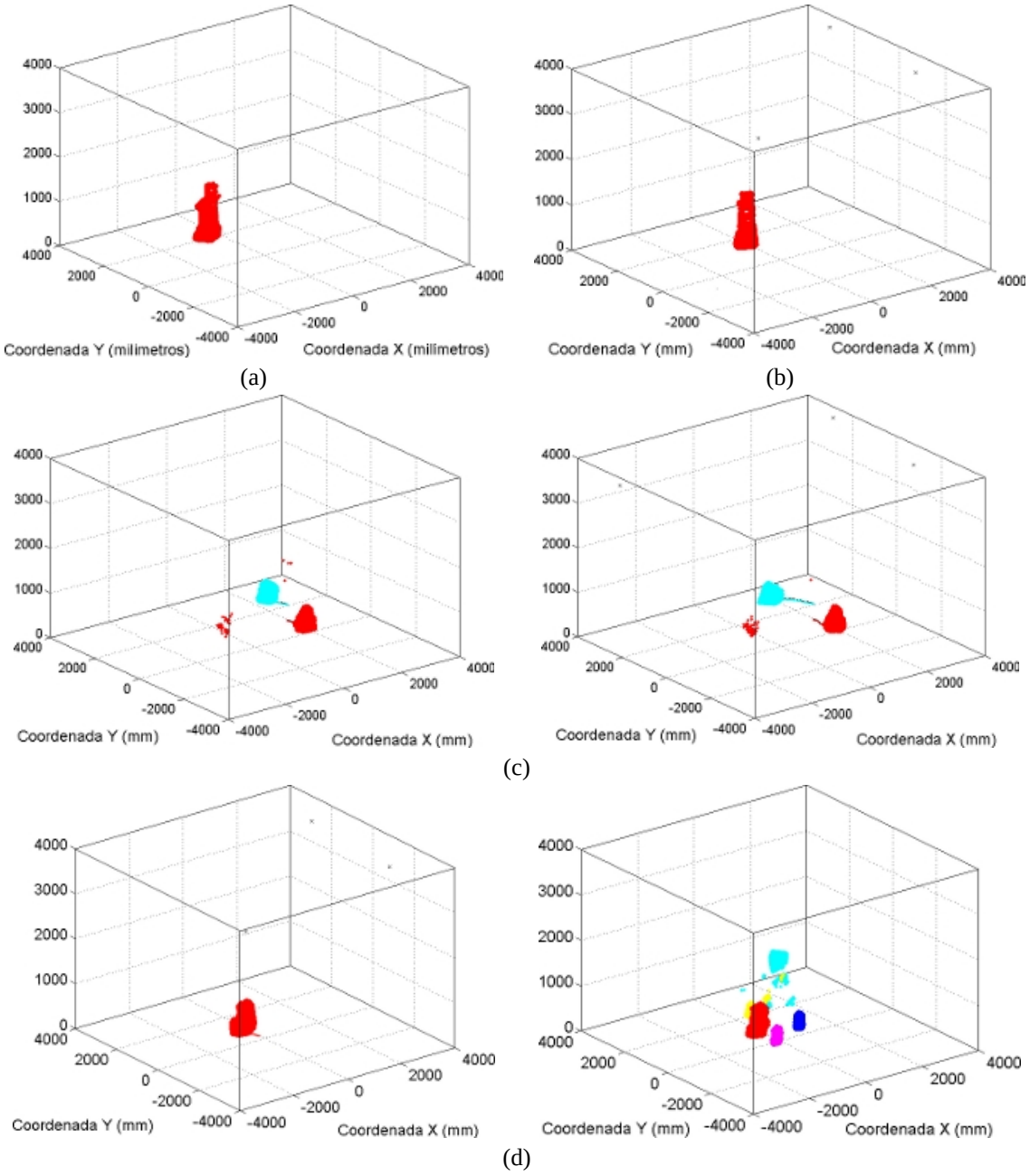


Figure 10. 3D reconstruction for two images belonging to each test sequences. (a) Sequence 1 (Figure 8(a)). (b) Sequence 2 (Figure 8(b)). (c) Sequence 3 (Figure 8(c)). (d) Sequence 4 (Figure 8(d)).

With regard to 3D positioning, Figure 11 shows the projection, onto the image plane, of the 3D trajectory of the mobile robot estimated by the algorithm and measured by the odometric sensors on board the robots. The represented trajectory has been calculated using 100 images belonging to sequences 1, 2 and 3. The trajectories shown

in Figure 11 are obtained by projecting the estimated trajectory in Γ_w, obtained using the proposed algorithm, onto the image plane of the camera 1.

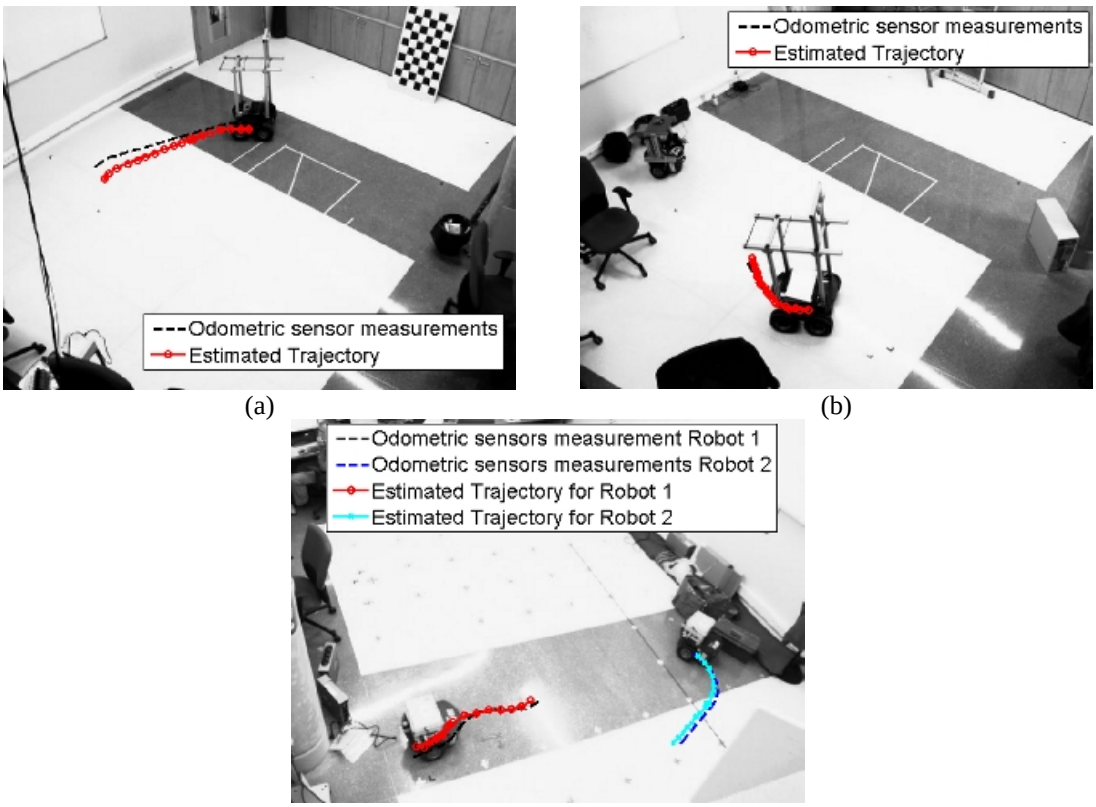


Figure 11. 3D trajectory estimated by the algorithm and measured by the odometric sensors on board the robots projected onto the image plane. (a) Images belonging to the sequence 1 (Figure 8(a)). (b) Images belonging to the sequence 2 (Figure 8(b)). (c) Images belonging to the sequence 3 (Figure 8(c))

These trajectories can also be represented in the world coordinate system. The coordinates of the centroid of the points belonging to each robot are projected onto the plane (X_w, Y_w) in Γ_w to obtain the 3D position. The result of this projection for a 500 images belonging to the sequence 3 is shown in Figure 12.

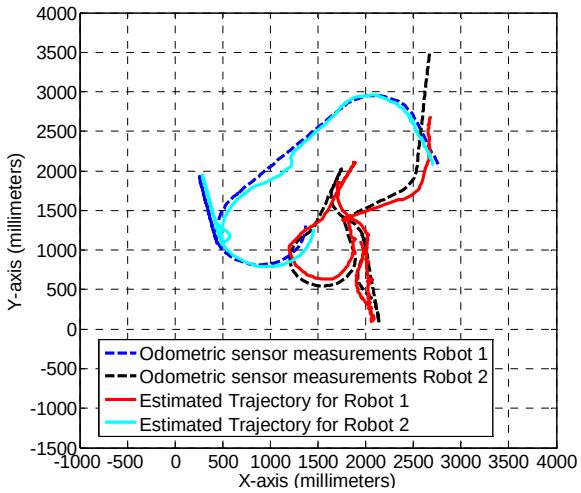


Figure 12. 3D trajectory estimated by the algorithm and measured by the odometric sensors on board the robots on the X_w, Y_w plane obtained from 500 images belonging to the sequence 3.

As can be observed in Figure 12, the estimated trajectories are closer to the measurements of the odometric sensors as the number of cameras increases. This fact can also be observed in the positioning error calculated as the difference between the estimated and the measured positions along X_w and Y_w axis, using equation (36):

$$\varepsilon_p = \sqrt{\varepsilon_{px}^2 + \varepsilon_{py}^2} \tag{36}$$

The positioning error, calculated for 100 images belonging to the sequences 1, 2 and 3, has been represented in Figure 13.

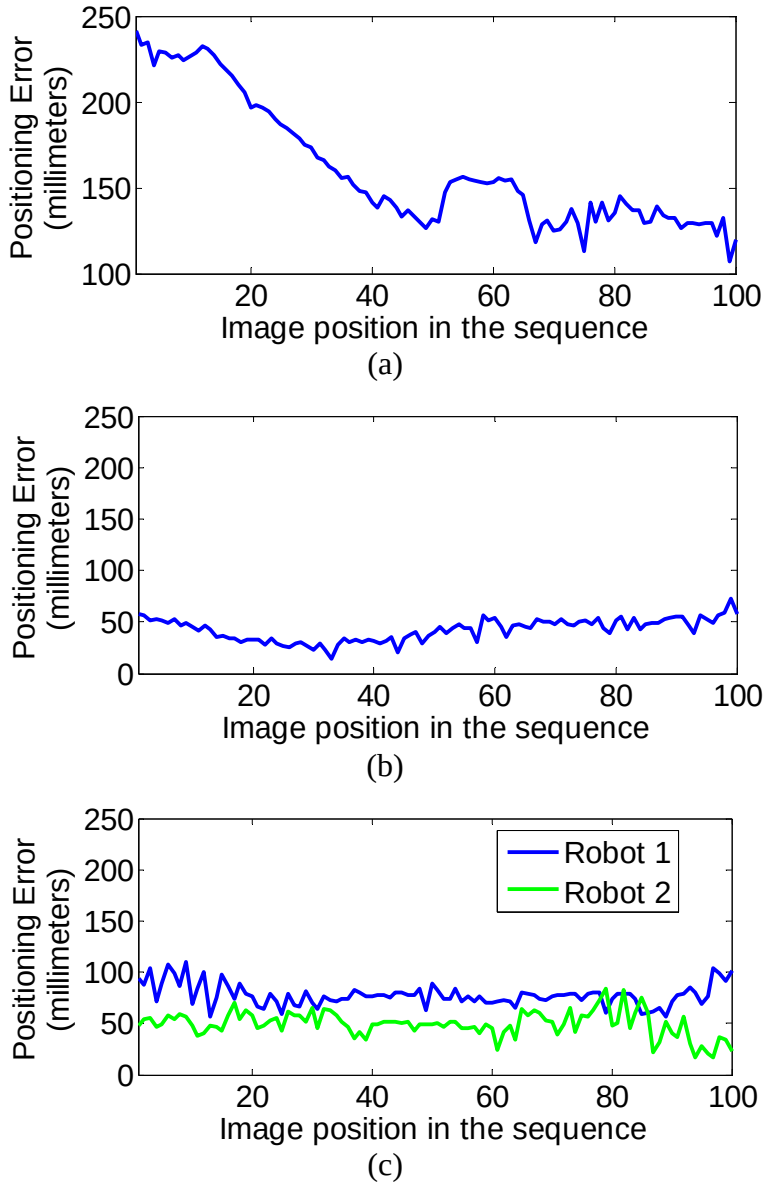


Figure 13. Positioning error (in millimetres) of the mobile robots, calculated using equation (36). (a) Robot in the image sequence 1 (b) Robot in the image sequence 2 (c) Robots in the image sequence 3.

Finally, the average value of the positioning error for 100 images represented in Figure 13 is shown in Figure.

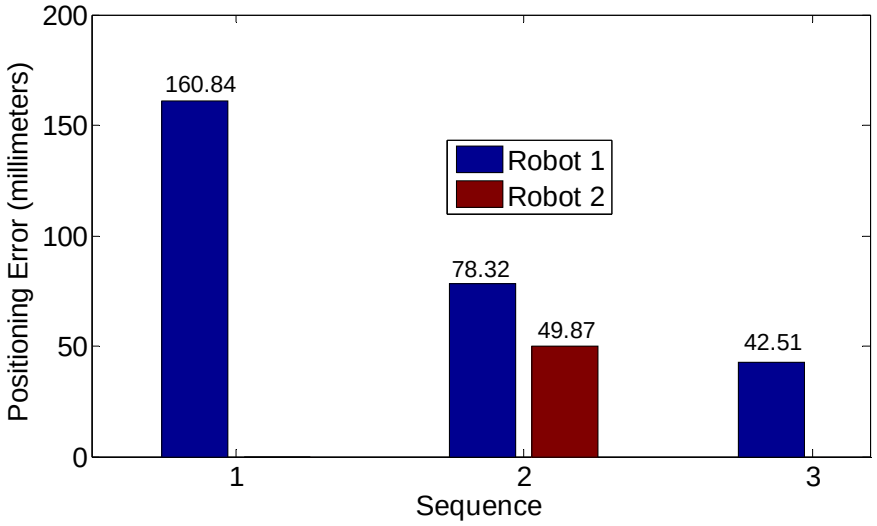


Figure 14. Average value of the positioning error for 100 images belonging to the sequences 1, 2 and 3.

6. Summary of the conclusions and contributions of the thesis

After describing each stage that form the proposed solution, the main conclusions and contributions of this thesis are summarized.

A method for obtaining the motion segmentation and 3D localization of multiple mobile robots in an intelligent space using a multi-camera sensor system has been proposed. The sensor system consists of a set of calibrated and synchronized cameras placed in fixed positions within the environment (in our case, the ISPACE-UAH). Motion segmentation and 3D position of the mobile robots are obtained through the minimization of an objective function that incorporates information from the multi-camera sensor.

In this thesis, the importance of the correct initialization of variables (curves that define the contours of the segmentation and depth) has been exposed. The initial values of these variables determine both the processing time, and the results of the segmentation and 3D positioning. The initial curves are obtained using GPCA. GPCA allows to model the background of the scene and to compare each input image with the generated background model. Regarding to depth, Visual Hull 3D allows us to relate the information from all the cameras. It provides a good alternative for estimating the initial value of the depth at each point of the image.

After the motion segmentation and 3D position estimation of all mobile objects in the scene, the identification of the robots is carried out by comparing the linear velocity

components along X_w and Y_w axis estimated by the motion segmentation algorithm and the linear velocity components along x and y axis measured by the odometric sensors onboard the robots. It is possible because the mobile robots are controlled by the intelligent space. Moreover, in this stage, it is also possible to discriminate between the controlled agents and other mobile elements (users or obstacles) that may be present in the ISPACE-UAH.

The proposed solution is completed with a tracking step based on a eXtended Particle Filter with Classification Process (XPFCP). The XPFCP allows tracking a variable number of mobile robots by using a single estimator, without increasing the state vector length. The information provided by XPFCP is incorporated into the initialization step. This feedback allows reducing significantly the time spent in the initialization of variables.

Several experimental tests have been carried out in the ISPACE-UAH and the obtained results validate the proposal. It has been demonstrated that the use of a multi-camera sensor increases significantly the accuracy of the 3D localization of the mobile robots against the use of a single camera.

The most outstanding contributions of the thesis are listed below.

- Proposal of an objective function for multiple cameras, and the corresponding minimization algorithm.
- Proposal of a solution that allows carrying out the fusion of the results when the objective function minimization is run independently for each camera.

Because of the importance of a correct initialization of the variables involved in the objective function, an important part of the efforts has been focused on finding alternatives for initialization. In this sense, several contributions have been made. These contributions allow obtaining motion segmentation and 3D positioning of the mobile robots in a smaller number of iterations of the minimization algorithm. They also allow increasing the robustness of the proposal solutions against lighting changes.

- It has been proposed a solution based on GPCA, that allows obtaining a set of initial curves that are close to the real contours of mobile robots.
- It has been set out an adaptive threshold for the reconstruction error. This adaptive threshold is based on the difference between the average intensity of

the input image and the average intensity of the images used to obtain the background model of the scene. It allows increasing the robustness of the curve initialization process against lighting changes.

- It has been proposed to use VH3D to relate the information from all available cameras in the world coordinates system Γ_w. It provides a good estimation of the initial depth.
- It has been proposed a solution, based on an extended version of k-means, to estimate the number of mobile robots present in the scene.
- Finally, it has been proposed the feedback of the information obtained from the XPFCP (estimation of the position of each mobile robot in the next image) to the stage of initialization of variables. It allows reducing the processing time of the initialization stage.

7. Future work

Upon finishing the thesis, it is interesting to emphasize some of the future lines that can be raised from this work.

- The most immediate task is the implementation of the different proposed solutions in real time.
- The solution proposed for motion segmentation and 3D positioning is restricted to rigid objects, another interesting line of future work is the creation of a new algorithm that also considers deformable objects.
- The proposal in this thesis has been evaluated in a small space (ISPACE-UAH) and it has not been considered the possibility of covering a larger space with multiple cameras, so that robots can be located even if they leave the ISPACE-UAH. This is a line of future work that has a special interest towards the implementation of such systems in buildings with multiple rooms.
- Finally, another interesting line of future work is to replace the present cameras by Time-Of-Flight (TOF) cameras. These cameras allow acquiring both intensity and range data into a single device and simplify the determination of geometrical properties of the scene. In this line, the use of Time-Of-Flight

(TOF) cameras can also be considered in order to improve the accuracy of the 3D positioning.

8. Publications arising from the thesis

- Cristina Losada, Manuel Mazo, Sira Palazuelos, Daniel Pizarro, Marta Marrón. *Multi-camera sensor system for 3D segmentation and localization of multiple mobile robots*. Sensors (Accepted for its publication).
- C. Losada, M. Mazo, S. Palazuelos, J. L. Martín, J. J. García. *Motion segmentation using GPCA techniques and optical flow*. Euro American Conference on Telematics and Information Systems 2007 (EATIS 2007). EATIS'07 ACM-DL Proceedings (ISBN: 978-1-59593-598-4). 2007.
- C. Losada, M. Mazo, S. Palazuelos, D. Pizarro, M. Marrón, F. Redondo. *3D Motion Segmentation and 3D Localization of Mobile Robots Using an Array of Static Cameras and Objective Function Minimization*. Proc. of the 2009 IEEE International Symposium on Intelligent Signal Processing. WISP 2009. (ISBN: 978-1-4244-5059-6). 2009.
- C. Losada, M. Mazo, S. Palazuelos, F. Redondo. *Adaptive Threshold for Robust Segmentation of Mobile Robots from Visual Information of their own Movement*. Proc. of the 2009 IEEE International Symposium on Intelligent Signal Processing. WISP 2009. (ISBN: 978-1-4244-5059-6). 2009.
- Cristina Losada, Manuel Mazo, Sira Palazuelos, Daniel Pizarro, Marta Marron. *Motion Segmentation and 3D Positioning of Multiple Mobile Robots Using an Array of Static Cameras in an Intelligent Space*. IEEE International Symposium on Industrial Electronics 2010. ISIE 2010. (Accepted for its publication)
- Cristina Losada, Juan Jesús García, Manuel Mazo, Jesús Ureña, Álvaro Hernández, Mª Jesús Díaz, Carlos De Marziani. *Uso de la técnica PCA para la validación de la detección de objetos en entornos ferroviarios*. Actas del XIII Seminario Anual de Automática, Electrónica Industrial e Instrumentación (SAAEI'06). ISBN: 84-8317-564-9 (Edición en CD). 2006.

- Cristina Losada, Manuel Mazo, Sira Palazuelos. *Segmentación de objetos en movimiento utilizando GPCA, técnicas algebraicas y flujo óptico para aplicaciones en espacios inteligentes*. Actas del XIV Seminario Anual de Automática, Electrónica Industrial e Instrumentación (SAAEI'07). 2007.
- Cristina Losada, Manuel Mazo, Sira Palazuelos, Francisco Redondo. *Segmentación robots móviles en espacios inteligentes utilizando técnicas GPCA y minimización de funciones de energía*. Actas del XV Seminario Anual de Automática, Electrónica Industrial e Instrumentación (SAAEI'08). ISBN: 13:978-84-96997-04-2. 2008.
- Losada, Cristina; Mazo, Manuel; Palazuelos, Sira; Blanco, Edward. *Segmentación y posicionamiento de sillas de ruedas en espacios inteligentes mediante minimización de funciones de energía*. Memorias del V Congreso de tecnologías de apoyo a la discapacidad (IBERDISCAP 2008). ISBN: 978-958-8316-63-5. pp: 271-274. 2008.

References

[Fernandez et al. 2007]

Fernandez, I.; Mazo, M; Lázaro, J.L.; Pizarro, D.; Santiso, E; Martín, P.; Losada, C. *Guidance of a mobile robot using an array of static cameras located in the environment*. Autonomous Robots. Vol. 23, Issue 4, pp. 305-324. 2007.

[Finlayson et al. 2002]

Finlayson, G.D.; Hordley, S.D. and Drew, M.S. *Removing shadows from images*. Proc. of the 7th European Conference on Computer Vision 2002. Part IV. pp: 823-836. 2002.

[Finlayson et al. 2004]

Finlayson, G.D.; Drew, M.S. and Lu, C. *Intrinsic Images by Entropy Minimization*. Proc. of the 8th European Conf. on Computer Vision 2004. pp: 582-595. 2004.

[Finlayson et al. 2006]

Finlayson, G.D.; Hordley, S.D.; Lu, C. and Drew, M.S. *On the removal of shadows from images*. IEEE Pattern Analysis and Machine Intelligence (PAMI) Jan 2006. Vol.28, Nº1, Pp: 59-68. 2006.

[Jieping et al. 2004]

Jieping, Y.; Ravi, H.; Qi, L. *GPCA: an efficient dimension reduction scheme for image compression and retrieval*. Proc. of the 10th ACM SIGKDD international conference on Knowledge discovery and data mining 2004. pp: 354-363. 2004.

[Kanungo et al. 2002]

Kanungo, T.; Netanyahu, N.S.; Wu, A.Y. *An efficient k-means clustering algorithm: analysis and implementation*. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 24, nº 7, July 2002.

[Laurentini 1991]

Laurentini, A. *The Visual Hull: a new tool for contour-based image understanding*. Proc. of the 7th Scandinavian Conference on Image Processing. pp: 993-1002. 1991.

[Lee et al. 2001]

Lee, J., Ando, N., Yakushi, T., & Nakajima, K. *Adaptive guidance for Mobile robots in intelligent infrastructure*. Proc. IEEE/RSJ International conference on robots and Systems 2001, pp. 90-95.

[Lloyd 1982]

Lloyd, S. *Least squares quantization in PCM*. IEEE Transactions on Information Theory. Vol. 28, nº 2, pp. 129-137. 1982

[Marron 2008]

Marron, M. *Seguimiento de múltiples objetos en entornos interiores muy poblados basado en la combinación de métodos probabilísticos y determinísticos*. PhD. Thesis. Universidad de Alcalá, 2008

[Mumford & Shah 1989]

Mumford, D. & Shah, J. *Optimal Approximations by Piecewise Smooth Functions and Associated Variational Problems*. Communications on Pure and Applied Math. Vol. 42, no. 5, pp: 577-684. 1989.

[Pizarro et al. 2009]

Pizarro, D. and Mazo, M. and Santiso, E. and Marron, M. and Fernandez, I. *Localization and Geometric Reconstruction of Mobile Robots Using a Camera Ring*. IEEE Transactions on Instrumentation and Measurement. Vol. 58, n. 8, 2009.

[Sekkati & Mitiche 2006a]

Sekkati, H. and Mitiche, A. *Joint Optical Flow Estimation, Segmentation, and Interpretation with Level Sets*. Computer Vision and Image Understanding 2006. Vol. 103, Nº 2, Pp: 89-100. 2006.

[Sekkati & Mitiche 2006b]

Sekkati, H. and Mitiche, A. *Concurrent 3D Motion Segmentation and 3D Interpretation of Temporal Sequences of Monocular Images*. IEEE Transactions on Image Processing 2006. Vol. 15, Nº 3, Pp: 641-653. 2006.

[Sogo et al. 1999]

Sogo, T., Ishiguro, H; Ishida, T. *Acquisition of qualitative spatial representation by visual observation*. Proc. IJCAI, pp.1054-1060. 1999.

[Steinhaus et al. 2004]

Steinhaus, P., Walther, M., Giesler, B., & Dillmann, R. *3D global and Mobile sensor data fusion for Mobile platform navigation*. Proc. of the IEEE International conference on robotics and automation (ICRA 2004). Vol. 4, pp. 3325-3330.

[Weiser 1999]

Weiser, M. *The computer for the twenty-first century*. Scientific American, Vol. 265. Issue. 3 pp. 94–104. 1991