Single Pan and Tilt Camera Indoor Positioning and Tracking System

T. Gaspar and P. Oliveira
IST/ISR, Lisboa, Portugal

Abstract— **An inexpensive single pan and tilt camera based indoor positioning and tracking system is proposed, supported on an architecture where three main modules can be identified: one related to the interface with the camera, tackled with parameter estimation techniques; other, responsible for isolating and identifying the target, based on advanced image processing techniques, and a third, that resorting to nonlinear dynamic system suboptimal state estimation techniques, performs the tracking of the target and estimates its position, and linear and angular velocities. The contributions of this work are fourfold: i) a new indoor positioning and tracking system architecture; ii) a new lens distortion calibration method, that preserves generic straight lines in images; iii) suboptimal nonlinear multiple-model adaptive estimation techniques, for the adopted target model, to tackle the positioning and tracking tasks, and iv) the implementation and validation in real time of a complex tracking system, based on a low cost single camera. To assess the performance of the proposed system, a series of indoor experimental tests for a range of operation of up to ten meter were carried out. An accuracy of 20 cm was obtained under realistic conditions.**

I. Introduction

With the development and the widespread use of robotic systems, localization and tracking have become fundamental issues that must be addressed in order to provide autonomous capabilities to a robot. The availability of reliable estimates is essential to its navigation and control systems, which justifies the significant effort that has been put into this domain, see [1], [2] and [3] and the references therein.

Successfully exploited alternative techniques have been reported such as infrared radiation, ultrasound, radio frequency, and vision (see a comparison in [1] and [5]). The indoor tracking system proposed in this work uses vision technology, since this technique has a growing domain of applicability and allows to achieve interesting results with very low investment. This system estimates in real time the position, velocity, and acceleration of a target that evolves in an unknown trajectory, in the 3D world, as well as its angular velocity. In order to accomplish this purpose, a new positioning and tracking architecture is detailed, based on suboptimal stochastic multiple-model adaptive estimation techniques. The complete process of synthesis, analysis, implementation, and validation in real time is presented next.

This document is organized as follows: in section II, the architecture and the main methodologies and algorithms used in the positioning and tracking system proposed are described. In section III, the camera and lens models are studied in detail. To isolate and identify the target, advanced image processing algorithms are discussed in section IV, and in section V a multiple-model nonlinear estimation technique is proposed. Section VI analyzes the experimental results obtained and in VII some concluding remarks are addressed.

This work was partially supported by Fundação para a Ciência e a Tecnologia (ISR/IST plurianual funding) through the POS_Conhecimento Program that includes FEDER funds and by the project PDCT/MAR - RUMOS of FCT and GREX of EC.

II. System Architecture

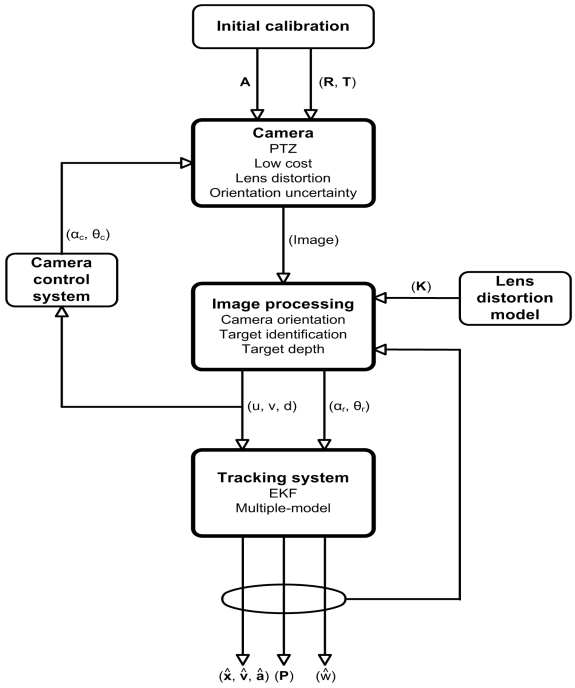

Fig. 1. Tracking system architecture.

In this project a new indoor positioning system architecture is proposed, based on three main modules: one that addresses the interface with the camera, the second that implements the image processing algorithms, and a third responsible for dynamic systems state estimation. The proposed architecture is presented in Fig. 1, where some quantities are introduced informally to augment the legibility of the document.

The extraction of physical information from an image acquired by a camera requires the knowledge of its intrinsic ($\mathbf{A}$) and extrinsic ($\mathbf{R}$ and $\mathbf{T}$) parameters, which are computed during the initial calibration process. In this work, calibration was preceded by an independent determination of a set of

ISBN 978-963-311-369-1
© Copyright EUCA 2009

parameters ($\mathbf{K}$) responsible for compensating the distortion introduced by the lens of the camera. Since the low cost camera used has no orientation sensor, the knowledge of its position in each moment required the development of an external algorithm capable of the estimation of its instantaneous pan and tilt angles (α_r and θ_r, respectively).

The target identification is the main purpose of the image processing module. An active contour method, usually denominated as snakes, was selected to track the important features in the image. The approach selected consists in estimating the target contour, providing the necessary information to compute its center coordinates (u, v) and its distance (d) to the origin of the world reference frame. These quantities correspond to the measurements that are used to estimate the position ($\widehat{\mathbf{x}}$), velocity ($\widehat{\mathbf{v}}$), and acceleration ($\widehat{\mathbf{a}}$) of the body to be tracked. Note that the computation of d requires the knowledge of the real dimensions of the target, since the proposed system uses one single camera instead of a stereo configuration.

To obtain estimates on the state and parameters of the underlying dynamic system, an estimation problem is formulated and solved. However, the dynamic model adopted and the sensor used have nonlinear characteristics. Extended Kalman filters included in a multiple-model adaptive estimation architecture were selected to provide estimates on the system state ($\widehat{\mathbf{x}}$, $\widehat{\mathbf{v}}$, and $\widehat{\mathbf{a}}$), to identify the unknown target angular velocity $\widehat{w}$, and the estimation error covariance P, as depicted in Fig. 1.

The command of the camera is the result of solving a decision problem, with the purpose of maintaining the target close to the image center. Since the range of movements available is very restricted, the implemented decision system is very simple and consists in computing the pan and tilt angles (α_c and θ_c), that should be sent to the camera at each moment. Large distances between the referred centers are avoided, thus the capability of the overall system to track the targets is increased.

III. Sensor: PTZ camera

A. Camera model

Given the high complexity of the camera optical system, and the consequent high number of parameters required to model the whole image acquisition process, it is common to exploit a linear model to the camera. In this architecture it was considered the classical pinhole model [4].

In order to determine the camera intrinsic and extrinsic parameters, the classical approach proposed by Faugeras [4] was selected and implemented, with the major advantages that only one image is required and reliable results can be obtained. A separate algorithm that compensates for lens distortion was implemented, as described in section III-C.

B. PTZ camera internal geometry

The camera used in this project has the ability to describe pan and tilt movements, which makes possible the variation over time of its extrinsic parameters. Thus, the rigorous definition of the rigid body transformation between camera and world reference frames implies the adoption of a model to the camera internal geometry and the study of its direct kinematics. The *Creative WebCam Live! Motion* camera used has a closed architecture, thus its internal geometry model was estimated from the analysis of its external structure, based on a small number of experiments.

The proposed model considers five transformations, that include the pan, tilt, and roll angles between the world and camera reference frames; the offset between the origin of the world reference frame and the camera rotation center, and the offset between the camera rotation and optical centers.

The composition of these transformations leads to the global transformation between world and camera reference frames ${}^{c}\mathbf{g}_M = {}^{M}\mathbf{g}_c^{-1}$, ${}^{M}\mathbf{g}_c = {}^{M}\mathbf{g}_0{}^{0}\mathbf{g}_1{}^{1}\mathbf{g}_2{}^{2}\mathbf{g}_3{}^{3}\mathbf{g}_c$, that is fundamental to determine the camera projection matrix over time.

The introduced geometry requires the knowledge of five parameters: pan, tilt and roll angles, the position of the camera optical center in the world coordinate frame, when these angles are zero, and the offset between this point and the camera rotation center. Since there is no position sensor in the camera, its orientation must be determined in real time using reference points in the 3D world. The position of the camera optical and rotation centers, when the pan and tilt angles are zero, can be computed on an initial stage resorting to points of the world with known coordinates.

C. Lens distortion

The mapping function of the pinhole camera between the 3D world and the 2D camera image is linear, when expressed in homogeneous coordinates. However, if a low-cost or wide-angle lens system is used, the linear pinhole camera model fails. In those cases, and for the camera used in this work, the radial lens distortion is the main source of errors (no vestige of tangential distortion was identified). Therefore, it is necessary to compensate this distortion resorting to a nonlinear inverse radial distortion function, which corrects measurements in the 2D camera image to those that would have been obtained with an ideal linear pinhole camera model.

The inverse radial distortion function is a mapping that recovers the coordinates (x, y) of undistorted points from the coordinates (x_d, y_d) of the correspondent distorted points, where both coordinates are related to a reference frame with origin in image distortion center (x_0, y_0). Since radial deformation increases with the distance to the distortion center, the inverse radial distortion function $f(r_d)$ can be approximated and parameterized by a Taylor expansion [6], that results in $x = x_d + x_d \sum_{i=0}^{\infty} k_i r_d^{i-1}$ and $y = y_d + y_d \sum_{i=0}^{\infty} k_i r_d^{i-1}$, where $r_d = \sqrt{x_d^2 + y_d^2}$.

In this project, it were only taken into account the parameters k_3 and k_5, that, as stated in [6] and according to practical tests, are sufficient to obtain good results. Using more parameters brings no major improvement to the approximation of $f(r_d)$ for images in video resolution, and an estimation of less parameters is more robust.

The proposed lens distortion compensation method is independent of the calibration process responsible for determining the pinhole model parameters, and is based on the *rationale* that straight lines in the 3D space must remain straight lines in 2D camera images. Ideally, if acquired images were not affected by distortion, 3D world straight lines would be preserved in 2D images. Hence, the inverse radial distortion model parameters estimation was based on the resolution of the following set of equations

$$\begin{cases} f_{i1} = (y_{i1} - \widehat{y}_{i1}(m_i, b_i, x_{i1}))^2 = 0 \\ \vdots \\ f_{iN_p} = (y_{iN_p} - \widehat{y}_{iN_p}(m_i, b_i, x_{iN_p}))^2 = 0 \end{cases} \quad i = 1, \ldots, N_r$$

with $\widehat{y}_{ij}(m_i, b_i, x_{ij}) = m_i x_{ij} + b_i$, where N_r and N_p are the number of straight lines and points per straight line acquired from the distorted image, respectively. A set of $N_r * N_p$ nonlinear equations results; its solution can be found resorting to the Newton's method, and estimates for parameters k_3, k_5, x_0, y_0, m_i, b_i, $i = 1, \ldots, N_r$, are obtained.

IV. Image Processing

In this section, advanced image processing algorithms are described to implement the target isolation and identification, leading to the measurements to be provided to the estimation system.

A. Target isolation and identification

The isolation and identification of the target to be tracked in each acquired image is proposed to be tackled resorting to an active contours method. Active contours [7], or snakes, are curves defined within an image domain that can move under the influence of internal forces coming from within the curve itself and external forces computed from the image data. The internal and external forces are defined so that the snake will conform to an object boundary or other desired features within an image. Snakes are widely used in several computer vision domains, such as edge detection [7], image segmentation [8], shape modeling [9], [10], or motion tracking [8], as happens in this application.

1) Parametric active contours (traditional method): in this project a *parametric active contour* method is used [7], in which a parameterized curve $\mathbf{x}(s) = [x(s), y(s)]$, $s \in [0, 1]$, evolves over time towards the desired image features, usually edges, attracted by external forces given by the negative gradient of a potential function. The evolution occurs in order to minimize the energy of the snake $E_{sk} = E_{int} + E_{ext}$, that includes a term related to its internal energy E_{int}, which has to do with its smoothness, and a term of external energy E_{ext}, based on forces extracted from the image [7]. Traditionally, this energy can be expressed in the form

$$E_{sk} = \int_0^1 \frac{1}{2}[\alpha|\mathbf{x}'(s)|^2 + \beta|\mathbf{x}''(s)|^2] + E_{ext}(\mathbf{x}(s))ds, \quad (1)$$

where parameters α and β control the snake tension and rigidity, respectively, and $\mathbf{x}'(s)$ and $\mathbf{x}''(s)$ denote the first and second derivatives of $\mathbf{x}(s)$ with respect to s.

Approximating the solution of the variational formulation (1) by the spacial finite differences method, with step h, yields

$$(\mathbf{x}_t)_i = \frac{\alpha}{h^2}(\mathbf{x}_{i+1} - 2\mathbf{x}_i + \mathbf{x}_{i-1}) - \frac{\beta}{h^4}(\mathbf{x}_{i+2} - 4\mathbf{x}_{i+1} + 6\mathbf{x}_i - 4\mathbf{x}_{i-1} + \mathbf{x}_{i-2}) + \mathbf{F}_{ext}^{(p)}(\mathbf{x}_i), \quad (2)$$

where $\mathbf{x}_i = \mathbf{x}(ih, t)$, and $\mathbf{F}_{ext}^{(p)}(\mathbf{x}_i)$ represents the image influence at the point $\mathbf{x}_i$.

The temporal evolution of the active contour in the image domain occurs according to $\mathbf{x}^{n+1} = \mathbf{x}^n + \tau\mathbf{x}_t^n$, where τ is the considered temporal step. The iterative process ends when the coordinates of each point of the snake remain approximately constant over time.

Traditional snakes have two main limitations: poor initializations may lead the snake towards boundaries other than the desired ones, and they have some difficulties in converging to boundary concavities. Therefore, in this project the *gradient vector flow* (GVF) approach proposed by Chenyang Xu and Jerry L. Prince in [11] was followed, where a new class of external forces for active contour models that addresses both problems referred above is introduced.

2) Gradient vector flow snakes: the overall GVF approach proposed in [11] consists in using a new external force, here denoted by $\mathbf{v}(x, y)$, which defines the *gradient vector flow* field. This force leads to the dynamic snake equation $\mathbf{x}_t(s, t) = \alpha\mathbf{x}''(s, t) - \beta\mathbf{x}''''(s, t) + \mathbf{v}$, whose solution corresponds to the GVF *snake*, and can be computed numerically by iterative processes, after discretization, in a procedure similar to the one followed in the traditional snake method.

GVF field is defined as the vector field $\mathbf{v}(x, y) = [u(x, y), v(x, y)]$ that minimizes the functional

$$\varepsilon = \int\int \mu(u_x^2 + u_y^2 + v_x^2 + v_y^2) + |\nabla f|^2|\mathbf{v} - \nabla f|^2 dxdy, \quad (3)$$

where the indices x and y represent the partial derivatives with respect to x and y, respectively; f is a scalar field $f(x, y) = -E_{ext}(x, y)$, and μ is a regularization parameter that should be set according to the amount of noise present in the image (images with more noise require the choice of larger values for the parameter μ).

Using the calculus of variations [12], it can be shown that the GVF field that minimizes (3) can be found by solving a pair of Euler equations, whose solution can be computed by means of an iterative numerical procedure, that, as deduced in [11], corresponds to propagate the GVF field components according to the iterative expressions

$$u_{i,j}^{n+1} = (1 - b_{i,j}\Delta t)u_{i,j}^n + r(u_{i+1,j}^n + u_{i,j+1}^n + u_{i-1,j}^n + u_{i,j-1}^n - 4u_{i,j}^n) + c_{i,j}^1\Delta t \quad (4)$$

$$v_{i,j}^{n+1} = (1 - b_{i,j}\Delta t)v_{i,j}^n + r(v_{i+1,j}^n + v_{i,j+1}^n + v_{i-1,j}^n + v_{i,j-1}^n - 4v_{i,j}^n) + c_{i,j}^2\Delta t, \quad (5)$$

where

$$b(x,y) = f_x(x,y)^2 + f_y(x,y)^2, \quad (6)$$
$$c^1(x,y) = b(x,y)f_x(x,y), \text{ and} \quad (7)$$
$$c^2(x,y) = b(x,y)f_y(x,y). \quad (8)$$

The notation adopted is the one proposed by Xu and Prince in [11], so that f_x and f_y correspond to the partial derivatives of f with respect to x and y; indices i, j and n correspond to x, y and t, respectively; Δt corresponds to the time step for each iteration, and $r = \frac{\mu \Delta t}{h^2}$, with $h = \Delta x = \Delta y$ (Δx and Δy correspond to the spacing between pixels).

According to numerical analysis theory [13], stability of equations (4) and (5) is guaranteed whenever the restriction

$$0 < \triangle t \leq \frac{h^2}{4\mu + h^2||b||}, \quad ||b|| = \max_{\forall_{i,j}} b_{i,j},$$

is verified. As can be concluded from the expressions above, convergence of the iterative process can be made faster on coarser images, i.e. for larger values of the spatial sampling h. On the other hand, smoother GVF fields, with larger values of the parameter μ, make the convergence rate slower. These last cases correspond to smaller values of the sampling period $\triangle t$.

B. Sensor measurements

Once defined the target contour identification procedure, it is important to make a brief overview on the way this information is used. The measurements that will be provided to the estimation process are the target center coordinates (u, v) and its distance (d) to the origin of world reference frame.

Target center coordinates in each acquired image are computed easily from its estimated contour, as being the mean of the coordinates of the points that belong to this contour. Target distance to the origin of world reference frame is computed from its estimated boundary. Its real dimensions in the 3D world, and the knowledge of the camera intrinsic and extrinsic parameters, allow to establish metric relations between image and world quantities. Estimates on the depth of the target can then be obtained. A complete stochastic characterization of these quantities can be found in [5] and these will be the measurements provided to the estimation method detailed next.

The use of triangulation methods for at least two cameras would allow the computation of the target distance without further knowledge on the target. However, the present tracking system uses a single camera. Thus, additional information must be available. In this work, it is assumed that the target dimensions are known.

V. Tracking System

In this section, the nonlinear estimation methods implemented are described. Estimates on the target position, velocity and acceleration, in the 3D world, are provided and angular velocity is identified. This estimator is based on measurements from the previously computed target center coordinates and distance to the origin of world reference frame.

A. Extended Kalman filter

The Kalman filter [14] provides an optimal solution to the problem of estimating the state of a discrete time process that is described by a linear stochastic difference equation. However, this approach is not valid when the process and/or the measurements are nonlinear. One of the most successful approaches, in these situations, consists in applying a linear time-varying Kalman filter to a system that results from the linearization of the original nonlinear one, along the estimates. This kind of filters are usually referred to as Extended Kalman filters (EKF) [14], and have the advantage of being computationally efficient, which is essential in real time applications.

Consider a nonlinear system with state $\mathbf{x} \in \Re^n$ expressed by the nonlinear stochastic difference equation

$$\mathbf{x}_k = f(\mathbf{x}_{k-1}, \mathbf{u}_{k-1}, \mathbf{w}_{k-1}),$$

and with measurements available $\mathbf{z} \in \Re^m$ given by

$$\mathbf{z}_k = h(\mathbf{x}_k, \mathbf{v}_k),$$

where the index k represents time, $\mathbf{u}_k$ the control input, and $\mathbf{w}_k \in \Re^n$ and $\mathbf{v}_k \in \Re^m$ are random variables that correspond to the process and measurement noise, respectively. These variables are assumed to be independent, i.e. $E[\mathbf{w}_k {\mathbf{v}_k}^T] = 0$, and with Gaussian probability density functions with zero mean and given covariance matrices.

In the case of linear dynamic systems, the estimates provided by the Kalman filter are optimal, in the sense that the mean square estimation error is minimized. Estimates computed by EKF are suboptimal. It is even possible that it does not converge to the system state in some situations. However, the good performance observed in many practical applications made this strategy the most successful and popular in nonlinear estimation.

The implementation of an EKF requires a mathematical model to the target and sensors used. The choice of appropriate models is extremely important since it improves significantly the target tracking system performance, reducing the effects of the limited observation data available in this kind of applications. Given the movements expected for the targets to be tracked, the 3D *Planar Constant-Turn Model* as presented in [15], was selected. This model considers the vector $\mathbf{x} = [\mathsf{x}, \dot{\mathsf{x}}, \ddot{\mathsf{x}}, \mathsf{y}, \dot{\mathsf{y}}, \ddot{\mathsf{y}}, \mathsf{z}, \dot{\mathsf{z}}, \ddot{\mathsf{z}}]^T$ as the state of the target, where $[\mathsf{x}, \mathsf{y}, \mathsf{z}]$, $[\dot{\mathsf{x}}, \dot{\mathsf{y}}, \dot{\mathsf{z}}]$, and $[\ddot{\mathsf{x}}, \ddot{\mathsf{y}}, \ddot{\mathsf{z}}]$ are the target position, velocity, and acceleration in the world, respectively.

The sensor measurements available in each time instant, that define function h, correspond to the target center coordinates (u, v) and target distance (d) to the origin of world reference frame, and are given by

$$u = \frac{p_{11}x + p_{12}y + p_{13}z + p_{14}}{p_{31}x + p_{32}y + p_{33}z + p_{34}} + v_u$$
$$v = \frac{p_{21}x + p_{22}y + p_{23}z + p_{24}}{p_{31}x + p_{32}y + p_{33}z + p_{34}} + v_v$$
$$d = \sqrt{x^2 + y^2 + z^2} + v_d, \quad (9)$$

where p_{ij} is the projection matrix element in the line i and column j, and $\mathbf{v} = [v_u, v_v, v_d]^T$ is the measurement

noise (the time step subscript k was omitted for simplicity of notation). The measurement vector is given by $\mathbf{z} = [u, v, d]^T$.

The complete measurement process characterization requires also the definition of the measurement noise covariance matrix. This matrix can be obtained from an accurate study of the available sensors, which, in this project, consisted in executing a set of experiments aiming to compute the standard deviation of the estimation error in the image coordinates of a 3D world point, and the standard deviation of the error in target depth estimation, as detailed in [5].

B. Multiple-model

The model considered to the target requires the knowledge of its angular velocity. However, this value is not known in real applications, which led us to the application of a multiple model based approach, identifying simultaneously some parameters of the system and estimating its state.

The implemented method, known as *Multiple-Model Adaptive Estimator (MMAE)* [16], considers several models to a system that differ in a parameters set (in this case the target angular velocity). Each one of these models includes an extended Kalman filter, whose state estimates are mixed properly. The individual estimates are combined using a weighted sum with the *a posteriori* hypothesis probabilities of each model as weighting factors, leading to the state estimate

$$\widehat{\mathbf{x}}_k = \sum_{j=1}^{N} p_k^j \widehat{\mathbf{x}}_k^j,$$

and the global covariance matrix

$$\mathbf{P}_k = \sum_{j=1}^{N} p_k^j [\mathbf{P}_k^j + (\widehat{\mathbf{x}}_k^j - \widehat{\mathbf{x}}_k)(\widehat{\mathbf{x}}_k^j - \widehat{\mathbf{x}}_k)^T],$$

where $\mathbf{P}_k^j$ is the estimate error covariance of the model j and p_k^j is its *a posteriori* probability.

It should be stressed that the methods used to compute the *a posteriori* probability for each model and the final state estimate are optimal, if each one of the individual estimates is optimal. However, this is not the case in this application, since the known solutions to nonlinear estimation problems at present can only provide sub-optimal results.

VI. Experimental Results

In this section some brief considerations about the developed system are made, and experimental results of its application to real time situations are analyzed.

A. Application description

The positioning and tracking system proposed in this project was implemented in *Matlab*, and can be divided into three main modules: one that addresses the interface with the camera, other that implements the image processing algorithms, and a third related to the estimation process.

1) Interface with the camera: since the camera used in this project has a discrete and limited range of movements, its orientation in each time instant is determined according to a decision system whose aim is to avoid that the distance between the image and the target centers exceed certain values.

The CCD sensor built-in the camera acquires images with a maximum dimension of $640 \times 480\, pixels$, which is the resolution chosen for this application. Despite its higher computational requirements, smaller targets can be tracked with an increase on the accuracy of the system.

2) Image processing: the active contour method was implemented with the values of α and β equal to 0.5 and 0.05, respectively, since these values were the ones that led to better results.

The developed application is optimized to follow red targets, whose identification in acquired images is easy, since image segmentation is itself a very complex domain, and does not correspond to the main focus of this work.

3) Estimation process: the adopted MMAE approach was based on the utilization of four initially equiprobable target models, that differ on target angular velocity values: $2\pi\frac{1}{50}[0, 1, 2, 3]\, rad/s$.

Each one of the referred models requires the knowledge of the power spectral density matrix of the process noise, that is not available. The matrix considered to this quantity was $diag[0.1, 0.1, 0.1]$, since it led to the best experimental results.

The sampling interval of the application was made variable, however the use of the previously referred parameters imposed an inferior bound of approximately $0.5\, s$.

B. Application performance

The results presented in this section are relative to the tracking of a red balloon attached to a robot *Pioneer P3-DX*, as depicted in Fig. 2, programmed to describe a circular trajectory.

Fig. 2. Real time target tracking. Left: Experimental setup; Right: Target identification, where the initial snake is presented in black, its temporal evolution is presented in red, and the contour final estimate is presented in blue.

In Fig. 3, the 3D nominal and estimated target trajectories are presented. The position, velocity, and acceleration estimation errors are presented in Fig. 4. These quantities have large transients in the beginning of the experiment, due to the initial state estimation error, and decrease quickly to values beneath $20\, cm$, $4\, cm/s$, and $0.5\, cm/s^2$, respectively. There

are several reasons that can justify the errors observed: i) the uncertainty associated with the characterization of the real trajectory described by the target; ii) possible mismatches in the models considered for the camera and target, and iii) incomplete measurement and sensor noise characterization.

Moreover, given the suboptimal nature of the results produced by the extended Kalman filter in nonlinear applications, in some experiments, where an excessively poor initial state estimate was tested, divergence of the filter occurred.

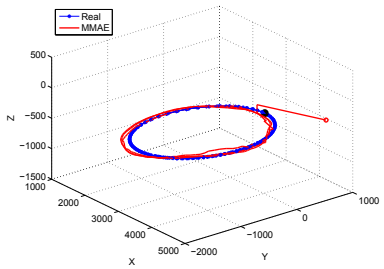

Fig. 3. 3D position estimate of a real target. The real position of the target in the initial instant is presented in black.

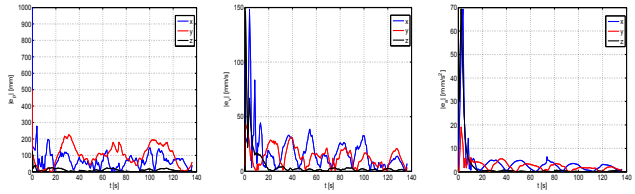

Fig. 4. Position (left pan), velocity (center pan), and acceleration (right pan) estimation error of a real target in the world.

The results of the adopted MMAE approach are presented in Fig. 5. For the trajectory reported the real target angular velocity is $2\pi 0.0217\, rad/s$. Thus, the probability associated to the model closer to the real target tends to 1 along the experiment, as depicted on the left panel of Fig. 5. On the right panel of that figure, the real and estimated angular velocities are plotted.

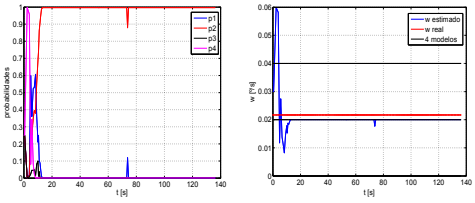

Fig. 5. MMAE evolution over time. On the left are the *a posteriori* hypothesis probabilities. On the right are the real (red) and estimated (blue) target angular velocities.

In what concerns the operation range for the proposed system, it depends significantly on the camera used and on the size of the target. In the experiments reported, an elliptic shape with axes of length $106\, mm$ and $145\, mm$ was located with the mentioned accuracies up to distances of approximately $7\, m$ from the camera. The lower bound of the range of distances in which the application works properly is limited by the distance at which the target stops being completely visible, filling the camera field of vision. For the target considered, this occurs at distances bellow $40\, cm$.

VII. Conclusions and Future Work

A new indoor positioning and tracking system architecture is presented, supported on suboptimal stochastic multiple-model adaptive estimation techniques. The proposed approach was implemented using a single low cost pan and tilt camera, estimating the real time location of a target which moves in the 3D real world with accuracies on the order of $20\, cm$.

The main limitations of the implemented system are the required knowledge on the target dimensions, and the inability to identify targets with colors other than red.

In the near future, an implementation of the developed architecture in **C** will be pursued, which will allow for the tracking of more unpredictable targets. Also, an extension of the proposed architecture to a multiple camera based system is thought. Distances to targets will then be computed resorting to triangulation methods, thus avoiding the requirement on the precise knowledge of their dimensions.

Finally, it is also recommended the integration of a sensor in the vision system that retrieves camera orientation in each time instant, and the implementation of an image segmentation algorithm that can identify a wider variety of targets.

References

[1] K. Kolodziej and J. Hjelm, *Local Positioning Systems: LBS Applications and Services*. CRC Press, 2006.

[2] Y. Bar-Shalom, X. Rong-Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation: Theory Algorithms and Software*. John Wiley & Sons, Inc., 2001.

[3] J. Borenstein, H. R. Everett, and L. Feng, *Where am I? Sensors and Methods for Mobile Robot Positioning*. Editado e compilado por J. Borenstein, 1996.

[4] O. Faugeras and Q. Luong, *The geometry of multiples images*. MIT Press, 2001.

[5] T. Gaspar, "Sistemas de seguimento para aplicações no interior," Master's thesis, Instituto Superior Técnico, 2008.

[6] T. Thormahlen, H. Broszio, and I. Wassermann, "Robust line-based calibration of lens distortion from a single view," *Mirage 2003*, pp. 105–112, 2003.

[7] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *Int. J. Comput. Vis.*, vol. 1, pp. 321–331, 1987.

[8] F. Leymarie and M. D. Levine, "Tracking deformable objects in the plane using an active contour model," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 15, pp. 617–634, 1993.

[9] D. Terzopoulos and K. Fleischer, "Deformable models," *Vis. Comput.*, vol. 4, pp. 306–331, 1988.

[10] T. McInerney and D. Terzopoulos, "A dynamic finite element surface model for segmentation and tracking in multidimensional medical images with application to cardiac 4d image analysis," *Comput. Med. Imag. Graph*, vol. 9, pp. 69–83, 1995.

[11] C. Xu and J. Prince, "Snakes, shapes, and gradient vector flow," *IEEE Trans. Image Processing*, vol. 7, pp. 359–269, 1998.

[12] I. M. Gelfand and S. V. Fomin, *Calculus of Variations*. Dover Publ., 2000.

[13] W. F. Ames, *Numerical Methods for Partial Differential Equations*, 3rd ed. New York: Academic, 1992.

[14] A. Gelb, *Applied Optimal Estimation*. Cambridge, Massachusetts: MIT Press, 2001.

[15] X. R. Li and V. P. Jilkov, "Survey of maneuvering target tracking. part i: Dynamic models," *IEEE Transactions on Aerospace and Electronic Systems*, pp. 1333–1364, 2003.

[16] M. Athans and C. Chang, *Adaptive Estimation and Parameter Identification using Multiple Model Estimation Algorithm*. Lexington, Mass.: MIT Lincoln Lab., June 1976.