2011 IEEE International Conference on Robotics and Automation
Shanghai International Conference Center
May 9-13, 2011, Shanghai, China

Monocular Depth from Focus Estimation with Complementary Filters

Tiago Gaspar and Paulo Oliveira

Abstract— **In this paper, new methodologies for the estimation of the depth of a target with unknown dimensions, based on depth from focus strategies, are proposed. The image-based measurements are detailed, through the minimization of a new functional, deeply rooted on optical characteristics of the lens system, namely the point spread function. This work complements an inexpensive single pan and tilt camera-based indoor positioning and tracking system, resorting to complementary filters for depth estimation. A motivation example is provided, where the target dimensions are assumed as known. Then, an extension corresponding to an higher-order filter is presented, that tackles the problem at hand. 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. A centimetric accuracy was obtained under realistic conditions.**

I. Introduction

With the development and widespread use of autonomous robotic vehicles, 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 for the position of a robot is essential to its navigation and control systems, which justifies the significant effort that has been put into this domain, see [1]–[3].

Successfully exploited techniques have been reported, such as infrared radiation, ultrasound, radio frequency, and vision, see details in [1]. The indoor tracking system addressed in this work resorts to vision technology, since this technique has a growing domain of applicability and allows to achieve interesting results with very low investment, see the comprehensive survey on monocular 3D tracking in [4]. This system estimates in real time the position, velocity, and acceleration of a target that evolves along an unknown trajectory in the 3D world, as well as its angular velocity. These estimates are obtained resorting to suboptimal stochastic multiple-model adaptive estimation techniques that explore information provided by a single camera.

In monocular configurations, the estimation of the depth of the target in relation to the camera is a key factor, since the use of triangulation methods, typical in multi-camera approaches, is not possible. When a single camera is used, the depth of a point in the 3D world can be estimated by exploring the relation between this quantity and the amount of blur that corrupts the projection of the point into acquired images. This is done by modelling the influence that some of the camera intrinsic parameters have on images acquired with a small depth of field. Based upon this principle, there are three main strategies that have been explored: depth from blur by focusing [5], [6], zooming [7], and irising [8]. In this paper, we are mainly concerned with depth estimation from blur by focusing. Two different techniques based upon this approach can be found in the literature: depth from defocus [6], [8], and depth from focus [5], [9], [10]. The depth estimation strategy that will be proposed is based on this latter method, since this type of approach does not require a mathematical model for the blurring process of the camera, i.e. the point spread function responsible for the blurring does not need to be modeled.

In this work, depth measurements, obtained according to the strategy described in the previous paragraph, are combined with additional information extracted from acquired images, by means of a complementary filter. Complementary filters have been used in a wide variety of sensor fusion problems, such as attitude estimation [11] or flight control [12]. These distortionless filters merge information provided by a given sensor suite over distinct, yet complementary, frequency regions. In the linear time-invariant setting, the filter design reduces to the problem of decomposing the identity operator into stable low- and high-pass transfer functions, which operate on complementary sensor information. The cutoff frequency of the transfer functions becomes a tuning parameter that must match the physical characteristics of the sensors. By exploring sensors redundancy, this strategy rejects measurement disturbances in complementary frequency regions without distorting the original signal [13].

This work is an evolution of a framework recently proposed for target tracking and positioning [14], where a low cost single pan and tilt camera-based indoor positioning and tracking system was presented. The focus of this paper is on the depth estimation module of that system. A novel strategy to estimate the depth of a target is proposed, which consists in a complementary filter that combines two different sources of information extracted from images acquired with a single camera: i) a measure of the target depth and ii) a biased measure of the target depth derivative over time (velocity of the target along the camera optical axis). Typically, the measurement of the depth of a target requires the use of at least two cameras, or the availability of further information about the target, such as its dimensions, for instance. However, it is possible to estimate this quantity using information from a single camera resorting to strategies based on the concept of depth from focus [8], [9]. The measurement of the depth derivative is obtained from the derivative of the dimensions of the target in acquired images. As before, establishing a relation between these two quantities would require further knowledge on the target dimensions. However, the use of a second-order complementary filter tackles this problem by estimating the bias that corrupts this measure, which is the

This work was partially funded by FCT (ISR/IST plurianual funding) through the PIDDAC Program and by the FCT project PTDC/EEA-CRO/111197/2009. The work of Tiago Gaspar was supported by the PhD Student Scholarship SFRH/BD/46860/2008, from FCT.

The authors are with the Institute for Systems and Robotics, Instituto Superior Técnico, Av. Rovisco Pais, 1049-001 Lisboa, Portugal (e-mails: {tgaspar,pjcro}@isr.ist.utl.pt).

result of assuming an incorrect value for the target unknown dimensions. Given the considerations above, this complementary filter estimates the instantaneous depth of targets describing arbitrary trajectories in the 3D world, without requiring the availability of further information about its dimensions and shape. A new monocular indoor positioning and tracking system results, which estimates in real time the target position, linear and angular velocities, and linear acceleration, for targets with unknown dimensions, see [14].

This document is organized as follows. The required measurements and the design and analysis of the proposed complementary filter, responsible for estimating the depth of the target, are described in sections II and III, respectively. In section IV, experimental results illustrating the performance of the proposed depth estimation algorithm are presented, and in section V, concluding remarks and directions of future work are addressed.

II. Depth Measurements

In this section, the process of obtaining the measurements used by the depth complementary filter is described. Estimates of the depth of the target are obtained using a depth from focus strategy, and the target depth derivative is estimated based on the variation of the target boundary in acquired images.

A. Target depth

The idea of inferring depth from focus is based on the concept of depth of field, which is a consequence of the inability of cameras to simultaneously focus planes on the scene at different depths.

Considering a thin model for the lens of the camera [15], it is possible to establish a nonlinear relation between the distance z from the lens to the plane that the camera can exactly focus at each instant of time, and the distance v between the lens and the image plane at which the projection of objects in the scene appears sharply focused. To complete the relation, the focal length f of the lens must be considered. This relation is known as the Gaussian Lens Formula [15], and can be rearranged in the form

$$z = \frac{fv}{v - f}. \tag{1}$$

The use of expression (1) to estimate the depth of a target moving in the scene requires the knowledge of both the focal length of the camera and the value of v, i.e. the value of the camera focus that minimizes the amount of blur that corrupts the projection of the target in acquired images. The estimation of this quantity requires the definition of a metric that quantifies the sharpness of a transition in an image. Metrics related with high-frequency energy contents in the image, Fourier transform, image gradient, or Laplacian, are detailed in [9]. Our goal is to estimate the depth of a target, therefore the metric proposed aims to maximize the image gradient magnitude across lines orthogonal to the target boundary, which, as described in [14], is obtained resorting to active contours, see [16] for details. This approach considers that the real target boundary is on a plane perpendicular to the camera optical axis, which is the plane that appears sharply focused when the camera focus value v_0 (i.e. the distance between the lens and the plane of the camera CCD sensor) is the one that optimizes the metric proposed. The plane in which the target boundary is considered to be is the plane that specifies the depth of the target. The problem at hand can be formulated as $\min_{v_0} \; g(v_0)$, where the cost function

$$g(v_0) = \frac{1}{\frac{1}{N_l}\sum_{i=1}^{N_l} \max_{(x,y)\in l_i} ||\nabla I_{v_0}(x,y)||^2} \tag{2}$$

is the inverse of the mean of the square of the image gradient magnitude maximum values across lines orthogonal to the target boundary. Moreover, N_l denotes the number of lines used, l_i the i-th line, ∇ the gradient operator, $||\cdot||$ the Euclidean norm, and $I_{v_0}(x,y)$ the intensity of the image acquired with the focus value v_0 at point (x,y). The formulation of this problem as the minimization of $g(v_0)$, instead of the maximization of its inverse, is based on the model that will be proposed for this function in the sequel.

In order to gain some insight into how to model the cost function proposed, consider that, for a given focus value v_0, acquired images are obtained from the convolution of the corresponding sharply focused image $I^f_{v_0}(x,y)$ with the point spread function $h(x,y)$ of the lens system, i.e. with the function that models the blurring process of the camera: $I_{v_0}(x,y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} I^f_{v_0}(\alpha,\beta)h(x-\alpha,y-\beta)\mathrm{d}\alpha\mathrm{d}\beta$.

A common model for the point spread function (PSF) is a circle of constant intensity. Let, in this situation, the PSF be

$$h(x,y) = \begin{cases} \frac{1}{\pi R_c^2} & x^2+y^2 \le R_c^2 \\ 0 & x^2+y^2 > R_c^2 \end{cases},$$

where R_c denotes the radius of the circle, and consider the existence of a vertical step in the sharply focused image of the form $I^f_{v_0}(x,y) = a_1 + a_2\, u(x-x_0)$, where $u(x-x_0)$ is the standard *unit step function* centered at point x_0, a_1 is the intensity of the image when $x < x_0$, and a_2 is the magnitude of the step. Thus, this approach profits from the target segmentation method used.

In this situation, it is straightforward to show that the partial derivative of $I_{v_0}(x,y)$ with respect to y is 0, since $I^f_{v_0}(x,y)$ does not depend on this variable, and differentiation and convolution are linear operations, thus they commute. Using this fact, and after some mathematical manipulation, it is also possible to show that the partial derivative of $I_{v_0}(x,y)$ with respect to x is 0, if $|x-x_0| > R_c$, and $\frac{2a_2}{\pi R_c^2}\sqrt{R_c^2-(x-x_0)^2}$, if $|x-x_0| \le R_c$. By considering a line l orthogonal to the boundary of the target, yields

$$\max_{(x,y)\in l} ||\nabla I_{v_0}(x,y)||^2\Big|_{x=x_0} = \left(\frac{2a_2}{\pi R_c}\right)^2.$$

Assuming a geometrical optics framework [15], which disregards the wave nature of light, and resorting to some trigonometric manipulations, it is possible to write the value of R_c as a function of the already defined quantities f, z, and v_0, and the diameter of the lens L, see [8] for details. Replacing the value of R_c in $(\frac{2a_2}{\pi R_c})^2$ by its expression, the cost function proposed in (2) may be rewritten as

$$g(v_0) = \frac{(f-z)^2 v_0^2 + 2fz(f-z)v_0 + (fz)^2}{[4fza_2/(L\pi)]^2}.$$

According to the discussion above, the cost function in (2) is expected to depend quadratically on v_0. Therefore, a quadratic model was considered for this function. Since three coefficients are enough to define the shape of a quadratic function, the acquisition of at least three images with different focus values provides at least three measurements of $g(v_0)$, one per focus value, which are enough to estimate the three coefficients of the cost function model. If three or more images are acquired, a system of linear equations results, which can be solved resorting to the standard linear least squares method [13]. The linear dependence of this model on the parameters that must be estimated is the reason why the minimization of $g(v_0)$ was considered, instead of the maximization of its inverse, which seemed more intuitive. The estimated coefficients can be easily converted into estimates of $v = \arg\min_{v_0} g(v_0)$, i.e. estimates of the camera focus value that minimizes the cost function for a given depth of the target, since this value corresponds to the one that minimizes the quadratic function. By repeating this procedure over time, successive estimates of the value of v result, and, as a consequence, estimates of the instantaneous depth z of the target can be computed resorting to (1).

The measurements of the target depth provided by the algorithm described in this section can be written in the form $z_m = z + z_d$, where z is the real target depth and z_d is the noise that corrupts its measurement.

B. Target depth derivative

Considering a pinhole model for the camera [17], the cartesian coordinates of a point in the camera reference frame (x, y, z) are related to the coordinates (x_p, y_p) of its projection into the image plane by expressions

$$x_p = f\frac{x}{z} \qquad \text{and} \qquad y_p = f\frac{y}{z}, \tag{3}$$

where the origin of the camera reference frame was considered to be coincident with the camera optical centre, and the origin of the image frame is in the image centre.

From the relations in (3), it is straightforward to show that the distance R, between two points in a plane at a distance z from the camera, and the distance r, between the projection of these points into the image plane, are related by

$$r = \frac{f}{z}R. \tag{4}$$

In particular, if two points of the real target, lying in the plane in which the target boundary is considered to be, are used to obtain a measure of the real target dimensions, they will verify this relation. However, the use of a distance between two points as a measure of the target dimensions would require a precise identification of those points in each image, which is a very difficult problem to solve, especially when the projection of the target appears with different orientations in different images.

In order to obtain a measure of the target dimensions invariant to rotations of the image of the target, consider that the coordinates $\mathbf{x} \in \mathbb{R}^2$, of a point of the curve that describes the target boundary, consist of two discrete random variables, and that the covariance of $\mathbf{x}$ is $\Sigma_{\mathbf{x}}$. Moreover, let $\mathbf{x}_a \in \mathbb{R}^2$ be the coordinates of a point of the curve that describes the boundary of a target in an image, and $\mathbf{x}_b = \mathbf{R}_{\mathbf{x}}\mathbf{x}_a$ the coordinates of the same point when the target boundary is rotated by an amount $\mathbf{R}_{\mathbf{x}}$, where $\mathbf{R}_{\mathbf{x}}$ is an element of the Special Orthogonal group $SO(2)$. Consider also that both quantities are random variables with covariance matrices $\Sigma_{\mathbf{x}_a}$ and $\Sigma_{\mathbf{x}_b}$. If $r_a = \sqrt{\text{tr}(\Sigma_{\mathbf{x}_a})}$ and $r_b = \sqrt{\text{tr}(\Sigma_{\mathbf{x}_b})}$ are the dimensions of the image of the target associated with $\mathbf{x}_a$ and $\mathbf{x}_b$, respectively, then

$$r_b = \sqrt{\text{tr}(\Sigma_{\mathbf{x}_b})} = \sqrt{\text{tr}(\mathbf{R}_{\mathbf{x}}\Sigma_{\mathbf{x}_a}\mathbf{R}_{\mathbf{x}}^T)} = \sqrt{\text{tr}(\Sigma_{\mathbf{x}_a}\mathbf{R}_{\mathbf{x}}^T\mathbf{R}_{\mathbf{x}})} = r_a,$$

since $\mathbf{R}_{\mathbf{x}}^T\mathbf{R}_{\mathbf{x}} = \mathbf{I}_{2\times 2}$, where $\mathbf{I}_{2\times 2}$ is the identity matrix of dimensions 2×2. Therefore, the square root of the trace of the covariance matrix associated with the boundary of the image of the target was used as a measure of its dimensions, since this quantity is invariant to rotations of the boundary of the target.

According to (4), and assuming that the focal length of the lens remains constant, it is possible to write the derivative of the depth of the target with respect to time in the form

$$\dot{z} = -\frac{\dot{r}}{r^2}Rf, \tag{5}$$

where r and $\dot{r}$ denote the square root of the trace of the covariance matrix associated with the boundary of the image of the target and its derivative with respect to time, respectively. Both quantities follow directly from the boundary of the target in the image, and their measurements are here denoted r_m and $\dot{r}_m$.

Relation (5) is a function of the value of R, which depends on the dimensions of the real target. However, when the dimensions of the target are not available, this quantity is not known. Therefore, an extra term γ, that takes this uncertainty into account, must be added to the value of R, resulting in the expression

$$\dot{z}' = \underbrace{-\frac{\dot{r}}{r^2}Rf}_{\dot{z}} \underbrace{-\frac{\dot{r}}{r^2}\gamma f}_{\beta}$$

for the target depth derivative. The value of $\dot{z}$ corresponds to the real target velocity in the direction of the camera optical axis, and β corresponds to a bias term that results from taking γ into account.

The measurements ψ_m of the target depth derivative provided by the method described can be written in the form

$$\psi_m = \psi + \beta + \psi_d + \beta_d, \tag{6}$$

where ψ denotes the real target depth derivative over time, ψ_d the noise that corrupts the measurements of this quantity, and β_d is a disturbance related to the bias value.

III. Depth Complementary Filter

In this section, a complementary filter that provides estimates for the depth of a moving target is proposed. Initially, for motivation, a simple continuous-time complementary structure for situations where the dimensions of the target are known is presented. Afterwards, this structure is modified to address the same problem when the dimensions of the target are not known. A rigorous formulation of the problem addressed in this section is presented next.

Problem statement 1: Consider a moving target with unknown dimensions and unknown position $\mathbf{p} = (x, y, z)$. Suppose that measurements

$$\begin{cases} z_m &= z + z_d \\ \psi_m &= \psi + \beta + \psi_d + \beta_d \end{cases}$$

of the target depth and its derivative are provided by a single camera, and that both quantities are corrupted by noise (z_d and ψ_d, respectively) in complementary frequency regions. These quantities are measured in relation to the camera reference frame. The value of the target depth derivative is affected by a bias term β, which results from the unknown nature of the target dimensions, and which is corrupted by a disturbance β_d. Given these assumptions, design a filter that provides an optimal solution in the minimum mean square error sense for the problem of estimating the instantaneous depth of the moving target.

A. First-order: known target dimensions

When the real target dimensions R are known, the measurements of the target depth derivative (6) are not biased, since the value of γ, and as a consequence the value of β, are null. A filter that estimates the target depth using measurements z_m and ψ_m is deduced below.

Let $z(s)$ and $\psi(s)$ denote the Laplace transforms of z and ψ, respectively. Then, for every $k > 0$, $z(s)$ admits the stable decomposition

$$z(s) = \underbrace{\frac{k}{s+k}}_{T_1(s)} z(s) + \underbrace{\frac{s}{s+k}}_{T_2(s)} z(s), \tag{7}$$

with $T_1(s)$ and $T_2(s)$ satisfying the equality $T_1(s)+T_2(s) = I$, where I denotes the identity operator.

Using relation $\psi(s) = sz(s)$, it follows from (7) that $z(s) = F_z(s)z(s) + F_\psi(s)\psi(s)$, which suggests a filter with the structure

$$\hat{z} = \mathcal{F}_z z_m + \mathcal{F}_\psi \psi_m, \tag{8}$$

where $\mathcal{F}_z$ and $\mathcal{F}_\psi$ are linear time-invariant operators with transfer functions $F_z(s)$ and $F_\psi(s)$, respectively. From the equations above, it is straightforward to deduce that the filter admits the state-space realization $\mathcal{F}$

$$\dot{\hat{z}} = \psi_m + k(z_m - \hat{z}). \tag{9}$$

Considering that $\mathcal{T}_1$ and $\mathcal{T}_2$ denote linear time-invariant operators with transfer functions $T_1(s)$ and $T_2(s)$, respectively, it is possible to rewrite (8) in the form

$$\hat{z} = (\mathcal{T}_1 + \mathcal{T}_2)z + \mathcal{F}_z z_d + \mathcal{F}_\psi \psi_d, \tag{10}$$

that shows that the estimate $\hat{z}$ provided by the filter consists of an undistorted copy $(\mathcal{T}_1 + \mathcal{T}_2)z = z$ of the original signal z, corrupted by the measurement noises z_d and ψ_d.

From the deduction above, it is possible to conclude that the filter proposed relies on information provided by the depth from focus algorithm at low frequencies only, since $T_1(s)$ corresponds to a low-pass filter. Moreover, the complementary filter derived blends the previous information with that from the target depth derivative at high frequencies, since $T_2(s) = I - T_1(s)$ corresponds to a high-pass filter. This decomposition into different frequency regions, that results from the complementary filter structure, holds the key to its practical success, as it mimics the natural frequency decomposition induced by the physical nature of the sensors. In this situation, for instance, the target depth measurement, provided by the depth from focus algorithm, provides reliable information at low frequencies only, whereas the target depth derivative measurements may be corrupted by a bias in the same frequency region (as exemplified in next section), which makes it useful at higher frequencies.

The complementary filter design corresponds to the choice of the parameter k, i.e. to the choice of the cutoff frequency of the low- and high-pass filters, which is entirely dictated by the physical characteristics of the sensors. Therefore, the emphasis, that in Wiener and Kalman filtering is put into describing process and measurement noises [13], is shifted from a statistical framework to a deterministic framework, where the aim is to shape the filter closed-form transfer function. The design of the filter can be done resorting to any efficient method, and the analysis of the filter can be performed in the frequency domain using Bode plots.

In the simple case described, the stochastic underlying process model, here called $\mathcal{M}$, can be written relying on the realization

$$\Sigma_{\mathcal{M}} := \begin{cases} \dot{z} &= \psi_m - \psi_d \\ z_m &= z + z_d \end{cases},$$

where ψ_d and z_d play the roles of process and measurement noises, respectively. In an $\mathcal{H}_2$ setting, the objective is to minimize the estimation error $z - \hat{z}$ for given values of the covariances of ψ_d and z_d. The optimal solution to this problem has the complementary structure described in relation (9). The covariances of ψ_d and z_d are simply viewed as design parameters to vary the cutoff frequency of the filter.

B. Second-order: unknown target dimensions

In most situations, there is no information about the dimensions of the target. Therefore, the value of γ, and as a consequence the value of β, are not known, and the measurements of the target depth derivative (6) are biased. The simple complementary structure described previously does not allow steady-state bias estimation. However, a modified version of its structure, augmented with an extra integrator, will meet this additional constraint. This strategy results in a new complementary filter, depicted in Fig. 1 and described in the remainder of this section, with the realization

$$\Sigma_{\mathcal{M}} := \begin{cases} \begin{bmatrix} \dot{\hat{x}}_1 \\ \dot{\hat{x}}_2 \end{bmatrix} = \begin{bmatrix} -k_1 & 1 \\ -k_2 & 0 \end{bmatrix} \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} + \begin{bmatrix} k_1 \\ k_2 \end{bmatrix} z_m + \begin{bmatrix} 1 \\ 0 \end{bmatrix} \psi_m \\ \hat{z} = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} \end{cases}, \tag{11}$$

where x_1 and x_2 denote the states associated with the depth of the target and with the bias term, respectively, and k_1 and k_2 are filter gains.

From (11), it is simple to show that the estimation of the target depth can be rewritten as in (10), where the transfer functions of $\mathcal{T}_1$ and $\mathcal{T}_2$ take the form $T_1(s) = (k_1 s + k_2)/(s^2 + k_1 s + k_2)$ and $T_2(s) = s^2/(s^2 + k_1 s + k_2)$, respectively, and the intensity of the noise term $\mathcal{F}_z z_d + \mathcal{F}_\psi \psi_d$

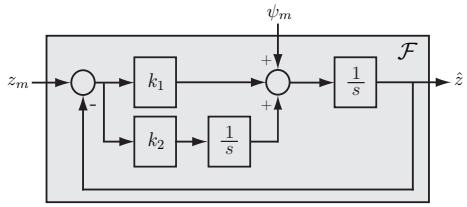


Fig. 1. Complementary filter with bias estimation.

is given by $F_z(s) = T_1(s)$ and $F_\psi(s) = s/(s^2 + k_1 s + k_2)$. As before, $T_1(s) + T_2(s) = I$, where $T_1(s)$ and $T_2(s)$ correspond to low- and high-pass filters, respectively. The second-order complementary filter proposed blends the information provided by the depth from focus algorithm at low frequency regions, with that of the target depth derivative in the complementary frequency range, leaving the original signal z undistorted. Therefore, low frequency bias in the disturbance that corrupts measurements ψ_m will be naturally rejected at the output. Note also that the filter rejects high frequency noise present in measurements z_m.

In this situation, the underlying process model can be written relying on the realization

$$\Sigma_{\mathcal{M}} := \begin{cases} \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} \psi_m - \begin{bmatrix} \psi_d \\ \beta_d \end{bmatrix} \\ z_m = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + z_d \end{cases}, \tag{12}$$

where ψ_d and β_d correspond to the process noise and z_d to the measurement noise. One of the two measurements available (ψ_m) is used as an input to the differential equation that models the process, and the other (z_m) as the observation. If the process and measurement noises are stationary, white, and Gaussian processes with zero mean, then, as stated in Lemma 1, the complementary filter described in this section corresponds to a stationary Kalman filter for the realization presented in (12). Therefore, according to the properties of Kalman filters [13], the proposed complementary filter provides a stable and optimal solution, in the minimum mean square error sense, for the problem of estimating the depth of a target evolving according to the underlying process model presented.

Lemma 1: Let the stationary process and observation noises in realization (12) correspond to stationary white Gaussian noises with zero mean and spectral densities σ_ψ^2, σ_β^2, and σ_z^2, respectively (i.e. $\psi_d \sim \mathcal{N}(0, \sigma_\psi^2)$, $\beta_d \sim \mathcal{N}(0, \sigma_\beta^2)$, and $z_d \sim \mathcal{N}(0, \sigma_z^2)$), and β denote a low frequency bias that corrupts the measurements of the target depth derivative. Then the complementary filter in (11) is the stationary Kalman filter for the system (12) if $k_1 = \sqrt{2\sigma_\beta/\sigma_z + (\sigma_\psi/\sigma_z)^2}$ and $k_2 = \sigma_\beta/\sigma_z$.

Proof: The proof of this lemma is omitted here due to space constraints. ■

In an $\mathcal{H}_2$ setting, the objective is to minimize the state estimation error for given values of the covariances of ψ_d, β_d, and z_d. As seen, the optimal solution to this problem has the complementary structure described in relation (11). The covariances of ψ_d, β_d, and z_d are simply viewed as design parameters to vary the cutoff frequency of the filter.

IV. Experimental Results

In this section, some brief considerations about the implementation of the proposed system and experimental results illustrating the performance of the depth estimation algorithm are presented.

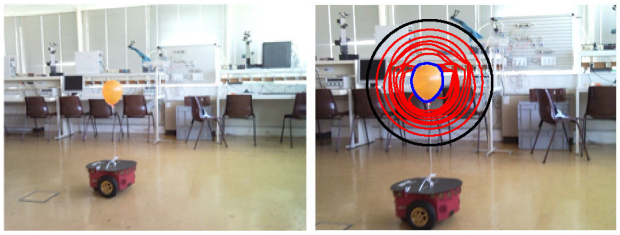
The results in this section were obtained with the 215 PTZ camera from AXIS. Images with the spatial resolution 704×576 pixels were used. For the sake of simplicity, only the red component of acquired images was considered, since the target in the experiments described in this section was red. However, the algorithm proposed in this paper is straightforward adapted to targets with other colours.

As in most cameras, the value of the distance v_0, between the plane of the CCD sensor of the used camera and the lens of the camera, is not accessible to the operator. Instead, a different parameter ranging from 1 to 9999 is available. This parameter is specified by the manufacturer and is usually known as the camera focus setting. The use of the depth estimation algorithm proposed requires the calibration of the relation between these two quantities, see [18] for details about this procedure.

In practice, the strategies described in section II lead to discrete measurements, z_{m_k} and ψ_{m_k}, of the depth of the target and its derivative with respect to time, respectively. The index k represents the time instant kT, $k = k_0, k_0 + 1, \ldots$, where the index k_0 is associated with the initial instant $k_0 T$ and $T > 0$ is the sampling interval. The measurements of the target depth z_{m_k} are obtained directly from the depth from focus algorithm, and the measurements of the target depth derivative ψ_{m_k} are computed according to $\psi_{m_k} = -fR'\dot{r}_{m_k}/r_{m_k}^2$, with $r_{m_k} = \sqrt{\text{tr}(\Sigma_{\mathbf{x}_k})}$ and $\dot{r}_{m_k} = (r_{m_k} - r_{m_{k-1}})/T$, where $R' = R + \gamma$ is the value considered for the target unknown dimensions and $\Sigma_{\mathbf{x}_k}$ is the covariance matrix associated with the boundary of the projection of the target into the image acquired at instant kT. These two measurements were provided to a discrete-time version of the filter with the realization in (11). The discretization of this expression is not detailed in this document due to lack of space, however, details about this procedure, in which a strategy referred by some authors as *emulation* was used, can be found in [19].

In the sequel, two experiments are reported: one in which the target, a balloon attached to a robot *Pioneer P3-DX* as in Fig. 2, moves along a straight line, and other in which the target describes a circumference. In both experiments, the nominal sampling interval T for the application was set to 1.3 s, due to limitations imposed by the resources available, the focal length of the lens was set to its maximum, $f = 45.6$ mm, and the value considered for the target unknown dimensions was $R' = 2.5$ mm.

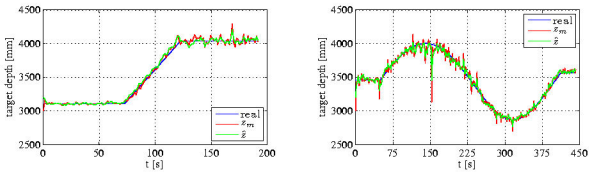
The performance of the depth estimates provided by the discrete-time complementary filter in both experiments is illustrated in Figs. 3, 4, and 5. The target nominal (plot in blue) and estimated (plot in green) depths are depicted in Fig. 3. As can be seen, the estimates provided by the filter converge to the target real depth, i.e. the depth estimation error, depicted in Fig. 4, converges to zero. From the standard deviations σ_{ss} of the steady-state depth estimation errors



(a) (b)

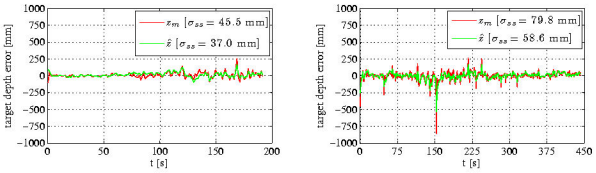
Fig. 2. Real time target tracking. Left: experimental setup; right: target identification, where the initial snake is presented in black, its temporal evolution is presented in red, and the final contour estimate is presented in blue.

presented in this figure, it is possible to confirm that the depth estimates $\hat{z}$ provided by the filter perform better than the measurements z_m obtained directly from the depth from focus strategy. The standard deviations of the steady-state errors associated with the depth estimates provided by the complementary filter (37.0 mm in the straight line trajectory and 58.6 mm in the circular trajectory) are smaller than the ones associated with the depth measurements provided by the depth from focus algorithm (45.5 mm in the straight line trajectory and 79.8 mm in the circular trajectory). There are several reasons that can justify these errors: i) uncertainty associated with the characterization of the real trajectory described by the target; ii) errors resulting from the fitting of the cost function, and iii) uncertainty associated with the calibration of the relation between the focus value and focus setting of the camera.



(a) Straight line trajectory. (b) Circular trajectory.

Fig. 3. Depth estimation ($k_1 = 0.7$ and $k_2 = 0.07$).



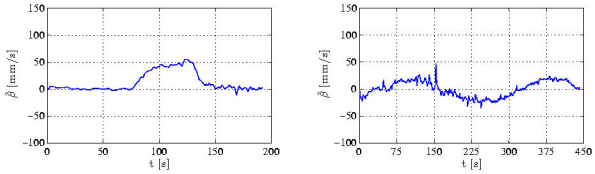
(a) Straight line trajectory. (b) Circular trajectory.

Fig. 4. Depth estimation error ($k_1 = 0.7$ and $k_2 = 0.07$).

The value of the bias estimates $\hat{\beta}$ provided by the second-order complementary filter in both experiments, which result from the unknown nature of the dimensions of the target, is depicted in Fig. 5.

V. Conclusions and Future Work

In this paper, new methodologies for the estimation of the depth of a target with unknown dimensions were proposed. Depth from focus techniques, rooted on optical characteristics of the lens system, namely the point spread function, were used. This work complements an inexpensive single pan and tilt camera-based indoor positioning and tracking system, resorting to a complementary filter for depth estimation. The analysis and synthesis of the filter were proposed, solving the problem at hand. The performance of the system was assessed resorting to a series of indoor experimental tests for a range of operation of up to ten meter. A centimetric accuracy was obtained under realistic conditions. In the near future, the overall system will be used to track and locate small indoor Unmanned Aerial Vehicles, and to generate real time 3D trajectories of marine animals under captivity, for behavioral studies.



(a) Straight line trajectory. (b) Circular trajectory.

Fig. 5. Bias estimation ($k_1 = 0.7$ and $k_2 = 0.07$).

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] M. Linderoth, A. Robertsson, K. Åström, and R. Johansson, "Object tracking with measurements from single or multiple cameras," in *Proceedings of the IEEE International Conference on Robotics and Automation*, May 2010, pp. 4525–4530.

[4] V. Lepetit and P. Fua, "Monocular model-based 3d tracking of rigid objects," *Foundations and Trends in Computer Graphics and Vision*, vol. 1, no. 1, pp. 1–89, 2005.

[5] H. Q. H. Viet, M. Miwa, H. Maruta, and M. Sato, "Recognition of motion in depth by a fixed camera," in *VII Digital Image Computing: Techniques and Applications*, Dec 2003, pp. 205–214.

[6] A. P. Pentland, "A new sense for depth of field," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 9, no. 4, pp. 523–531, Jul 1987.

[7] N. Asada, M. Baba, and A. Oda, "Depth from blur by zooming," in *Proceedings of the Vision Interface Annual Conference*, May 2001, pp. 165–172.

[8] J. Ens and P. Lawrence, "An investigation of methods for determining depth from focus," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 2, pp. 97–108, 1993.

[9] E. Krotkov, "Focusing," *International Journal of Computer Vision*, vol. 1, pp. 223–237, Oct 1987.

[10] S. Nayar and Y. Nakagawa, "Shape from focus," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 8, pp. 824–831, 1994.

[11] J. Vasconcelos, C. Silvestre, P. Oliveira, P. Batista, and B. Cardeira, "Discrete time-varying attitude complementary filter," in *Proceedings of the American Control Conference*, 2009, pp. 4056–4061.

[12] S. Merhav, *Aerospace Sensor Systems and Applications*. Springer-Verlag, 1996.

[13] R. Brown and P. Hwang, *Introduction to Random Signals and Applied Kalman Filtering*. John Wiley & Sons, 1997.

[14] T. Gaspar and P. Oliveira, "Single pan and tilt camera indoor positioning and tracking system," *European Journal of Control*, in press, 2011, Preprint available in http://users.isr.ist.utl.pt/~tgaspar/temp/EJC2010pp.pdf.

[15] E. Hecht, *Optics*, 4th ed. Addison-Wesley, 2001.

[16] A. Blake and M. Isard, *Active Contours*, 1st ed. Springer, 2000.

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

[18] K. Tarabanis, R. Tsai, and D. Goodman, "Modeling of a computer-controlled zoom lens," in *Proceedings of the IEEE International Conference on Robotics and Automation*, vol. 2, May 1992, pp. 1545–1551.

[19] W. Rugh, *Linear System Theory*, 2nd ed. Prentice Hall, 1996.