A SINGLE PAN AND TILT CAMERA ARCHITECTURE FOR INDOOR POSITIONING AND TRACKING

T. Gaspar and P. Oliveira
IST/ISR, Lisboa, Portugal
tgaspar@isr.ist.utl.pt, pjcro@isr.ist.utl.pt

Keywords: Indoor positioning and tracking systems, Camera calibration, GVF snakes, Multiple-model adaptive estimation, Single camera vision systems.

Abstract: A new architecture for indoor positioning and tracking is proposed, based on a single low cost pan and tilt camera, where three main modules can be identified: one related to the interface with the camera, supported on 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. To assess the performance of the proposed methods and this new architecture, a software package was developed. An accuracy of $20\,cm$ was obtained in a series of indoor experimental tests, for a range of operation of up to ten meter, under realistic real time conditions.

1 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 (Kolodziej and Hjelm, 2006), (Bar-Shalom et al., 2001) and (Borenstein et al., 1996).

In outdoor applications, the NAVSTAR Global Positioning System (GPS) has been widely explored with satisfactory results for most of the actual needs. Indoor positioning systems based on this technology however face some undesirable effects, like multipath and strong attenuation of the electromagnetic waves, precluding their use.

Alternative techniques, such as infrared radiation, ultrasounds, radio frequency, vision has been successfully exploited as reported in detail in (Kolodziej and Hjelm, 2006), and summarized in (Gaspar, 2008).

The indoor tracking system proposed in this project uses vision technology, since this technique has a growing domain of applicability and allows to achieve acceptable 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 exceeds the objectives of this paper, due to space limitations. The reader interested can found these issues discussed in detail in (Gaspar, 2008).

This document is organized as follows. In section 2 the architecture of the developed positioning and tracking system is introduced, as well as the main methodologies and algorithms developed. In section 3 the camera and lens models are briefly introduced. To isolate and identify the target, advanced image processing algorithms are discussed in section 4, and in section 5, the used multiple-model nonlinear estimation technique is introduced. In the last two sections, 6 and 7, experimental results of the developed system, and concluding remarks and comments on future work, respectively, are presented.

2 SYSTEM ARCHITECTURE

In this project a new architecture for indoor positioning and tracking is proposed, based on three main

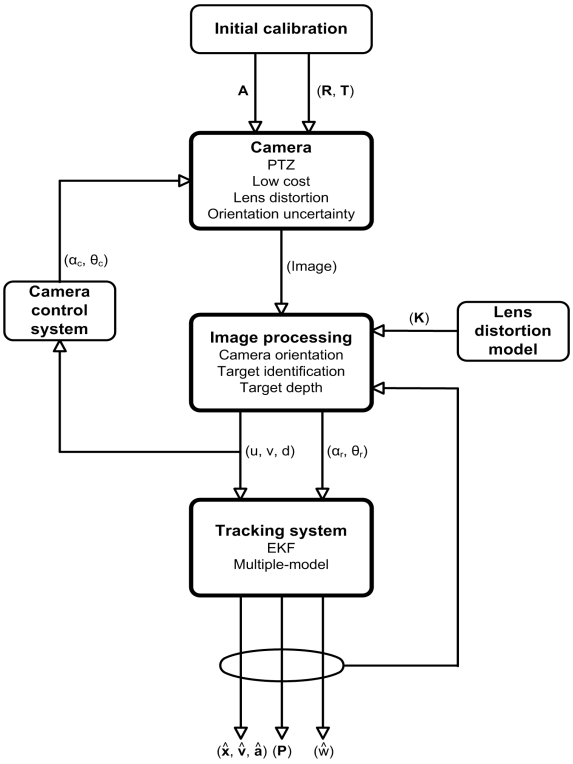

Figure 1: Tracking system architecture.

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, and is described next [1].

The extraction of physical information from an image acquired by a camera, requires the knowledge of its intrinsic (**A**) and extrinsic (**R** and **T**) parameters, which are computed during the initial calibration process. In this paper, calibration was preceded by an independent determination of a set of parameters (**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 requires the development of an external algorithm capable of estimate its instantaneous pan α_r and tilt θ_r angles.

The target identification is the main purpose of the image processing block. An active contour method, usually denominated as snakes, was selected to track the important features in the image. The approach selected consists of 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 a 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 were included in a multiple-model adaptive estimation methodology, that provides estimates on the system state ($\widehat{\mathbf{x}}$, $\widehat{\mathbf{v}}$, and $\widehat{\mathbf{a}}$), identifies the unknown target angular velocity $\widehat{w}$, and the estimation error covariance P, as depicted in Fig. 1.

The command for 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 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.

3 SENSOR: PTZ CAMERA

3.1 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 (Faugeras and Luong, 2001).

Let $\mathbf{M} = [x,y,z,t]^T$ be the homogeneous coordinates of a visible point, in the world reference frame, and $\mathbf{m} = [u,v,s]^T$ the corresponding homogeneous coordinates of the same point in the image frame. According to this model, the relation between the coordinates expressed in these two coordinate frames is given by

$$\lambda \mathbf{m} = \mathbf{PM}, \tag{1}$$

where λ is a multiplicative constant, related with the distance from the point in space to the camera, and **P** the projection matrix that relates 3D world coordinates and 2D image coordinates. The transformation given by this matrix can be decomposed into

[1] In this section some quantities are presented informally to augment the legibility of the whole document.

three others: one between world and camera coordinate frames, expressed by ${}^{c}\mathbf{g}_{M}$ in homogeneous coordinates; other responsible for projecting 3D points into the image plane, represented by π, and a third one that changes the origin and units of the coordinate system used to identify each point in the acquired images, denoted as **A**. The product of the three previous transformations results in the overall expression for the matrix **P**, which is given by $\mathbf{P} = \mathbf{A}.\pi.{}^{c}\mathbf{g}_{M}$, and establishes the relation between a point in the 3D world and its correspondent in the acquired images.

The use of the previous model implies the determination of the intrinsic and extrinsic parameters referred before. In this work, the classical approach proposed by Faugeras (Faugeras and Luong, 2001) was selected and implemented. The disadvantages of this method are: i) the required preparation of the scene in which the camera is inserted, and ii) the distortion of the lens is disregarded. However, the impact of these requirements is moderate since the camera in this application is supposed to be placed in a fixed location in the world (the calibration needs to be performed just once). A separate algorithm that compensates for lens distortion is implemented, see section 3.3 for details. The major advantages are that only one image is required and reliable results can be obtained.

The classical method proposed by Faugeras consists in performing an initial estimation of the projection matrix, that is done from a set of points with known coordinates in world and camera reference frames. Writing (1) and reorganizing the expression obtained to every one of the n points used in the calibration process, and considering that the index i identifies the coordinates of the i^{th} used point, yields, for each point,

$$\begin{bmatrix} x_i & y_i & z_i & 1 & 0 & 0 & 0 & 0 & -u_i x_i & -u_i y_i & -u_i z_i & -u_i \\ 0 & 0 & 0 & 0 & x_i & y_i & z_i & 1 & -v_i x_i & -v_i y_i & -v_i z_i & -v_i \end{bmatrix} \cdot \mathbf{p} = 0,$$

with

$$\mathbf{p} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} & p_{21} & p_{22} & p_{23} & p_{24} & p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix}^{T},$$

where p_{jk} is the **P** element whose line and column are j and k, respectively.

The previous equations, when applied to the entire set of used points, lead to a system of the form $\mathbf{Lp} = 0$, where **L** is a $2n \times 12$ matrix. The solution of this system corresponds to the eigenvector associated with the smallest eigenvalue of $\mathbf{L}^{T}\mathbf{L}$, or, equivalently, to the singular vector of **L** associated with the smallest singular value of its Single Value Decomposition. Since the projection matrix has 12 elements, and each point considered contributes with two equations, there is a minimum of 6 points that must be used in the calibration process. The intrinsic and extrinsic parameters of the camera can then be computed from the estimated **p** vector as

$$\begin{array}{rclrcl} u_0 & = & \mathbf{p}_1 \cdot \mathbf{p}_3, & v_0 & = & \mathbf{p}_2 \cdot \mathbf{p}_3, \\ |\alpha_u| & = & ||\mathbf{p}_1 - u_0 \mathbf{p}_3||, & |\alpha_v| & = & ||\mathbf{p}_2 - v_0 \mathbf{p}_3||, \\ \mathbf{r}_3 & = & \mathbf{p}_3, & \mathbf{r}_2 & = & \frac{\mathbf{p}_2 - v_0 \mathbf{r}_3}{\alpha_v}, \\ \mathbf{r}_1 & = & \frac{\mathbf{p}_3 - u_0 \mathbf{r}_3}{\alpha_u}, & t_z & = & p_{34}, \\ t_x & = & \frac{p_{14} - u_0 t_z}{\alpha_u}, & t_y & = & \frac{p_{24} - v_0 t_z}{\alpha_v}, \end{array}$$

where $\mathbf{p}_k = \begin{bmatrix} p_{k1} & p_{k2} & p_{k3} \end{bmatrix}$, and $\mathbf{p}_i.\mathbf{p}_j$ represents the internal product of the vectors $\mathbf{p}_i$ and $\mathbf{p}_j$, see (Faugeras and Luong, 2001) and (Gaspar, 2008) for details.

3.2 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.

Since the used *Creative WebCam Live! Motion* camera has a closed architecture, its internal geometry model was estimated from the analysis of its external structure and 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 this transformations leads to the global transformation between world and camera reference frames:

$${}^{c}\mathbf{g}_{M} = {}^{M}\mathbf{g}_{c}^{-1}, \qquad {}^{M}\mathbf{g}_{c} = {}^{M}\mathbf{g}_{0}\,{}^{0}\mathbf{g}_{1}\,{}^{1}\mathbf{g}_{2}\,{}^{2}\mathbf{g}_{3}\,{}^{3}\mathbf{g}_{c},$$

that is fundamental to determine the camera projection matrix over time.

The expressions introduced require, however, 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 performed on an initial stage resorting to points of the world with known coordinates.

3.3 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 and no vestige of tangential distortion was identified. Therefore, it is necessary to compensate this distortion by 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 (Thormahlen et al., 2003), that results in

$$x = x_d + x_d \sum_{i=0}^{\infty} k_i r_d^{i-1} \quad and \quad y = y_d + y_d \sum_{i=0}^{\infty} k_i r_d^{i-1},$$

where

$$r_d = \sqrt{x_d^2 + y_d^2}.$$

The lens distortion compensation method adopted in this project 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 the parameters k_3, k_5, x_0, y_0, m_i, b_i, $i = 1,\ldots,N_r$ are obtained. vfill

4 IMAGE PROCESSING

4.1 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 (Kass et al., 1987), 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 (Kass et al., 1987), image segmentation (Leymarie and Levine, 1993), shape modeling (Terzopoulos and Fleischer, 1988), (McInerney and Terzopoulos, 1995), or motion tracking (Leymarie and Levine, 1993), as happens in this application.

In this project a *parametric active contour* method is used (Kass et al., 1987), 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, as can be seen, 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. Traditionally, this energy can be expressed in the form

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

where the 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 (2) by the spacial finite differences method, with step h, yields

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

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 the expression

$$\mathbf{x}^{n+1} = \mathbf{x}^n + \tau\mathbf{x}_t^n,$$

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.

4.2 Sensor Measurements

Once obtained the target contour, it is possible to compute the measurements that will be provided to the estimation process: 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 as being the mean of the coordinates of the points that belong to the target 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, allows to establish metric relations between image and world quantities. Estimates on the depth of the target can then be obtained. A complete stochastic characterization can be found in (Gaspar, 2008) and will be the measurements considered as inputs to the estimation method used.

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.

5 TRACKING SYSTEM

In this section, the implemented nonlinear estimation methods is 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.

5.1 Extended Kalman Kilter

The Kalman filter (Gelb, 2001) 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 nod 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) (Gelb, 2001), 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 covariance matrices $\mathbf{Q}_k$ and $\mathbf{R}_k$, respectively.

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, revealed this strategy as 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 (Li and Jilkov, 2003), was selected. This model considers the vector $\mathbf{x} = [\mathrm{x}, \dot{\mathrm{x}}, \ddot{\mathrm{x}}, \mathrm{y}, \dot{\mathrm{y}}, \ddot{\mathrm{y}}, \mathrm{z}, \dot{\mathrm{z}}, \ddot{\mathrm{z}}]^T$ as the state of the target, where $[\mathrm{x}, \mathrm{y}, \mathrm{z}]$, $[\dot{\mathrm{x}}, \dot{\mathrm{y}}, \dot{\mathrm{z}}]$, and $[\ddot{\mathrm{x}}, \ddot{\mathrm{y}}, \ddot{\mathrm{z}}]$ are the target position, velocity, and acceleration in the world, respectively.

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

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

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 $\mathbf{z} = [u, v, d]^T$.

Next, a standard notation is used (see (Gelb, 2001) for details) to describe each Kalman filter:

Predict step

$$\begin{aligned} \widehat{\mathbf{x}}_k^- &= f(\widehat{\mathbf{x}}_{k-1}, \mathbf{u}_{k-1}, 0) \\ \mathbf{P}_k^- &= \mathbf{A}_k \mathbf{P}_{k-1} \mathbf{A}_k^T + \mathbf{W}_k \mathbf{Q}_{k-1} \mathbf{W}_k^T \end{aligned}$$

Update step

$$\begin{aligned} \mathbf{K}_k &= \mathbf{P}_k^- \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_k^- \mathbf{H}_k^T + \mathbf{V}_k \mathbf{R}_k \mathbf{V}_k^T)^{-1} \\ \widehat{\mathbf{x}}_k &= \widehat{\mathbf{x}}_k^- + \mathbf{K}_k (\mathbf{z}_k - h(\widehat{\mathbf{x}}_k^-, 0)) \\ \mathbf{P}_k &= (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_k^-, \end{aligned}$$

where $\mathbf{K}_k$ is the Kalman filter gain.

The complete measurement process characterization requires also the definition of the measurement noise covariance matrix $\mathbf{R}$. 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.

5.2 Multiple-model

The model considered for 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 Estimation (MMAE)* (Athans and Chang, 1976), considers several models to a system that differ in a parameter 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

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

with 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 p_k^j corresponds to the *a posteriori* probability of the model j, at the time instante k, and N to the number of considered models.

It should be stressed that the methods used to compute the *a posteriori* probabilities of 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 do not provide optimal results.

6 EXPERIMENTAL RESULTS

In this section some brief considerations about the developed positioning and tracking system are advanced, and the experimental results of its application to real time situations are presented.

6.1 Application Description

The architecture for positioning and tracking 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.

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 640×480 *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.

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.

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 models requires the knowledge of the power spectral density matrix of the process noise, that is not available. After some preliminary tuning, the matrix considered for this quantity was set to $diag[0.1, 0.1, 0.1]$.

The sampling interval of the developed application was made variable, however for the parameters previously discussed, a lower bound of approximately $0.5\,s$ was found.

6.2 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.

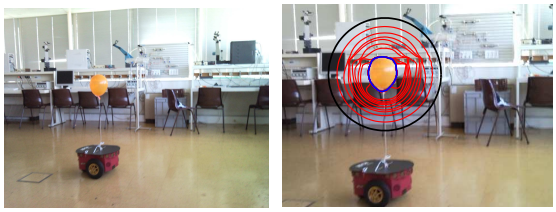

Figure 2: Real time target tracking. Left: Experimental setup; Right: Target identification, where the initial snake is presented in black, its temporal evolution is presented in red, and the contour final estimate is presented in blue.

In Fig. 3, the 3D nominal and estimated target trajectories are presented. The target position, velocity and acceleration along time are depicted in Fig. 4. Despite the significant initial uncertainty in the state estimate, the target position, velocity, and acceleration estimates converge to the vicinity of the real values. Moreover, given the suboptimal nature of the results produced by the extended Kalman filter in nonlinear applications, in some experimental cases where an excessively poor initial state estimate was tested, divergence of the filter occurred.

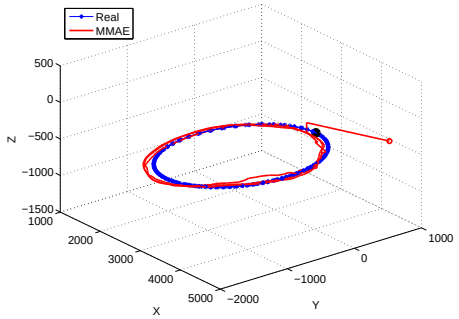

Figure 3: 3D position estimate of a real target. The real position of the target in the initial instant is presented in black.

The position, velocity, and acceleration estimation errors are presented in Fig. 5. These quantities

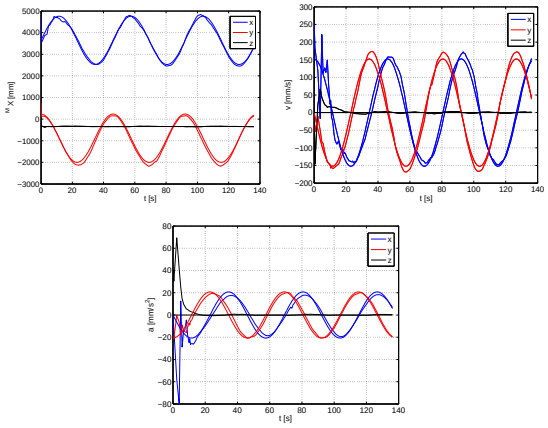

Figure 4: Position (top left pan), velocity (top right pan), and acceleration (bottom pan) estimates of a real target in the world. The slender and tickler lines correspond to the estimated and real values, respectively.

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, and ii) possible mismatches between the models considered for the camera and target, and iii) incorrect measurement and sensor noise characterization.

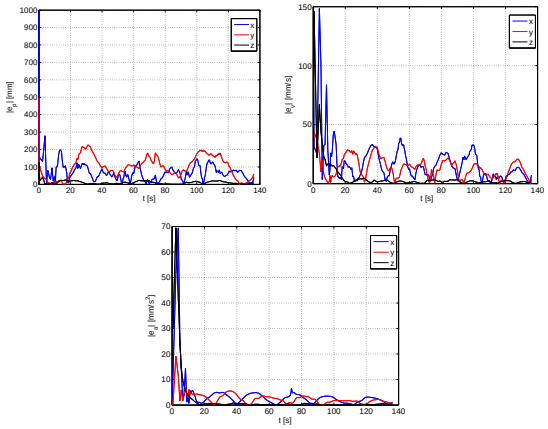

Figure 5: Position (top left pan), velocity (top right pan), and acceleration (bottom pan) estimation error of a real target in the world.

The results of the adopted MMAE approach are presented in Fig. 6. 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. 6. On the right panel of that figure, the real and estimated angular velocities are plotted.

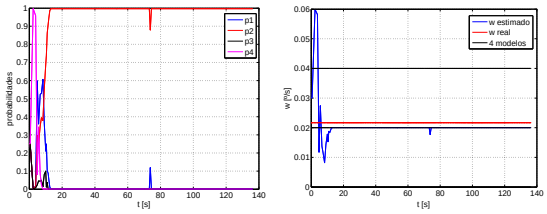

Figure 6: MMAE evolution over time. On the left, the *a posteriori* hypothesis probabilities. On the right, real (red) and estimated (blue) target angular velocity.

In what concerns the range of operation for the proposed system, it depends significantly on the camera used and on the size of the target to be tracked. In the experiments reported, an elliptic shape with axes of length 106 *mm* and 145 *mm*, was identified and located, with the mentioned accuracies for distances up to 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*.

7 CONCLUSIONS AND FUTURE WORK

A new architecture for indoor positioning and tracking 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 advised 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.

ACKNOWLEDGEMENTS

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/55609/2004 - RUMOS of the FCT.

REFERENCES

Athans, M. and Chang, C. (1976). *Adaptive Estimation and Parameter Identification using Multiple Model Estimation Algorithm*. MIT Lincoln Lab., Lexington, Mass.

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

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

Faugeras, O. and Luong, Q. (2001). *The geometry of multiples images*. MIT Press.

Gaspar, T. (2008). Sistemas de seguimento para aplicações no interior. Master's thesis, Instituto Superior Técnico.

Gelb, A. (2001). *Applied Optimal Estimation*. MIT Press, Cambridge, Massachusetts.

Kass, M., Witkin, A., and Terzopoulos, D. (1987). Snakes: Active contour models. *Int. J. Comput. Vis.*, 1:321–331.

Kolodziej, K. and Hjelm, J. (2006). *Local Positioning Systems: LBS Applications and Services*. CRC Press.

Leymarie, F. and Levine, M. D. (1993). Tracking deformable objects in the plane using an active contour model. *IEEE Trans. Pattern Anal. Machine Intell.*, 15:617–634.

Li, X. R. and Jilkov, V. P. (2003). Survey of maneuvering target tracking. part i: Dynamic models. *IEEE Transactions on Aerospace and Electronic Systems*, pages 1333–1364.

McInerney, T. and Terzopoulos, D. (1995). 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*, 9:69–83.

Terzopoulos, D. and Fleischer, K. (1988). Deformable models. *Vis. Comput.*, 4:306–331.

Thormahlen, T., Broszio, H., and Wassermann, I. (2003). Robust line-based calibration of lens distortion from a single view. *Mirage 2003*, pages 105–112.