

Position based Structure from Motion using a Moving Calibrated Camera

Nitendra Nath*, David Braganza, and Darren M. Dawson

Abstract: In this paper, a 3D Euclidean position estimator using a single moving calibrated camera whose position is known is developed that asymptotically recovers the structure of a static object. To estimate the unknown structure an adaptive least squares estimation strategy is employed based on a novel prediction error formulation and a Lyapunov stability analysis. Numerical simulation results are presented to illustrate the effectiveness of the proposed algorithm.

I. INTRODUCTION

Three-dimensional (3D) reconstruction of an object, where the Euclidean coordinates of feature points on a moving or fixed object are recovered from a sequence of two-dimensional (2D) images, known as ‘Structure from Motion (SFM)’ [1], [2], [3], or ‘Simultaneous Localization and Mapping (SLAM)’ [4], [5], has been a mainstream research problem for a long time. The recovery of 3D structure from its 2D projection is usually done by mounting a camera on a moving vehicle such as unmanned aerial vehicle (UAV) or a mobile robot that travels through the environment and takes images of static objects or features. SFM and SLAM have significant impact for several applications including autonomous vehicle navigation, aerial tracking, path planning, surveillance of ground based, stationary or moving objects [6], [7], [8], [9], and terrain mapping systems [10], [11], [12].

Although, the problem of Euclidean reconstruction is inherently nonlinear, linearization based techniques, such as the extended Kalman filter [3], [4], [5], have been used quite frequently. The linearized motion models can cause significant inconsistencies in solutions, as noted in [13]. Some of the past research focused on utilizing nonlinear system analysis and estimation tools to develop nonlinear state observers for the problem [14], [15], [16]. In recent work, Dahl et al. [17] designed a structure estimator for the system state and motion, based on a parameterization of the nonlinear perspective dynamic system. Dupree et al. [18] presented an algorithm for reconstruction of an object’s Euclidean coordinates even if the tracked feature points leave the camera’s field of view, but his work required that at least one geometric length of the object be known *a priori*. The work presented by Laganieri et al. [19] utilized a neural network based approach to recover the structure of a moving

scene from an image sequence where the maximal rigidity principle is applied, allowing a rigid transformation in the structure from one time instant to another. Chitrakaran et al. [20], [21] proposed a nonlinear estimation technique, using a monocular camera, for identification of the Euclidean structure of an object. The homography between two different views, the current frame and a constant reference frame, was utilized to model the camera motion. This algorithm required at least one distance between two feature points on the object to be known for the recovery of 3D Euclidean coordinates. Furthermore, the normal vector to the object must be known [20] and the rotation between the object frame and the camera at the reference position must be known [21]. In recent work [22], where the velocity of the moving camera was assumed to be known for the reconstruction of Euclidean coordinates of a static object, the requirements of [20], [21], that information about the scene be known *a priori* were eliminated. The work in [22] utilized a nonlinear integral observer to estimate the velocity of each feature point in the image plane and then an estimator for the unknown depth variable was developed which facilitated the estimation of 3D Euclidean coordinates of the feature. In [23], De Luca et al. developed an on-line depth estimator based on nonlinear observer theory by using a state space realization of the perspective camera model and utilizing measurements of the camera velocity.

In this work, our objective is to estimate the 3D Euclidean structure of a static object using a single calibrated camera mounted on a moving platform whose position is measurable. In the past, several researchers have addressed the structure identification with constrained camera motion problem, for example [2], [24], [25]. Chaumette et al. [2] proposed a method where the presence of noise and discrete sampling, forced constraints on the motion of the camera. Larry et al. [24] used Kalman filter based algorithms but again various constraints were imposed on the camera motion to make the problem simpler. For instance the camera motion was assumed to have a constant depth, thus making the problem simpler. In similar work, Smith et al. [25] allowed only lateral motion to the camera.

In this paper, we present a method to estimate the Euclidean structure of feature points on a static object using a moving calibrated camera whose position is measurable. The estimator is designed by first developing a geometric model to relate the fixed feature points on the object with the moving camera. The novelty of this work lies in the parameterization of a nonlinear static model which relates the projected pixel coordinates with the Euclidean coordinates of the feature points. A prediction error formulation is

This work is supported in part by a DOE Contract, and a Honda Corporation Grant.

* To whom all correspondence should be addressed.

Nitendra Nath and Darren M. Dawson are with the Department of Electrical & Computer Engineering, Clemson University, Clemson, SC 29634-0915 (e-mail: nnath@clemson.edu; darren.dawson@ces.clemson.edu). David Braganza is with OFS, 50 Hall Road, Sturbridge, MA 01566 (email: dbraganza@ofsoptics.com).

then presented which allows us to utilize well established nonlinear estimation theory to design an adaptive least squares estimator. We show that the developed estimator asymptotically identifies the Euclidean coordinates of the feature points subject to a persistency of excitation condition similar to that of [23]. The real-time estimation technique that is presented has shown good robustness to noise and also provides fast convergence which is demonstrated by our simulation.

The rest of this paper is organized as follows, in Section II the geometric model which relates Euclidean coordinates of visual features on the static object with their corresponding projected pixel coordinates is developed based on the perspective projection. Section III describes the Euclidean structure estimation with the prediction error formulation and the stability analysis. Numerical simulation, presented in Section IV, demonstrates the performance of the proposed estimator.

II. GEOMETRIC MODEL

To develop a geometrical relationship between a moving camera and a stationary object, an orthogonal coordinate frame, denoted by \mathcal{C} , whose origin coincides with the optical center of the moving camera, an inertial coordinate frame, denoted by \mathcal{W} , and an orthogonal coordinate frame, denoted by \mathcal{B} are considered (see Figure 1). To make the following discussion more tractable, n feature points located on a static object, denoted by $\mathcal{F}_i \forall i = 1, \dots, n$ are considered. Let the 3D coordinates of the i^{th} feature point on the object be denoted as the constant $x_{fi} \in \mathbb{R}^3$ relative to the base frame \mathcal{W} , and $\bar{m}_i(t) \in \mathbb{R}^3$ relative to \mathcal{C} , which is defined as follows

$$\bar{m}_i \triangleq [x_i \quad y_i \quad z_i]^T. \quad (1)$$

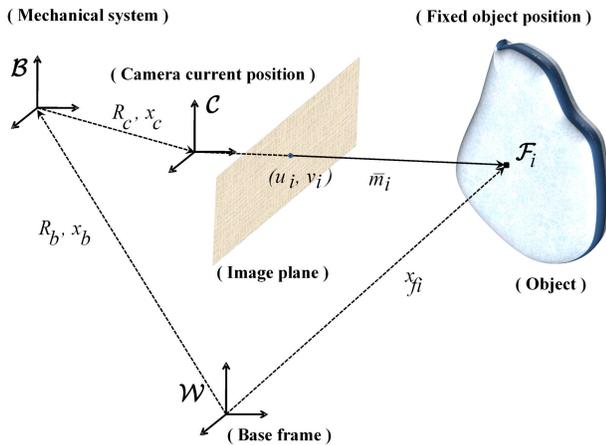


Fig. 1. Geometric relationships between the fixed object, mechanical system, and the camera.

In the subsequent development, it is assumed that the object is always in the field of view of the camera, hence the distances from the origin of \mathcal{C} to all feature points are always positive and bounded. To relate the coordinate systems,

let $R_b(t) \in SO(3)$ and $x_b(t) \in \mathbb{R}^3$ denote the rotation matrix and the translation vector respectively, from \mathcal{B} to \mathcal{W} , expressed in \mathcal{W} . Let $R_c(t) \in SO(3)$ and $x_c(t) \in \mathbb{R}^3$ denote the rotation matrix and the translation vector respectively, from \mathcal{C} to \mathcal{B} , expressed in \mathcal{B} . Let $m_i(t) \in \mathbb{R}^3$ denote the normalized Euclidean coordinates for the i^{th} feature point, which is defined as follows

$$m_i \triangleq \frac{1}{z_i} \bar{m}_i = [x_i/z_i \quad y_i/z_i \quad 1]^T. \quad (2)$$

In the image captured by the camera, each of these feature points has corresponding projected pixel coordinates, denoted by $p_i(t) \in \mathbb{R}^2$, defined as follows

$$p_i \triangleq [u_i \quad v_i]^T \quad (3)$$

where $u_i(t), v_i(t) \in \mathbb{R}$. The projected pixel coordinates of the feature points are related to the normalized Euclidean coordinates by the pin-hole model [26] such that

$$p_i = A m_i \quad (4)$$

where $A \in \mathbb{R}^{2 \times 3}$ is a known constant intrinsic camera calibration matrix defined as follows [27]

$$A \triangleq \begin{bmatrix} f k_u & f k_u \cot \phi & u_0 \\ 0 & \frac{f k_v}{\sin \phi} & v_0 \end{bmatrix} \quad (5)$$

where $k_u, k_v \in \mathbb{R}$ denote camera scaling factors, $u_0, v_0 \in \mathbb{R}$ represent the pixel coordinates of the principal point, $\phi \in \mathbb{R}$ is the angle between the camera axes, and $f \in \mathbb{R}$ is the camera focal length. From (2) and (4), $p_i(t)$ can be written as follows

$$p_i = \frac{1}{z_i} A \bar{m}_i. \quad (6)$$

The Euclidean coordinates of the i^{th} feature point $\bar{m}_i(t)$ relative to the camera along with the corresponding depth $z_i(t)$ are unknown and unmeasurable signals. The corresponding projected pixel coordinates $p_i(t)$ along with $R_b(t)$, and $x_b(t)$ are measurable signals and R_c , and x_c are known constant parameters. The objective of this work is to accurately identify the unknown constant Euclidean coordinates of the feature x_{fi} relative to the world frame in order to recover the 3D structure of the object.

III. EUCLIDEAN STRUCTURE ESTIMATION

In this section, a prediction error formulation for the unknown parameters will be used to parameterize the interaction matrix and the unknown depth variable. A stability analysis will be shown that ensures the estimation error signals go to zero.

A. Prediction Error Formulation

To facilitate the error formulation, the geometric model shown in Figure 1 is considered, where the vector representing the distance between the origins of \mathcal{C} and \mathcal{W} , expressed in \mathcal{W} , be denoted as $x_{cw}(t) \in \mathbb{R}^3$. From Figure 1, it is easy to see that

$$x_{cw} = R_b x_c + x_b. \quad (7)$$

Similarly, from the triangle formed by \mathcal{F}_i and the origins of \mathcal{C} and \mathcal{W} , the following expression can be obtained

$$x_{fi} = x_{cw} + R_b R_c \bar{m}_i. \quad (8)$$

After substituting (7) in (8) and solving for $\bar{m}_i(t)$, we obtain

$$\bar{m}_i = R_c^T [R_b^T (x_{fi} - x_b) - x_c]. \quad (9)$$

After utilizing (6) and (9), the pixel coordinates for the i^{th} feature point can be written as follows

$$p_i = \frac{1}{z_i} A R_c^T [R_b^T (x_{fi} - x_b) - x_c]. \quad (10)$$

The corresponding depth $z_i(t)$ can be written as follows

$$z_i = R_{c3}^T [R_b^T (x_{fi} - x_b) - x_c] \quad (11)$$

where $R_{c3}^T \in \mathbb{R}^{1 \times 3}$ is the last row of R_c^T . It should be noted that, in (10) and (11), A , R_c , x_c are known constant parameters, $R_b(t)$, $x_b(t)$ are measurable signals, and x_{fi} is an unknown constant parameter. Based on these facts $p_i(t)$ can be parameterized as follows

$$p_i = \frac{1}{\Pi \Theta_i} W \Theta_i \quad (12)$$

where

$$\Pi \Theta_i = z_i = R_{c3}^T [R_b^T (x_{fi} - x_b) - x_c] \quad (13)$$

$$W \Theta_i = A R_c^T [R_b^T (x_{fi} - x_b) - x_c]. \quad (14)$$

We note that $z_i(t)$ is assumed to satisfy the following inequalities

$$\rho_i(\cdot) \geq z_i(t) = \Pi \Theta_i \geq \varepsilon_i \quad (15)$$

where $\rho_i(\cdot) \in \mathbb{R} \forall i = 1, \dots, n$ is a positive function and $\varepsilon_i \in \mathbb{R} \forall i = 1, \dots, n$ is a positive constant. In (13) and (14), $\Pi(t) \in \mathbb{R}^{1 \times 4}$, $W(t) \in \mathbb{R}^{2 \times 4}$ are measurable regression matrices, and $\Theta_i \in \mathbb{R}^4$ is an unknown constant parameter vector¹, which is defined as

$$\Theta_i \triangleq [x_{fi1} \quad x_{fi2} \quad x_{fi3} \quad 1]^T \quad (16)$$

where $x_{fij} \in \mathbb{R} \forall j = 1, 2, 3$, is the unknown Euclidean coordinate of the i^{th} feature point, relative to the world frame. It should be noted that in (13) and (14), x_{fi} is the only unknown. After multiplying both sides of (12) with the term $\Pi(t) \Theta_i$, the following expression can be obtained

$$p_i \Pi \Theta_i = W \Theta_i. \quad (17)$$

The estimate of (17) can be defined as follows

$$\hat{p}_i \Pi \hat{\Theta}_i = W \hat{\Theta}_i \quad (18)$$

where $\hat{\Theta}_i(t) \in \mathbb{R}^4$ is the estimate for $\Theta_i \in \mathbb{R}^4$. After subtracting (18) from (17), the following expression is obtained

$$p_i \Pi \Theta_i - \hat{p}_i \Pi \hat{\Theta}_i = W \Theta_i - W \hat{\Theta}_i. \quad (19)$$

¹The reader is referred to [28] for derivations of $\Pi(t)$, $W(t)$, Θ_i , $\forall i = 1, \dots, n$.

After adding and subtracting the term $\hat{p}_i \Pi \Theta_i$ to the left-hand-side of (19) and simplifying, the following expression can be obtained [29]

$$\tilde{p}_i = \frac{1}{\Pi \Theta_i} (W - \hat{p}_i \Pi) \tilde{\Theta}_i \quad (20)$$

where $\tilde{\Theta}_i(t) \in \mathbb{R}^4$ is the estimation error defined as follows

$$\tilde{\Theta}_i \triangleq \Theta_i - \hat{\Theta}_i \quad \forall i = 1, \dots, n \quad (21)$$

and the combined prediction error for the i^{th} feature point $\tilde{p}_i(t) \in \mathbb{R}^2$ is defined as follows

$$\tilde{p}_i \triangleq p_i - \hat{p}_i \quad \forall i = 1, \dots, n. \quad (22)$$

The combination of the pixel coordinates and their estimates for all the feature points, denoted by $p(t) \in \mathbb{R}^{2n}$ and $\hat{p}(t) \in \mathbb{R}^{2n}$ respectively, are defined as follows

$$p \triangleq [p_1^T \quad p_2^T \quad \dots \quad p_n^T]^T \quad (23)$$

$$\hat{p} \triangleq [\hat{p}_1^T \quad \hat{p}_2^T \quad \dots \quad \hat{p}_n^T]^T \quad (24)$$

and the prediction error $\tilde{p}(t) \in \mathbb{R}^{2n}$ is defined as follows

$$\tilde{p} \triangleq p - \hat{p} = [\tilde{p}_1^T \quad \tilde{p}_2^T \quad \dots \quad \tilde{p}_n^T]^T. \quad (25)$$

Based on (20), the prediction error $\tilde{p}(t)$ can be written as

$$\tilde{p} = B \bar{W}_p \tilde{\Theta} \quad (26)$$

where $\bar{W}_p(t) \in \mathbb{R}^{2n \times 4n}$ is a measurable signal defined as follows

$$\bar{W}_p \triangleq \begin{bmatrix} W - \hat{p}_1 \Pi & 0_{2 \times 4} & \dots & 0_{2 \times 4} \\ 0_{2 \times 4} & W - \hat{p}_2 \Pi & \dots & 0_{2 \times 4} \\ \vdots & \vdots & \ddots & \vdots \\ 0_{2 \times 4} & 0_{2 \times 4} & \dots & W - \hat{p}_n \Pi \end{bmatrix} \quad (27)$$

where $0_{2 \times 4} \in \mathbb{R}^{2 \times 4}$ is a zero matrix, and $B(t) \in \mathbb{R}^{2n \times 2n}$ is an auxiliary matrix defined as

$$B \triangleq \text{diag} \left\{ \frac{1}{\Pi \Theta_1}, \frac{1}{\Pi \Theta_1}, \dots, \frac{1}{\Pi \Theta_n}, \frac{1}{\Pi \Theta_n} \right\}. \quad (28)$$

The combination of the estimation error, $\tilde{\Theta}(t) \in \mathbb{R}^{4n}$, is defined as follows

$$\tilde{\Theta} \triangleq \Theta - \hat{\Theta} = [\tilde{\Theta}_1 \quad \tilde{\Theta}_2 \quad \dots \quad \tilde{\Theta}_n]^T. \quad (29)$$

Based on the stability analysis, the adaptive update law $\dot{\hat{\Theta}}(t)$ is designed as follows

$$\dot{\hat{\Theta}} \triangleq \text{Proj} \{ \alpha \Gamma \bar{W}_p^T \tilde{p} \} \quad (30)$$

where $\text{Proj}\{\cdot\}$ ensures the positiveness of the term $\Pi(t) \hat{\Theta}_i(t)$ (see [30] for a detailed description), and $\alpha(t) \in \mathbb{R}$ is a positive scalar function defined as follows

$$\alpha \triangleq 1 + \frac{1}{\bar{\varepsilon}} \bar{\rho}(\cdot) \quad (31)$$

where $\bar{\rho}(\cdot) \triangleq \max_i \{\rho_i(\cdot)\} \in \mathbb{R}$ is a positive function and $\bar{\varepsilon} \triangleq \min_i \{\varepsilon_i\} \in \mathbb{R}$ is a positive constant. In (30), $\Gamma(t) \in$

$\mathbb{R}^{4n \times 4n}$ is the least-squares estimation gain matrix, designed as follows

$$\frac{d}{dt} \{\Gamma^{-1}(t)\} = 2\bar{W}_p^T \bar{W}_p. \quad (32)$$

B. Stability Analysis

Theorem 1: The update law defined in (30) ensures that $\|\hat{\Theta}(t)\| \rightarrow 0$ as $t \rightarrow +\infty$ provided that the following persistent excitation condition [31] holds

$$\gamma_1 I_{4n} \leq \int_{t_0}^{t_0+T} \bar{W}_p^T(\tau) \bar{W}_p(\tau) d\tau \leq \gamma_2 I_{4n} \quad (33)$$

where γ_1, γ_2 are positive constants, $I_{4n} \in \mathbb{R}^{4n \times 4n}$ is an identity matrix.

Proof: See [29] for a similar proof.

Remark 1: The parameter vector $\hat{\Theta}_i(t)$ provides a scaled estimate of the Euclidean coordinates of the feature points on the object relative to the world frame. Since the last element in the unknown constant parameter vector is equal to 1 (16), the scale factor can be computed as

$$\lambda_i = \hat{\Theta}_{i4} \quad (34)$$

where, $\lambda_i(t) \in \mathbb{R}$ is the scale factor for the i^{th} feature point and $\hat{\Theta}_{i4}(t) \in \mathbb{R}$ is the last entry of $\hat{\Theta}_i(t)$. It should be noted that $\hat{\Theta}_{i4}(t)$ is always nonzero which is guaranteed by the projection algorithm introduced in (30). The estimates of the Euclidean coordinates of the i^{th} feature point can now be recovered as follows

$$\begin{aligned} \hat{x}_{fi1} &= \frac{1}{\lambda_i} \hat{\Theta}_{i1} \\ \hat{x}_{fi2} &= \frac{1}{\lambda_i} \hat{\Theta}_{i2} \\ \hat{x}_{fi3} &= \frac{1}{\lambda_i} \hat{\Theta}_{i3} \end{aligned} \quad (35)$$

where $\hat{\Theta}_{ij}(t) \forall j = 1, 2, 3$, is the j^{th} element of the estimated parameter vector for i^{th} feature point.

IV. SIMULATION RESULTS

A simulation study was conducted to evaluate the performance of the proposed estimation algorithm using the Mathworks Simulink program. Simulations were performed using four static feature points whose Euclidean coordinates, with respect to the world frame, were selected as follows

$$\begin{aligned} x_{f1} &= [0 \ 1 \ 1]^T \\ x_{f2} &= [0 \ 0.5 \ 1]^T \\ x_{f3} &= [0 \ 0 \ 1]^T \\ x_{f4} &= [1 \ 1 \ 1]^T. \end{aligned} \quad (36)$$

The linear and angular positions of the mechanical system were chosen as

$$\begin{aligned} q_b &= [-0.1\cos(t) \ 0.1\sin(t) \ -0.1\sin(0.5t)]^T \text{ m} \\ \theta_b &= [0.1\sin(0.1t) \ 0 \ 0]^T \text{ rad}. \end{aligned} \quad (37)$$

Image space feature point trajectories were generated based on rigid body kinematics and known motion of the camera.

In addition, the camera's intrinsic calibration matrix and extrinsic parameters were selected as follows

$$\begin{aligned} A &= \begin{bmatrix} 825 & 0 & 320 \\ 0 & 835 & 240 \\ 0 & 0 & 1 \end{bmatrix} \\ Rc &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad x_c = [0.5 \ 0 \ 0.1]^T \text{ m}. \end{aligned} \quad (38)$$

Initial condition for the unknown vector to be estimated was taken as 100 for all its entries. The estimator gains were chosen to give the best performance were selected as follows

$$\alpha = 300, \quad \Gamma_i = \text{diag}\{4000, 4000, 4000, 4000\}. \quad (39)$$

In the simulations, three different cases were taken into consideration using the above parameters. For case 1, the pixel coordinates had no noise added to them. In case 2, Gaussian noise of variance 200, having a mean of 0, was added to the pixel coordinates and in case 3, Gaussian noise with a variance of 400, having a mean of 0, was added to the pixel coordinates.

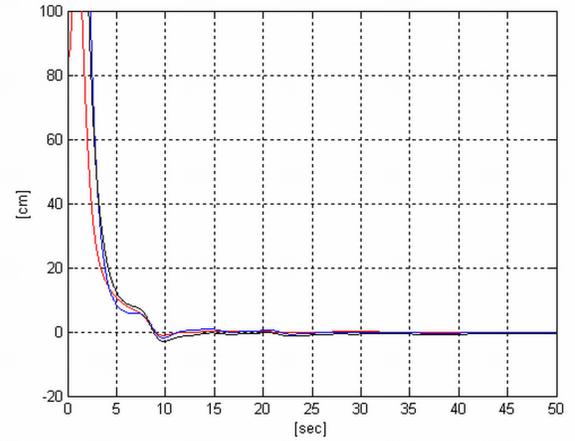


Fig. 2. Simulation case 1: Distance estimation error without additive noise.

TABLE I
SIMULATION RESULTS FOR ESTIMATED EUCLIDEAN DISTANCES

Object	Actual distance (cm)	Estimated distance (cm)	Error (cm)	%Error
Case 1				
Length I (x_{f1}, x_{f2})	50.0	49.94	0.06	0.12
Length II (x_{f2}, x_{f4})	111.8	111.25	0.55	0.49
Length III (x_{f1}, x_{f4})	100.0	99.86	0.14	0.14
Case 2				
Length I (x_{f1}, x_{f2})	50.0	49.90	0.10	0.20
Length II (x_{f2}, x_{f4})	111.8	111.15	0.65	0.58
Length III (x_{f1}, x_{f4})	100.0	99.74	0.26	0.26
Case 3				
Length I (x_{f1}, x_{f2})	50.0	49.88	0.12	0.24
Length II (x_{f2}, x_{f4})	111.8	111.08	0.72	0.64
Length III (x_{f1}, x_{f4})	100.0	99.65	0.35	0.35

The simulation results for the Euclidean distance estimation for each of the cases is shown in Table I. Here it should

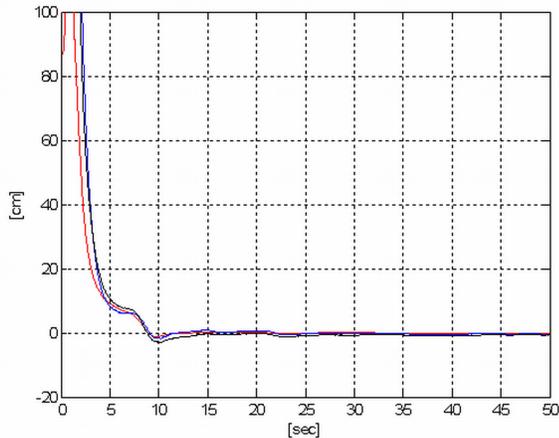


Fig. 3. Simulation case 2: Distance estimation error with Gaussian noise of variance 200.

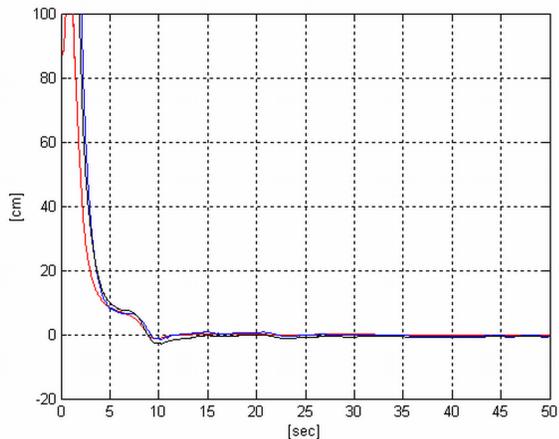


Fig. 4. Simulation case 3: Distance estimation error with Gaussian noise of variance 400.

be noted that, for all the estimated lengths, the percentage error is below 1% for all the cases and below 0.5% for most of the lengths. It can be easily inferred from Table I that even the image points with noise (case 2 and case 3) do not make any significant effect on the estimator. All the above mentioned cases demonstrate the robustness of the estimator. The estimated Euclidean distance error without additive noise is shown in Figure 2. Figures 3 and 4 show distance estimation error for cases 2 and 3 respectively. Note that the error signals converge in just a few seconds.

V. CONCLUSION

An estimation technique for reconstruction of 3D Euclidean coordinates of a static object, with a moving calibrated camera whose position is measurable, was presented. An adaptive update law was designed after utilizing a unique prediction error formulation. It was proven that Euclidean distance estimation signals are driven to zero, upon satisfaction of a persistent excitation condition. Numerical simulation results were presented demonstrating the robustness and

accuracy of the estimator. The estimator accurately identifies the Euclidean distances between the feature points without having any information of the object.

REFERENCES

- [1] J. Oliensis, "A critique of structure from motion algorithms," *Computer Vision and Image Understanding*, vol. 80, no. 2, pp. 172–214, 2000.
- [2] F. Chaumette, S. Boukir, and D. Juvin, "Structure from controlled motion," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 18, no. 5, pp. 492–504, May 1996.
- [3] A. Chiuso, P. Favaro, H. Jin, and S. Soatto, "Structure from motion casually integrated over time," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 24, no. 4, pp. 523–535, Apr. 2002.
- [4] H. Durrant-Whyte and T. Bailey, "Simultaneous localization and mapping (slam): Part ii," *IEEE Robot. and Automat. Mag.*, vol. 13, no. 3, pp. 108–117, Sep. 2006.
- [5] A. J. Davison, I. D. Reid, N. D. Molton, and O. Stasse, "Monoslam: Real-time single camera slam," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 29, no. 6, pp. 1052–1067, Jun. 2007.
- [6] T. Fukao, K. Fujitani, and T. Kanade, "An autonomous blimp for a surveillance system," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Las Vegas, NV, 2003, pp. 1820–1825.
- [7] T. Kanade, O. Amidi, and Q. Ke, "Real-time and 3d vision for autonomous small and micro air vehicles," in *Proc. IEEE Int. Conf. Decision and Control*, Dec. 2004, pp. 1655–1662.
- [8] J. D. Redding, T. W. McLain, R. W. Beard, and C. N. Taylor, "Vision-based target localization from a fixed-wing miniature air vehicle," in *Proc. American Control Conf.*, Minneapolis, MN, Jun. 2006, pp. 2862–2867.
- [9] V. N. Dobrokhodov, I. I. Kaminer, K. D. Jones, and R. Ghabelloo, "Vision based tracking and motion estimation for moving targets using small uavs," in *Proc. American Control Conf.*, Minneapolis, MN, Jul. 2006, pp. 1428–1433.
- [10] J. H. Kim and S. Sukkarieh, "Airborne simultaneous localisation and map building," in *Proc. IEEE Int. Conf. Robot. Autom.*, Taipei, Taiwan, 2003, pp. 406–411.
- [11] I. K. Jung and S. Lacroix, "High resolution terrain mapping using low altitude aerial stereo imagery," in *Proc. IEEE Int. Conf. Computer Vision*, Nice, France, 2003, pp. 1820–1825.
- [12] I. Miyagawa and K. Arakawa, "Motion and shape recovery based on iterative stabilization for modest deviation from planar motion," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 28, no. 7, pp. 1176–1181, Jul. 2006.
- [13] S. J. Julier and J. K. Uhlmann, "A counter example to the theory of simultaneous localization and map building," in *Proc. IEEE Int. Conf. Robot. Autom.*, Seoul, Korea, 2001, pp. 4238–4243.
- [14] M. Jankovic and B. K. Ghosh, "Visually guided ranging from observations of points, lines and curves via an identifier based nonlinear observer," *Systems and Control Letters*, vol. 25, pp. 63–73, 1995.
- [15] X. Chen and H. Kano, "State observer for a class of nonlinear systems and its application to machine vision," *IEEE Trans. Automat. Contr.*, vol. 49, no. 11, pp. 2085–2091, Nov. 2004.
- [16] X. Hu and T. Ersson, "Active state estimation of nonlinear systems," *Automatica*, vol. 40, pp. 2075–2082, 2004.
- [17] O. Dahl, F. Nyberg, and A. Heyden, "Nonlinear and adaptive observers for perspective dynamic systems," in *Proc. American Control Conf.*, New York, NY, Jul. 2007, pp. 1966–971.
- [18] K. Dupree, N. R. Gans, W. MacKunis, and W. E. Dixon, "Euclidean calculation of feature points of a rotating satellite: a daisy chaining approach," in *Proc. American Control Conf.*, New York, NY, Jul. 2007, pp. 3874–3879.
- [19] R. Laganieri and P. Cohen, "Gradual perception of structure from motion: a neural approach," *IEEE Trans. Neural Networks*, vol. 6, no. 3, pp. 736–748, 1995.
- [20] V. K. Chitrakaran, D. M. Dawson, J. Chen, and H. Kannan, "Velocity and structure estimation of a moving object using a monocular camera," in *Proc. American Control Conf.*, Minneapolis, MN, Jun. 2006, pp. 5159–5164.
- [21] V. K. Chitrakaran and D. M. Dawson, "A Lyapunov-based method for estimation of Euclidean position of static features using a single camera," in *Proc. American Control Conf.*, New York, NY, Jul. 2007, pp. 1988–1993.

- [22] D. Braganza, D. M. Dawson, and T. Hughes, "Euclidean position estimation of static features using a moving camera with known velocities," in *Proc. IEEE Int. Conf. Decision and Control*, New Orleans, LA, Dec. 2007, pp. 2695–2700.
- [23] A. De Luca, G. Oriolo, and P. R. Giordano, "On-line estimation of feature depth for image-based visual servoing schemes," in *Proc. IEEE Int. Conf. Robot. Autom.*, Roma, Italy, Apr. 2007, pp. 2823–2828.
- [24] L. Matthies, T. Kanade, and R. Szeliski, "Kalman filter-based algorithms for estimating depth from image sequences," *Int. Journal of Computer Vision*, vol. 3, no. 3, pp. 209–238, 1989.
- [25] C. E. Smith and N. P. Papanikolopoulos, "Computation of shape through controlled active exploration," in *Proc. IEEE Int. Conf. Robot. Autom.*, San Diego, CA, May 1994, pp. 2516–2521.
- [26] O. Faugeras, *Three-Dimensional Computer Vision*. Cambridge, MA: MIT Press, 1993.
- [27] E. Malis and F. Chaumette, "2 1/2 D visual servoing with respect to unknown objects through a new estimation scheme of camera displacement," *Int. Journal of Computer Vision*, vol. 37, no. 1, pp. 79–97, 2000.
- [28] D. Braganza, N. Nath, and D. M. Dawson, "Derivation of regression matrices for position based structure from motion for camera-in-hand configuration," Clemson University CRB, Tech. Rep. CU/CRB/5/9/07/1, May 2007. [Online]. Available: <http://www.ces.clemson.edu/ece/crb/publicn/tr.htm>
- [29] E. Tatlicioglu, D. Dawson, and B. Xian, "Adaptive visual servo regulation control for camera-in-hand configuration with a fixed-camera extension," in *Proc. IEEE Int. Conf. Decision and Control*, New Orleans, LA, Dec. 2007, pp. 2339–2344.
- [30] E. Tatlicioglu, D. M. Dawson, and B. Xian, "Adaptive visual servo regulation control for camera-in-hand configuration with a fixed-camera extension," Clemson University CRB, Tech. Rep. CU/CRB/3/9/06/1, Mar. 2007. [Online]. Available: <http://www.ces.clemson.edu/ece/crb/publicn/tr.htm>
- [31] J. J. E. Slotine and W. Li, *Applied Nonlinear Control*. Englewood Cliffs, NJ: Prentice Hall, 1991.