NEW DYNAMIC ESTIMATION OF DEPTH FROM FOCUS IN ACTIVE VISION SYSTEMS
Data Acquisition, LPV Observer Design, Analysis and Test

Tiago Gaspar and Paulo Oliveira

Institute for Systems and Robotics, Instituto Superior Técnico, Av. Rovisco Pais, 1049-001 Lisboa, Portugal

Keywords: Depth estimation, Depth from focus, LPV observers.

Abstract: In this paper, new methodologies for the estimation of the depth of a generic moving target with unknown dimensions, based upon depth from focus strategies, are proposed. A set of measurements, extracted from real time images acquired with a single pan and tilt camera, is used. These measurements are obtained resorting to the minimization of a new functional, deeply rooted on optical characteristics of the lens system, and combined with additional information extracted from images to provide estimates for the depth of the target. This integration is performed by a Linear Parameter Varying (LPV) observer, whose synthesis and analysis are also detailed. To assess the performance of the proposed system, a series of indoor experimental tests, with a real target mounted on a robotic platform, for a range of operation of up to ten meter, were carried out. A centimetric accuracy was obtained under realistic conditions.

1 INTRODUCTION

Depth estimation plays a key role in a wide variety of domains, such as target tracking (Bar-Shalom et al., 2001), 3D reconstruction (Bertelli et al., 2008), obstacle detection (Discant et al., 2007), and video surveillance (Haritaoglu et al., 2000). In 3D image applications, a common approach consists in using triangulation methods applied to the data collected by two or more cameras. However, there has been work on estimating depth resorting to a single camera, see (Krotkov, 1987) and (Ens and Lawrence, 1993). In addition to the main advantage of requiring just one camera, this technique reduces the impact of the image to image matching problem, as well as the impact of occlusion problems, see (Schechner and Kiryati, 1998). The idea is to explore the relation between the depth of a point in the 3D world and the amount of blur that affects its projection 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, see (Viet et al., 2003) and (Pentland, 1987); depth from blur by zooming, see (Asada et al., 2001); and depth from blur by irising, see (Ens and Lawrence, 1993).

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, see (Pentland, 1987) and (Ens and Lawrence, 1993), and depth from focus, see (Krotkov, 1987), (Nayar and Nakagawa, 1994), and (Viet et al., 2003). This work is based on the 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 (PSF) responsible for the blurring does not need to be modeled. This is not possible in depth from defocus strategies, where it is common to consider that this function is either a two-dimensional Gaussian, or a circle of constant intensity. Moreover, the amount of blur present in an image is a consequence of both the characteristics of the lens and the scene itself, which restricts the applicability of depth from defocus methods to step discontinuities in the scene. There are strategies that tackle this problem by using a minimum of two images of the same scene, acquired with a different depth of field (Pentland, 1987). Since the contribution of the scene to all images is the same, it can be removed. However, measuring the amount of blur with high precision is still a difficult problem, as it is an ill-posed inverse problem.

In this paper, two novelties are proposed: a new algorithm for the estimation of the depth of a target with unknown dimensions is developed, and a strat-

egy to estimate these dimensions is also described. The depth estimation problem is tackled by combining information present on the target boundary, namely the amount of blur that corrupts this region, with measurements of the dimensions of the projection of the target into acquired images. The dynamics of the depth of the target is written as a function of a parameter that depends on the dimensions of the image of the target, which leads to a LPV observer for the depth of moving targets with unknown dimensions. In what concerns the dimensions of the real target, they are estimated resorting to the depth estimates provided by the observer and to the measurements of the dimensions of the image of the target.

This document is organized as follows: in section 2, some background on theory of defocus is provided, and in section 3, a new method to estimate the camera focus value that minimizes the amount of blur in an image discontinuity is presented. In section 4, the design and analysis of the proposed LPV observer are detailed, and in section 5, experimental results illustrating the performance of the described depth estimation algorithms are provided. Finally, section 6 summarizes the main conclusions of this work and unveils challenging problems for the future.

2 BACKGROUND ON THEORY OF DEFOCUS

There are two traditional approaches to model the image formation process: one uses geometrical optics and the other physical optics. The first is an approximation that disregards behaviours specifically attributed to the wave nature of light, such as interference and diffraction, and relies on ray tracing. The great simplicity of this approach compensates for its inaccuracies. On the contrary, the second relies on diffraction theory, and its results are exact. In this work, only geometrical effects are considered since the spatial resolution of the used imaging system makes diffraction effects negligible.

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. The depth of field of a camera with a given focus value corresponds to the distance between the farthest and the nearest planes on the scene, in relation to the camera, whose points appear in acquired images with a satisfactory definition, according to a certain criterion.

At each instant, a lens can exactly focus points in only one plane, denominated object plane. Considering a thin model for the lens of the camera, see (Hecht,

2001), 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, see Fig. 1. To complete the
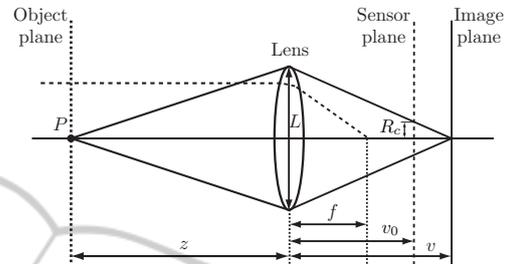

Figure 1: Geometrical optics model for the imaging process of a thin lens.

relation, the focal length f of the lens must be considered. This relation is known as the Gaussian Lens Formula, see (Hecht, 2001), and can be rearranged in the form

$$z = \frac{fv}{v - f}. \qquad (1)$$

Considering that the CCD sensor plane is located at a distance $v_0 < v$ from the lens, and using (1) and some trigonometric manipulations, it is possible to write the distance z from the lens to the object plane in the scene as

$$z = \frac{fv_0}{2R_cF + v_0 - f}, \qquad (2)$$

see (Ens and Lawrence, 1993) and (Pentland, 1987) for details, where F is the f-number of the lens and R_c is the effective radius of the point spread function. This expression is valid when $v > v_0$. An expression similar to this would be easily derived for the case $v < v_0$.

In practical applications, usually all parameters in the right-hand side of equation (2) are known, except for R_c. Depth from focus methods consist in finding the sensor plane position that minimizes the amount of blur present in image points of interest. This corresponds to finding the camera focus value that leads to $R_c = 0$, which is solved by optimizing a cost function that depends on the amount of blur present in the points of interest. Depth can then be computed using (1).

3 MINIMUM BLUR FOCUS VALUE

In this section, a method to estimate the camera focus

value that leads to the minimum amount of blur in an image discontinuity is proposed. The cost function used for this purpose is described, as well as the procedure used to search for its minimum.

3.1 Cost Function

The estimation of the camera focus value that minimizes the amount of blur in an image discontinuity 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, such as the Fourier transform, the image gradient, or the Laplacian, are detailed in (Krotkov, 1987).

There are some properties that are desirable for a cost function: it must preferably be unimodal, vary monotonically with the focus value on either side of the mode, and be robust in presence of noise. In (Krotkov, 1987), several cost functions were tested and the maximization of the magnitude of the image intensity gradient proved to achieve better results in what concerns the referred criteria.

The goal of our system is to estimate the depth of a target, therefore the metric proposed aims to maximize the image gradient magnitude across lines orthogonal to its boundary, which can be found resorting to active contours, see (Blake and Isard, 2000) 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 CCD sensor of the camera) 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

$$g(v_0) = \frac{1}{\frac{1}{N_l} \sum_{i=1}^{N_l} \max_{(x,y) \in l_i} ||\nabla I_{v_0}(x,y)||^2} \quad (3)$$

is the inverse of the mean 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.

3.2 Optimization of the Cost Function

The minimization of the cost function proposed in (3) is difficult. The data available is scarce and to get new information, the acquisition of new images is required. The problem is even more difficult as we want to estimate parameters related with the depth of a moving target. Therefore, a model for the cost function that allows to infer its minimum resorting only to a few images will be derived.

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 $I_{v_0}^f(x,y)$ with the point spread function $h(x,y)$ of the lens system, i.e. with the function that models the camera blurring process:

$$I_{v_0}(x,y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I_{v_0}^f(\alpha, \beta) h(x-\alpha, y-\beta) d\alpha d\beta.$$

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

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

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 $I_{v_0}^f(x,y) = a_1 + a_2 u(x-x_0)$, 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 $I_{v_0}^f(x,y)$ 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

$$\begin{cases} 0 & , \quad |x-x_0| > R_c \\ \frac{2a_2}{\pi R_c^2} \sqrt{R_c^2 - (x-x_0)^2} & , \quad |x-x_0| \leq R_c \end{cases}.$$

By considering a line l orthogonal to the boundary of the target, it is possible to conclude that

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

From Fig. 1, 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 (Ens and Lawrence, 1993) for details. The replacement of the value of R_c in $\left(\frac{2a_2}{\pi R_c}\right)^2$ by its expression, allows to write the cost function proposed in (3) in the form

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

According to the discussion above, which is confirmed by Fig. 2, the cost function in (3) is expected to depend parabolically on v_0. Therefore, the parabolic model $g(v_0) = a(v_0 - v)^2 + b$, where a, b, and v are parameters to be estimated, was considered for the cost function. In particular, v is the camera focus value that minimizes the cost function. This expression can also be written as $g(v_0) = a'v_0^2 + b'v_0 + c'$, where $a' = a$, $b' = -2av$, and $c' = av^2 + b$. In this form, the model of the cost function depends linearly on the parameters that must be estimated, which simplifies significantly the fitting process described next. This is the reason why the minimization of $g(v_0)$ was considered, instead of the maximization of its inverse, which seemed more intuitive.
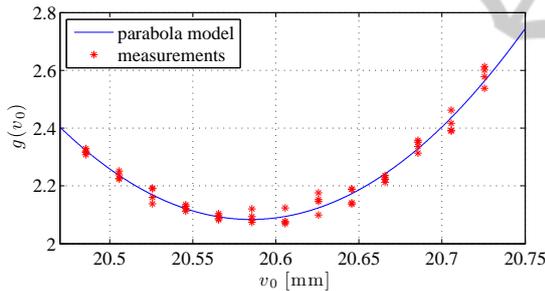

Figure 2: Cost function for an AXIS 215 PTZ, when the camera focal length is 29 mm and the target is 3 m away from the lens.

Consider that at instant k the focus value of the camera is v_{0_k}, and that a measurement of the cost function $g(v_{0_k})$ corrupted by additive white Gaussian noise is available. Stacking N of these measurements, a fitting problem can be formulated as $\min_{\mathbf{y}} ||\mathbf{Ay} - \mathbf{b}||$, where $|| \cdot ||$ is the Euclidean norm, $\mathbf{A}$ is a matrix with N rows of the form $[v_{0_k}^2 \ v_{0_k} \ 1]$, $\mathbf{b}$ is a column vector that stacks the N measurements, and $\mathbf{y} = [a' \ b' \ c']^T$ is the vector of parameters to be estimated. The solution to this problem is straightforward, using least squares method. Given the three unknowns of the model, each minimization of the cost function requires the acquisition of at least three images with different focus values. This procedure must be repeated over time since the cost function varies with the instantaneous depth of the target. Once estimated the three parameters, from which the camera focus value v that minimizes

the cost function is easily obtainable, the depth z of the target can be computed resorting to (1).

The depth estimation method proposed in this section is robust to variations in parameters such as scene illumination or camera zoom and aperture values, which may change the shape of the cost function, see Fig.3, since the implemented estimation process estimates new parabola coefficients in each iteration of the algorithm, leading to the adaptation of the cost function model to those values.
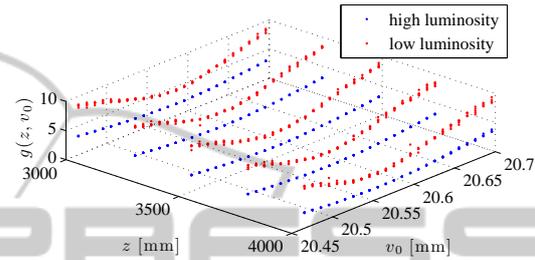

Figure 3: Luminosity influence on the cost function, for several target depths (results obtained with an AXIS 215 PTZ; $f = 45.6$ mm).

4 DEPTH LPV OBSERVER

In this section, an observer for the depth of a target with unknown dimensions is pursued. A state-space formulation for the evolution of the target depth is derived in continuous time, and an observer for the state of the LPV system that results is proposed. The analysis of the observer stability and its discrete-time version are also provided.

4.1 Continuous-time Observer

Considering a pinhole model for the camera, see (Faugeras and Luong, 2001), the relation between the cartesian coordinates of a point in the camera reference frame (x, y, z) and the coordinates (x_p, y_p) of its projection into the image plane is given by $x_p = fx/z$ and $y_p = fy/z$, 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 expressions of x_p and y_p, 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

$$r = fR/z. \tag{4}$$

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 $\mathbf{x} \in \mathcal{R}^2$, 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 $\mathbf{x}_a \in \mathcal{R}^2$ 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 $\Sigma_{\mathbf{x}_b}$, and that $\text{tr}(\cdot)$ denotes the trace of a matrix. 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 $\mathbf{R}_{\mathbf{x}}^T\mathbf{R}_{\mathbf{x}} = \mathbf{I}_{2\times2}$, where $\mathbf{I}_{2\times2}$ 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 relation (4), and assuming that the focal length f of the lens and the dimensions R of the real target do not vary over time, it is possible to write the derivative of the depth of the target with respect to time in the form

$$\dot{z} = -\frac{\dot{r}}{r}z, \qquad (5)$$

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 projection of the target into acquired images, and their measurements are here denoted r_m and $\dot{r}_m$. Assuming that z_m, r_m, and $\dot{r}_m$ are exact measurements of z, r, and $\dot{r}$, and denoting the quotient $-\dot{r}_m/r_m$ by a parameter α, a deterministic LPV system with the realization

$$\begin{cases} \dot{z} &= \alpha z \\ z_m &= z \end{cases}$$

results. An observer for the state z of this system can be written in the form

$$\dot{\hat{z}} = \alpha\hat{z} + h(z_m - \hat{z}), \quad \hat{z}(t_0) = \hat{z}_0, \qquad (6)$$

see (Rugh, 1996), where $\hat{z}$ and $\dot{\hat{z}}$ are the target depth estimate and its derivative with respect to time, respectively, h is the observer gain, t_0 is the initial time instant, and $\hat{z}_0$ is the initial estimate for the target depth.

From the considerations above, it is straightforward to show that the state estimation error $\tilde{z} = z - \hat{z}$ satisfies the linear state equation

$$\dot{\tilde{z}} = (\alpha - h)\tilde{z}, \quad \tilde{z}(t_0) = z_0 - \hat{z}_0, \qquad (7)$$

where $\dot{\tilde{z}}$ denotes the derivative of the estimation error with respect to time. The values of r_m and $\dot{r}_m$, and as a consequence the value of α, depend on several variables, such as the target dimensions, the target depth, and the target motion. Therefore, the gain of the observer must be chosen according to the experiment at hand to guarantee the stability of the observer, as shown in Proposition 1.

Proposition 1. *The linear state equation* (7) *is uniformly exponentially stable if the gain h of the observer verifies* $h \geq \alpha_{max} + \nu/2q$, *where* α_{max} *is the upper bound of* α, *and* ν *and* q *are finite positive constants.*

Proof. Consider the Lyapunov function $V(\tilde{z}) = q\tilde{z}^2$, where q is a finite positive constant. From the error dynamics in (7), it is possible to show that the derivative of this Lyapunov function with respect to time has the form $\dot{V}(\tilde{z}) = 2q(\alpha - h)\tilde{z}^2$. According to Lyapunov theory, see (Rugh, 1996), the linear state equation (7) is uniformly exponentially stable if there exists a q that, for all possible values of α, verifies $2q(\alpha - h) \leq -\nu$, where ν is a finite positive constant. This relation can be rewritten in the form $h \geq \alpha + \nu/2q$. Consider that α has an upper bound α_{max}, which is specified by the values that both r_m and $\dot{r}_m$ can assume. If the gain of the observer is chosen in such a way that $h \geq \alpha_{max} + \nu/2q$ is verified, for given values of ν and q, then the observer-error state equation (7) is guaranteed to be uniformly exponentially stable. $\square$

4.2 Discrete-time Observer

According to relation (6), the depth estimates provided by the proposed observer can be rewritten in the form

$$\dot{\hat{z}}(t) = \underbrace{(\alpha(t) - h)}_{a(t)}\hat{z}(t) + hz_m(t), \qquad (8)$$

where a is a new parameter. The time variable t, omitted in previous sections, was considered to distinguish the terms that depend on time from the ones that do not.

The solution of the homogeneous equation $\dot{\hat{z}}(t) = a(t)\hat{z}(t)$ is given by

$$\hat{z}(t) = \underbrace{e^{\int_\tau^t a(\sigma)d\sigma}}_{\Phi(t,\tau)}\hat{z}(\tau),$$

where τ is an arbitrary instant of time verifying $t \geq \tau$. Therefore, the solution of (8) has the form

$$\hat{z}(t) = \Phi(t,\tau)\hat{z}(\tau) + \int_\tau^t \Phi(t,\sigma)hz_m(\sigma)d\sigma, \quad t \geq \tau,$$

see (Rugh, 1996) for more details. Evaluating this expression for $t = (k+1)T$ and $\tau = kT$, where T is a fixed positive constant and $k = k_0, k_0 + 1, \ldots$, yields

$$\hat{z}_{k+1} = F_k\hat{z}_k + \Lambda_k u_k,$$

where u_k and $\hat{z}_k$ denote the input z_m and the state estimate $\hat{z}$, respectively, at instant kT. The values of the input z_m and parameter a were assumed constant over the integration range, and the index k_0 is associated with the initial time instant $k_0 T$. According to the considerations above, we have $F_k = e^{a_k T}$ and $\Lambda_k = h(e^{a_k T} - 1)/a_k$, where $a_k = -h - \dot{r}_{m_k}/r_{m_k}$. The variables associated with the subscript k are discrete, with values that correspond to the evaluation of their continuous-time versions at time instant kT.

The discrete-time LPV observer derived in this section provides estimates for the depth of a target, with unknown dimensions, moving in a 3D scene. Therefore, this observer is suitable for the tracking system proposed, since it is appropriate for implementation in a digital computer.

5 EXPERIMENTAL RESULTS

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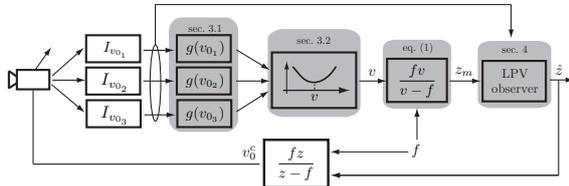

Figure 4: Scheme of the proposed depth estimation algorithms.

Figure 4 depicts a simplified version of the architecture of the proposed depth estimation strategies. In this figure, $I_{v_{0_i}}$ and $g(v_{0_i})$, $i = 1, 2, 3$, denote, respectively, the three images used by the depth from focus algorithm and the cost function measurements extracted from these images. The value of v_0^c corresponds to the focus value used to command the focus of the camera.

Results provided in this section were obtained with the 215 PTZ camera from AXIS. Images with the spatial resolution 704×576 pixels were used. Since image segmentation is itself a very complex domain, which does not correspond to the main focus of this work, targets with easily identifiable colours were considered.

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 algorithms proposed requires the calibration of the relation between these two quantities, see (Tarabanis et al., 1992) for details about this procedure.

The implementation of the proposed discrete-time observer requires the availability of discrete-time versions of the measurements extracted from images. The value of the target depth at time instant kT, $k = k_0, k_0 + 1, \ldots$, obtained from the depth from focus algorithm, is denoted z_{m_k}. The dimensions r_{m_k} of the projection of the target into the image acquired at instant kT, and its derivatie over time $\dot{r}_{m_k}$, are computed according to $\sqrt{\text{tr}(\Sigma_{\mathbf{x}_k})}$ and $(r_{m_k} - r_{m_{k-1}})/T$, respectively, where $\Sigma_{\mathbf{x}_k}$ denotes the covariance matrix associated with the boundary of the projection of the target into the image acquired at instant kT. As stated before, this boundary is estimated resorting to active contours.

In the sequel, two experiments are reported: one in which the target, a balloon attached to a robot *Pioneer P3-DX* as in Fig. 5, 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, and the focal length of the lens was set to its maximum, $f = 45.6$ mm.
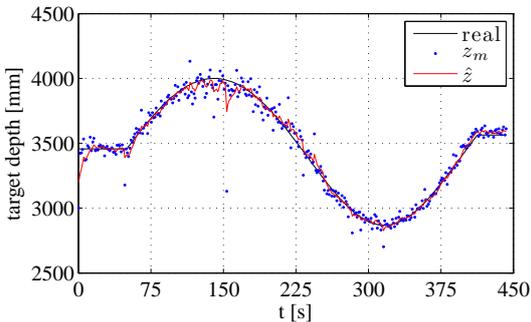
The performance of the depth estimates provided by the depth from focus algorithm and discrete-time observer in both experiments is illustrated in Figs. 6 and 7. In Fig. 6, the nominal depth of the target is plotted in blue and the value of the depth estimates provided by the depth form focus algorithm and LPV observer are plotted in red and green, respectively. As can be seen, both estimates converge to the target real depth, i.e. the depth estimation error, depicted in Fig. 7, converges to zero. From the standard de-

(a)

(b)

Figure 5: Real time target tracking. Left: experimental setup; right: target identification, where the initial guess for the target contour is presented in black, its temporal evolution is presented in red, and the final contour estimate is presented in blue.
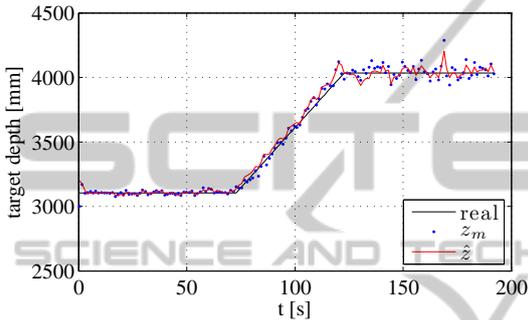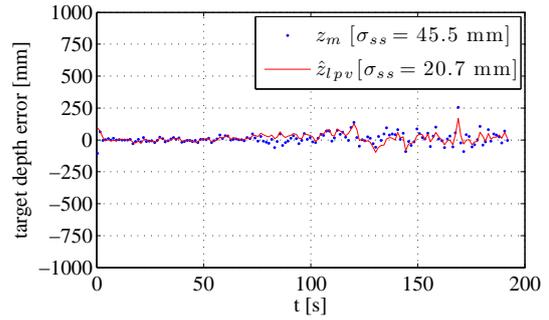

(a) Straight line trajectory.
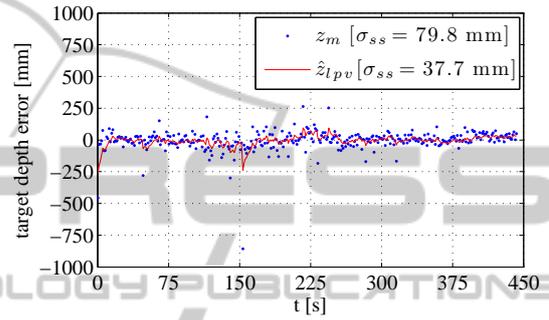

(b) Circular trajectory.

Figure 6: Depth estimation ($h = 0.4$).

viations σ_{ss} of the steady-state depth estimation errors presented in this figure, it is possible to confirm that the depth estimates $\hat{z}$ provided by the observer 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 LPV observer (20.7 mm in the straight line trajectory and 37.7 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 explain the er-

(a) Straight line trajectory.

(b) Circular trajectory.

Figure 7: Depth estimation error ($h = 0.4$).

rors observed in Fig. 7: 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 the focus setting of the camera. In particular, part of the errors associated with the estimates provided by the observer result from the fact that the measurements z_m, r_m, and $\dot{r}_m$ are not exact, as assumed in the derivation of the observer, but corrupted by noise.

For the experiments reported in this section, the dimensions of the target do not vary over time. Therefore, it is possible to infer from (4) that an estimate $\hat{R}$ of the target dimensions can be obtained according to $\hat{R} = E\left\{(r_m\hat{z})/f\right\}$, where E denotes the expected value operator. The quantities r and z, in (4), were replaced by the measurements r_m and by estimates $\hat{z}$ of the depth of the target, respectively, since their real values are not known. In discrete-time, this expression can be rewritten in the form

$$\hat{R} = \frac{1}{N_T} \sum_{k=1}^{N_T} \frac{r_{m_k}\hat{z}_k}{f},$$

where N_T denotes the number of iterations of the expriment, and r_{m_k} and $\hat{z}_k$ the values of r_m and $\hat{z}$, respectively, at time instant kT. The use of this strategy to estimate the dimensions of the target used in the

two experiments leads to $\hat{R} = 35.80$ mm, when the depth estimates provided by the depth from focus algorithm are used, and to $\hat{R} = 35.77$ mm, when the depth estimates are provided by the observer. Theses values are very close to the target real dimensions, $R = 35.76$ mm.

6 CONCLUSIONS

In this paper, new methodologies for the estimation of the depth of a moving target with unknown dimensions were proposed. Measurements of the target depth, extracted from real time images acquired with a single pan and tilt camera and based upon depth from focus techniques, were used. These measurements were processed resorting to a LPV observer, whose analysis and synthesis were also provided. The performance of the proposed algorithms 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, this system will be used to generate real time 3D trajectories of marine animals under captivity, for behavioural studies.

ACKNOWLEDGEMENTS

REFERENCES

Asada, N., Baba, M., and Oda, A. (2001). Depth from blur by zooming. In *Proceedings of the Vision Interface Annual Conference*, pages 165–172.

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.

Bertelli, L., Ghosh, P., Manjunath, B., and Gibou, F. (2008). Robust depth estimation for efficient 3d face reconstruction. *15th IEEE International Conference on Image Processing*, pages 1516–1519.

Blake, A. and Isard, M. (2000). *Active Contours*. Springer, 1st edition.

Discant, A., Rogozan, A., Rusu, C., and Bensrhair, A. (2007). Sensors for obstacle detection - a survey. *30th International Spring Seminar on Electronics Technology*, pages 100–105.

Ens, J. and Lawrence, P. (1993). An investigation of methods for determining depth from focus. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(2):97–108.

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

Haritaoglu, I., Harwood, D., and Davis, L. (2000). W^4: real-time surveillance of people and their activities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):809–830.

Hecht, E. (2001). *Optics*. Addison-Wesley, 4th edition.

Krotkov, E. (1987). Focusing. *International Journal of Computer Vision*, 1:223–237.

Nayar, S. and Nakagawa, Y. (1994). Shape from focus. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(8):824–831.

Pentland, A. P. (1987). A new sense for depth of field. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(4):523–531.

Rugh, W. (1996). *Linear System Theory*. Prentice Hall, 2nd edition.

Schechner, Y. Y. and Kiryati, N. (1998). Depth from defocus vs. stereo: How different really are they? In *Proceedings of the International Conference on Pattern Recognition*, pages 1784–1786.

Tarabanis, K., Tsai, R., and Goodman, D. (1992). Modeling of a computer-controlled zoom lens. In *Proceedings of the IEEE International Conference on Robotics and Automation*, volume 2, pages 1545–1551.

Viet, H. Q. H., Miwa, M., Maruta, H., and Sato, M. (2003). Recognition of motion in depth by a fixed camera. In *VII Digital Image Computing: Techniques and Applications*, pages 205–214.