2010 American Control Conference
Marriott Waterfront, Baltimore, MD, USA
June 30-July 02, 2010

Structure and Motion Estimation of a Moving Object Using a Moving Camera

Ashwin P. Dani, Zhen Kan, Nicholas R. Fischer, Warren E. Dixon

***Abstract*— A solution is presented to the problem of estimating the structure and motion of a moving object seen from a moving camera. A nonlinear observer is proposed, which asymptotically identifies the structure and motion of the moving object, when the camera motion is persistently exciting. The object is assumed to be moving with constant velocities. The proposed method makes no assumptions on the minimum number of views or point correspondences as required by the existing approaches.**

I. Introduction

The problem of recovering the structure of a static scene using a moving camera, called 'structure from motion (SfM)', is well understood. A number of solutions to the SfM problem are given in the form of batch methods [1]–[7] as well as online methods [8]–[15]. The solution to the SfM problem (e.g. see [12], [15]) can be used to self-localize a camera with respect to its environment. Triangulation is feasible if a stationary point in the scene can be viewed from two different camera locations (i.e. the scene is static). Since SfM techniques rely on triangulation, they cannot be used to recover the structure and motion of moving objects [16]. A need arises to answer the question: *Given observations of point correspondences in every image of a video stream with known camera motion, is it possible to recover the Euclidean structure and motion (i.e. linear and angular velocities) of independently moving objects observed by the moving camera?*

The problem stated above is referred to as "structure and motion from motion (SaMfM)" in this paper. In practice, the motivation to solve the SaMfM problem comes from scenarios such as to determine the range and speed of cars moving on a highway as observed from an airborne helicopter. In another example, consider an object-grabbing robotic arm, equipped with a hand-held camera, grabbing randomly placed objects moving on a conveyer belt. Estimation of the range and velocity of a lead vehicle using a camera mounted on a follower would help in formation control of a convoy of unmanned ground vehicles (UGVs). Camera velocities for these applications can be measured using sensors such as a global positioning system (GPS) or an inertial measurement unit (IMU).

Solutions exist in literature for the specific cases of the SaMfM problem where constraints are applied to the trajectories and velocities of the moving object. The pioneering work in [16] referred to the SaMfM problem as "trajectory triangulation" and provided a solution where at least five views are required if the motion of the object is constrained to a straight line and at least nine views are required if the object is moving with conic trajectories. However, convergence of the method is not guaranteed. In [17], the structure and motion of the objects moving with linear or conic trajectories are recovered from tangent projections, provided at least nine views are available and the motion of the camera is known. In [18], the SaMfM problem is solved with constant velocities, assuming an approximate orthographic projection camera model. In [19], a stereo camera is used to provide a solution to the SaMfM problem with at least four views. In [20], a method is developed based on Homography and the SaMfM problem is solved using five point correspondences in three views. The method in [20] does not allow moving points to be on different motion planes.

In this paper, a solution to a specific case of the SaMfM problem is presented where the objects are moving independently along straight lines in an arbitrary direction. The structure and motion for each object in the scene can be recovered independently using the camera velocities and the feature point data obtained from an image sequence. The proposed method has several advantages over the existing methods. There are no requirements of minimum number of point correspondences or number of views. Another advantage is that the nonlinear observer processes the data in every image as it arrives, and thus, can perform real-time computation of the structure and motion of a moving object. The batch methods collect data from multiple images and then process it. Hence, processing time for the proposed observer is less than that of the batch methods. A stability analysis of the proposed observer is presented which guarantees convergence of the observer, provided an observability condition based on the persistency of excitation (PE) of the camera motion is satisfied. Convergence of batch methods is not always guaranteed.

II. Euclidean to Image Space Mapping

Consider a scenario depicted in Fig. 1 where a moving camera views moving point objects. In Fig. 1, an inertial reference frame is denoted by $\mathcal{F}^{*1}$. After the initial time, a

[1]This research is supported in part by the NSF CAREER award number 0547448, NSF award number 0901491, the Department of Energy, grant number DE-FG04-86NE37967 as part of the DOE University Research Program in Robotics (URPR), ASTREC, and NSF I/UCRC.

[2]The authors are with the Department of Mechanical and Aerospace Engineering, University of Florida, Gainesville, FL 32611-6250, USA. Email: {ashwin31, kanzhen0322, moooink, wdixon}@ufl.edu

[1]w.l.o.g. $\mathcal{F}^*$ can be attached to the camera at the location corresponding to an initial point in time t_0 where the object is in the camera field of view (FOV) and $\mathcal{F}^*$ is identical to $\mathcal{F}_c(t_0)$.

978-1-4244-7427-1/10/$26.00 ©2010 AACC

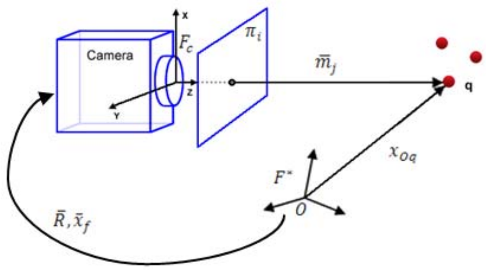

Fig. 1. Objects as seen from the camera and coordinate relationships.

reference frame $\mathcal{F}_c$ attached to a pinhole camera undergoes some rotation $\bar{R}(t) \in SO(3)$ and translation $\bar{x}_f(t) \in \mathbb{R}^3$ away from $\mathcal{F}^*$.

The Euclidean coordinates $\bar{m}_j(t) \in \mathbb{R}^3$ (where $j = \{1, 2, ..., n\}$ denotes a point number) of points observed by a camera expressed in the camera frame $\mathcal{F}_c$ and the respective normalized Euclidean coordinates $m_j(t) \in \mathbb{R}^3$ are defined as

$$\bar{m}_j(t) = \begin{bmatrix} x_{1j}(t), & x_{2j}(t), & x_{3j}(t) \end{bmatrix}^T, \quad (1)$$

$$m_j(t) = \begin{bmatrix} \frac{x_{1j}(t)}{x_{3j}(t)}, & \frac{x_{2j}(t)}{x_{3j}(t)}, & 1 \end{bmatrix}^T. \quad (2)$$

Consider a closed and bounded set $\mathcal{Y} \subset \mathbb{R}^3$. To facilitate the subsequent development, the state vector $y_j(t) = [y_{1j}(t), y_{2j}(t), y_{3j}(t)]^T \in \mathcal{Y}$ is constructed from (2) as

$$y_j = \begin{bmatrix} \frac{x_{1j}}{x_{3j}}, & \frac{x_{2j}}{x_{3j}}, & \frac{1}{x_{3j}} \end{bmatrix}^T. \quad (3)$$

Using projective geometry, the normalized Euclidean coordinates $m_j(t)$ can be related to the pixel coordinates in the image space as

$$q_j = A m_j \quad (4)$$

where $q_j(t) = \begin{bmatrix} u_j(t) & v_j(t) & 1 \end{bmatrix}^T$ is a vector of the image-space feature point coordinates $u_j(t)$, $v_j(t) \in \mathbb{R}$ defined on the closed and bounded set $\mathcal{I} \subset \mathbb{R}^3$, and $A \in \mathbb{R}^{3\times3}$ is a constant, known, invertible camera calibration matrix [21]. Since A is known, the expression in (4) can be used to recover $m_j(t)$, which can be used to partially reconstruct the state $y_j(t)$ so that the first two components of $y_j(t)$ can be determined.

Assumption 1: The relative Euclidean distance $x_{3j}(t)$ between the camera and the feature points observed on the target is upper and lower bounded by some known positive constants (i.e., the object remains within some finite distance away from the camera). Therefore, the definition in (3) can be used to assume that

$$\bar{y}_3 \geq y_{3j}(t) \geq \underline{y_3} \quad (5)$$

where $\bar{y}_3, \underline{y_3} \in \mathbb{R}$ denote known positive bounding constants. Likewise, since the image coordinates are constrained (i.e., the target remains in the camera field of view), the relationships in (2), (3), and (4) along with the fact that A is invertible can be used to conclude that

$$\bar{y}_1 \geq |y_{1j}(t)| \geq \underline{y_1} \qquad \bar{y}_2 \geq |y_{2j}(t)| \geq \underline{y_2}$$

where $\bar{y}_1$, $\bar{y}_2$, $\underline{y_1}$, $\underline{y_2} \in \mathbb{R}$ denote known positive bounding constants.

Assumption 2: The motion of the camera is assumed to be smooth such that the acceleration is bounded by a constant. Thus, $y_j(t)$ belongs to class C^2, which also implies that the second derivative of $y_j(t)$ is bounded by a constant.

For the remainder of this paper, the feature point subscript j is omitted to streamline the notation.

III. Camera Motion Model

Consider the moving camera viewing a moving point q. As seen from Fig. 1, the point q can be expressed in the coordinate system $\mathcal{F}_c$ as

$$\bar{m} = \bar{x}_f + \bar{R} x_{oq} \quad (6)$$

where x_{oq} is a vector from the origin of coordinate system $\mathcal{F}^*$ to the point q expressed in the coordinate system $\mathcal{F}^*$. Differentiating (6), the relative motion of q as observed in the camera coordinate system can be expressed by the following kinematics [21], [22]

$$\dot{\bar{m}} = [\omega]_\times \bar{m} - v_r \quad (7)$$

where $\bar{m}(t)$ is defined in (1), $[\omega]_\times \in \mathbb{R}^{3\times3}$ denotes a skew symmetric matrix formed from the angular velocity vector of the camera $\omega(t) = \begin{bmatrix} \omega_1 & \omega_2 & \omega_3 \end{bmatrix}^T \in \mathcal{W}$, and $v_r(t)$ represents the relative velocity of the camera with respect to the moving point, defined as

$$v_r = v_c - v_p. \quad (8)$$

In (8), $v_c(t)$ denotes the camera velocity in the inertial reference frame given by $v_c(t) = \begin{bmatrix} v_{cx} & v_{cy} & v_{cz} \end{bmatrix}^T \in \mathcal{V}_c$ and $v_p(t)$ denotes the velocity of the point in the camera reference frame given by $v_p(t) = \begin{bmatrix} v_{px} & v_{py} & v_{pz} \end{bmatrix}^T \in \mathcal{V}_p$. The sets $\mathcal{W}$, $\mathcal{V}_c$ and $\mathcal{V}_p$ are closed and bounded sets such that $\mathcal{W} \subset \mathbb{R}^3, \mathcal{V}_c \subset \mathbb{R}^3$ and $\mathcal{V}_p \subset \mathbb{R}^3$.

Assumption 3: For the subsequent development of an observer, the point velocities are assumed to be constant.

IV. Structure and Motion Estimation

A. Structure and Motion from Motion (SaMfM) Objective

The objective of SaMfM is to recover the structure (i.e. Euclidean coordinates) and motion (i.e. Euclidean linear and angular velocities) of moving objects observed by a moving camera, assuming that all the camera velocities are known. The object can be tracked as a single point or a collection of feature points, where the range (i.e., $\frac{1}{x_3(t)}$) and motion of each point should be estimated.

B. State Dynamics Formulation

The states defined in (3) contain unknown structure information of the object. To facilitate the observer design, states are defined in this section to incorporate unknown structure and velocity information. Specifically, an auxiliary state vector $p(t) = \begin{bmatrix} p_1(t) & p_2(t) & p_3(t) \end{bmatrix}^T \in \mathbb{R}^3$ is defined as

$$p \triangleq \begin{bmatrix} v_{px}y_3(t) & v_{py}y_3(t) & v_{pz}y_3(t) \end{bmatrix}^T \tag{9}$$

which incorporates the unknown object velocity information. To recover the 3D structure, the state $y_3(t)$ should be estimated since $y_3(t)$ contains range information. Since, the states $y_1(t), y_2(t)$ can be measured from the images, the estimated state $y_3(t)$ can be used to scale $y_1(t)$ and $y_2(t)$, and thus $\bar{m}(t)$, i.e. the 3D structure can be recovered. To recover the velocity information, the state $p(t)$ must be estimated. Once the states $y_3(t)$ and $p(t)$ are estimated, velocity information can be recovered by scaling the estimated $p(t)$ by the estimated $y_3(t)$. Using (3) and (7), the dynamics of the state vector $y(t)$ are expressed as

$$\begin{aligned} \dot{y}_1 &= \Omega_1 + (-v_{cx} + y_1 v_{cz})y_3 + p_1 - y_1 p_3, \\ \dot{y}_2 &= \Omega_2 + (-v_{cy} + y_2 v_{cz})y_3 + p_2 - y_2 p_3, \\ \dot{y}_3 &= -v_{cz}y_3^2 + (y_2\omega_1 - y_1\omega_2)y_3 + v_{pz}y_3^2 \end{aligned} \tag{10}$$

where $\Omega_1(t) \in \mathbb{R}$ and $\Omega_2(t) \in \mathbb{R}$ are defined as

$$\begin{aligned} \Omega_1(t) &\triangleq -\omega_2 + y_2\omega_3 + y_1 y_2 \omega_1 - y_1^2\omega_2, \\ \Omega_2(t) &\triangleq \omega_1 - y_1\omega_3 - y_1 y_2\omega_2 + y_2^2\omega_1. \end{aligned}$$

Differentiating (9) and using (10) along with Assumption 3, the dynamics of the state $p(t)$ can be represented by following set of differential equations

$$\begin{aligned} \dot{p}_1 &= -v_{cz}p_1 y_3 + (y_2\omega_1 - y_1\omega_2)p_1 + p_3 p_1, \\ \dot{p}_2 &= -v_{cz}p_2 y_3 + (y_2\omega_1 - y_1\omega_2)p_2 + p_3 p_2, \\ \dot{p}_3 &= -v_{cz}p_3 y_3 + (y_2\omega_1 - y_1\omega_2)p_3 + p_3^2. \end{aligned} \tag{11}$$

By defining the vector $z(t) \in \mathbb{R}^2$ and vector $\theta(t) \in \mathbb{R}^4$ as

$$\begin{aligned} z(t) &\triangleq \begin{bmatrix} y_1 & y_2 \end{bmatrix}^T, \\ \theta(t) &\triangleq \begin{bmatrix} y_3 & p_1 & p_2 & p_3 \end{bmatrix}^T \end{aligned}$$

the state dynamics given by (10) and (11) can be expressed as

$$\begin{aligned} \dot{z} &= \Omega(z,u) + J(z,u)\theta, \\ \dot{\theta} &= g(z,\theta,u) \end{aligned} \tag{12}$$

where $\Omega(t) = \begin{bmatrix} \Omega_1(t) & \Omega_2(t) \end{bmatrix}^T$, $u(t) = \begin{bmatrix} v_c(t) & \omega(t) \end{bmatrix}^T$ and the functions $J(z,u) \in \mathbb{R}^{2\times 4}$ and $g(z,\theta,u) \in \mathbb{R}^4$ are given by

$$J = \begin{bmatrix} (-v_{cx} + y_1 v_{cz}) & 1 & 0 & -y_1 \\ (-v_{cy} + y_2 v_{cz}) & 0 & 1 & -y_2 \end{bmatrix}, \tag{13}$$

and

$$g = \begin{bmatrix} v_{cz}y_3^2 + (y_2\omega_1 - y_1\omega_2)y_3 - p_3 y_3 \\ v_{cz}p_1 y_3 + (y_2\omega_1 - y_1\omega_2)p_1 - p_3 p_1 \\ v_{cz}p_2 y_3 + (y_2\omega_1 - y_1\omega_2)p_2 - p_3 p_2 \\ v_{cz}p_3 y_3 + (y_2\omega_1 - y_1\omega_2)p_3 - p_3^2 \end{bmatrix}. \tag{14}$$

A nonlinear observer is designed to estimate the parameters $\theta(t)$ which contain unknown depth and unknown velocity information of the moving object.

Assumption 4: The function $g(z,\theta,u)$ is locally Lipschitz with respect to the second argument.

Assumption 5: The signal $v_c(t)$ is of class $\mathcal{C}^2$, hence, $\dot{v}_c(t)$ and $\ddot{v}_c(t) \in \mathcal{L}_\infty$

Assumption 6: There exists a positive constant $\gamma \in \mathbb{R}$ and small positive constant $\tau \in \mathbb{R}$ such that the inequality

$$\int_t^{t+\tau} J^T(\beta)J(\beta)d\beta \geq \gamma I$$

is satisfied for all $t \geq 0$. This is a persistency of excitation condition for the camera motion.

Remark 1: Based on Assumptions 1-3 and 5, $v_c(t)$ and $v_p(t)$ belong to class $\mathcal{C}^2$. Thus, the following inequalities can be obtained

$$\|\theta(t)\| \leq \bar{\theta}, \quad \left\|\dot{\theta}(t)\right\| \leq \bar{\xi}_3, \quad \left\|\ddot{\theta}(t)\right\| \leq \bar{\xi}_4$$

where $\bar{\theta}$, $\bar{\xi}_3$, $\bar{\xi}_4 \in \mathbb{R}$ denote known bounding constants.

Remark 2: Using the fact that the camera and the point velocities $v_c(t)$, $\omega(t)$, and $v_p(t)$ are bounded above, along with Assumption 1, 2 and 5, an upper bound on $J(z,u)$, $\dot{J}(z,u)$, $\ddot{J}(z,u)$ can be established as

$$\|J(z,u)\| \leq \bar{\xi}_5, \quad \left\|\dot{J}(z,u)\right\| \leq \bar{\xi}_6, \quad \left\|\ddot{J}(z,u)\right\| \leq \bar{\xi}_7$$

where $\bar{\xi}_5$, $\bar{\xi}_6$, $\bar{\xi}_7 \in \mathbb{R}$ denote known bounding constants.

Remark 3: Even though the rank of $J^T(z,u)J(z,u)$ can be at most 2, the integration of $J^T(z,u)J(z,u)$ can achieve full rank [11], [23], [24]. The PE condition in Assumption 6 requires the camera velocities to be persistently exciting, which means that the camera should not be translating parallel to the projected ray or camera must be translating in all three directions simultaneously, i.e. $v_{cx} = v_{cy} = v_{cz} \neq 0$.

C. State Estimator

To quantify the objective of estimating $z(t)$ and $\theta(t)$, the errors in estimation denoted by $e(t) \in \mathbb{R}^2$ and $\tilde{\theta}(t) \in \mathbb{R}^4$ are defined as

$$e \triangleq z - \hat{z}, \qquad \tilde{\theta} \triangleq \theta - \hat{\theta}. \tag{15}$$

To facilitate the stability analysis, the filtered error $r(t) \in \mathbb{R}^2$ is defined as

$$r \triangleq \dot{e} + \alpha e. \tag{16}$$

where $\alpha \in \mathbb{R}$ denote a positive constant. Based on the structure of (12), a continuous nonlinear observer is designed as

$$\dot{\hat{z}} = \Omega(z,u) + J(z,u)\hat{\theta} + \eta,$$
$$\dot{\hat{\theta}} = proj(\hat{\theta}, \phi) \quad (17)$$

where $proj(\cdot)$ is a smooth projection operator [25], [26] and $\phi(z, \hat{\theta}, u, e) \in \mathbb{R}^4$ is defined as

$$\phi \triangleq g(z, \hat{\theta}, u) + \Gamma J^T(\eta - \alpha e) \quad (18)$$

where $\Gamma \in \mathbb{R}^{4\times4}$ is a gain matrix. In (17), $\hat{z}(t) \in \mathbb{R}^2$ denotes the estimate of the measurable state $z(t)$ given by $\hat{z}(t) \triangleq \begin{bmatrix} \hat{y}_1(t) & \hat{y}_2(t) \end{bmatrix}^T$ and $\hat{\theta}(t) \in \mathbb{R}^4$ denotes an estimate of $\theta(t)$ given by $\hat{\theta}(t) = \begin{bmatrix} \hat{y}_3(t) & \hat{p}_1(t) & \hat{p}_2(t) & \hat{p}_3(t) \end{bmatrix}^T$. The term $\eta(t) \in \mathbb{R}^2$ is defined as the generalized solution to

$$\dot{\eta} = (K + I_{2\times2})r(t) + \rho sgn(e(t)) - \alpha^2 e(\tau)d\tau \quad (19)$$

where $K, \rho \in \mathbb{R}^{2\times2}$ are diagonal gain matrices.

After utilizing the errors in (15) and the proposed estimator in (17), the error dynamics can be expressed as

$$\dot{e} = J\tilde{\theta} - \eta,$$
$$\dot{\tilde{\theta}} = g - \hat{g} - \Gamma J^T(\eta - \alpha e). \quad (20)$$

Differentiating (16), the open-loop error is given by

$$\dot{r} = J\dot{\tilde{\theta}} + \dot{J}\tilde{\theta} - \dot{p} + \alpha\dot{e},$$
$$= J\dot{\theta} - J\dot{\hat{\theta}} + \dot{J}\theta - \dot{J}\hat{\theta} - \dot{\eta} + \alpha\dot{e}. \quad (21)$$

The following upper bounds on $\hat{\theta}(t)$ and $\dot{\hat{\theta}}(t)$ can be established

$$\left\|\hat{\theta}(t)\right\| \leq \zeta_1, \qquad \left\|\dot{\hat{\theta}}(t)\right\| \leq \zeta_2 + \zeta_3 \|r\| \quad (22)$$

where ζ_1, ζ_2 and $\zeta_3 \in \mathbb{R}$ are bounding constants. The bound on $\hat{\theta}(t)$ comes from the smooth projection operator used in the estimator design (17). From the upper bounds of $\theta(t)$ and $\hat{\theta}(t)$, and using (15), an upper bound of $\tilde{\theta}(t)$ can be determined. The bound on $\dot{\hat{\theta}}(t)$ can be established by substituting $\eta(t)$ from (20) into $\dot{\hat{\theta}}(t)$ and utilizing bounds on $\hat{\theta}(t), \tilde{\theta}(t), J(t), z(t)$. Terms in (21) can be combined as

$$\dot{r} = \chi_1 + \chi_2 - \dot{\eta} + \alpha\dot{e} \quad (23)$$

where the auxiliary terms $\chi_1(t)$ and $\chi_2(t)$ are defined as

$$\chi_1 \triangleq J\dot{\theta} + \dot{J}\tilde{\theta} + \|J\|\zeta_2,$$
$$\chi_2 \triangleq -J\dot{\hat{\theta}} - \|J\|\zeta_2. \quad (24)$$

The following bounds can be established for $\chi_1(t)$, $\dot{\chi}_1(t)$ and $\chi_2(t)$ based on Remarks 1-3

$$\|\chi_1\| \leq \varsigma_1, \qquad \|\chi_2\| \leq \varsigma_2\|r\|, \quad (25)$$
$$\left\|\dot{\chi}_1\right\| \leq \varsigma_3 + \varsigma_4\|r\|$$

where $\varsigma_i \in \mathbb{R}$, $i = (1, .., 4)$ are known positive constants. The signals $\chi_1(t)$ and $\chi_2(t)$ are created to separate terms bounded by constants and by state dependencies in (21). The terms inside $\chi_2(t)$ which are upper bounded by constants are removed from $\chi_2(t)$ and combined with $\chi_1(t)$ as $\|J\|\zeta_2$. This segregation of terms is helpful in the stability analysis. Utilizing the robust term in (19) and the open-loop error system (23), the closed-loop error system is expressed as

$$\dot{r} = \chi_1 + \chi_2 - (K + I)r - \rho sgn(e) + \alpha r. \quad (26)$$

D. Stability Analysis

The stability of the observer in (17) is analyzed by first analyzing the stability of the $\dot{\hat{z}}(t)$ dynamics. The presence of $\tilde{\theta}(t)$ in the $\dot{e}(t)$ error dynamics can be treated as a bounded disturbance. Thus, a robust term $\eta(t)$ is used to crush the disturbances and to drive $e(t)$ and $\dot{e}(t)$ to zero. Once, $e(t)$ and $\dot{e}(t)$ are driven to zero, the designed term $\eta(t)$ identifies the disturbance term $J(z,u)\tilde{\theta}(t)$, which can be used to stabilize the $\dot{\tilde{\theta}}(t)$ dynamics. Thus, the stability analysis of error $e(t)$ is shown first and then the stability of error $\tilde{\theta}(t)$ is analyzed. The $\dot{\tilde{\theta}}(t)$ error dynamics is a linear differential equation with vanishing disturbances. Tools from linear systems theory are used to achieve $\tilde{\theta}(t) \in \mathcal{L}_\infty$ and $\left\|\tilde{\theta}(t)\right\| \to 0$ as $t \to \infty$, which is the ultimate goal of this paper.

Theorem: The observer in (17) is asymptotically stable in the sense

$$\|e(t)\| \to 0 \quad \text{as} \quad t \to \infty \text{ and } \left\|\tilde{\theta}(t)\right\| \to 0 \quad \text{as} \quad t \to \infty$$

provided Assumptions 1-6 and following sufficient conditions are satisfied

$$\rho \geq \varsigma_1 + \frac{1}{\alpha}\varsigma_3, \qquad \beta \geq \varsigma_4. \quad (27)$$

Proof: The proof is given in two parts. First the stability of $\dot{e}(t)$ is analyzed followed by the stability of $\dot{\tilde{\theta}}(t)$ dynamics.

1) Stability of $\dot{e}(t)$ dynamics: Consider a domain $\mathcal{D} \subset \mathbb{R}^5$ containing $\psi(t) = 0$, where $\psi(t) \in \mathbb{R}^5$ is defined as

$$\psi(t) \triangleq \begin{bmatrix} r^T & e^T & \sqrt{P(t)} \end{bmatrix}^T. \quad (28)$$

The auxiliary function $P(t) \in \mathbb{R}$ in (28) is defined as

$$P = \rho\|e(0)\| - e^T(0)\chi_1(0) - L(t) + L(0)$$

where the signal $L(t) \in \mathbb{R}$ is generated as

$$\dot{L}(t) = r^T(\tau)(\chi_1 - \rho sgn(e(\tau))) - \beta\|e\|\|r\|$$

and $\beta \in \mathbb{R}$ is chosen according to (27). It can be proven that $P(t) \geq 0$, in a similar manner as [27], [28] provided the sufficient conditions in (27) are satisfied. Let $V_e(y,t) : \mathcal{D} \times [0, \infty) \to \mathbb{R}$ be a continuously differentiable non-negative radially unbounded function defined as

$$V_e(y,t) \triangleq \frac{1}{2}r^T r + \frac{1}{2}e^T e + P \quad (29)$$

where $K_1 \in \mathbb{R}^{2\times 2}$ is a diagonal matrix. After utilizing the error derivatives from (16) and (23), the time derivative of (29) is given by

$$\dot{V}_e = r^T\chi_2 - r^T(K+I)r + \alpha r^T r + e^T r - \alpha e^T e + \beta \|e\| \|r\| .$$

Using the fact,

$$\|e\| \|r\| \leq \frac{1}{2}\|e\|^2 + \frac{1}{2}\|r\|^2 ,$$

and after utilizing the bound on $\chi_2(t)$ in (25), the following inequality can be obtained

$$\begin{aligned} \dot{V}_e &\leq \varsigma_2 \|r\|^2 - k\|r\|^2 + \alpha\|r\|^2 + \frac{\beta+1}{2}\|r\|^2 \\ &\quad + \frac{\beta+1}{2}\|e\|^2 - \alpha\|e\|^2 , \\ \dot{V}_e &\leq -(k - \varsigma_2 - \alpha - \frac{\beta+1}{2})\|r\|^2 - (\alpha - \frac{\beta+1}{2})\|e\|^2 \end{aligned}$$

where $k \in \mathbb{R} \triangleq \max\{k_i\}\ \forall i = (1,2)$, with k_i being non-zero entrees of the diagonal gain matrix K. Choosing $\alpha > \frac{\beta+1}{2}$ and $k > \varsigma_2 + \alpha + \frac{\beta+1}{2}$, the following inequality can be established

$$\dot{V}_e \leq -(k - \varsigma_2 - \alpha - \frac{\beta+1}{2})\|r\|^2 . \qquad (30)$$

Using inequalities (29) and (30) it can be inferred that $V_e(y,t) \in \mathcal{L}_\infty$; thus $r(t), e(t) \in \mathcal{L}_\infty$. Since, $r(t)$ and $e(t) \in \mathcal{L}_\infty$, using linear analysis (16) can be used to show that $\dot{e}(t) \in \mathcal{L}_\infty$. Since, $e(t), \dot{e}(t) \in \mathcal{L}_\infty$, (19) can be used to show that $\dot{\eta}(t) \in \mathcal{L}_\infty$. From $\dot{e}(t), \dot{\eta}(t) \in \mathcal{L}_\infty$, it can be shown that $\dot{r}(t) \in \mathcal{L}_\infty$. Also, (30) implies $r(t) \in \mathcal{L}_2$. Using the fact that $r(t) \in \mathcal{L}_2 \cap \mathcal{L}_\infty$ and $\dot{r}(t) \in \mathcal{L}_\infty$, Barbalat's lemma [29] can be invoked to prove that

$$\|r(t)\| \to 0 \quad \text{as} \quad t \to \infty. \qquad (31)$$

Based on the definition of $r(t)$, using linear analysis techniques, (31) can be used to prove that

$$\|e(t)\| \to 0 \quad \text{as} \quad t \to \infty.$$

2) Stability of $\dot{\tilde{\theta}}$ dynamics: From the equation (20), it can be observed that as $t \to \infty$, $\eta(t)$ identifies $J(t)\tilde{\theta}(t) - \dot{e}(t)$ asymptotically. Thus, substituting $\eta(t)$ from the $\dot{e}(t)$ dynamics into $\dot{\tilde{\theta}}(t)$ dynamics from (20), the $\dot{\tilde{\theta}}(t)$ dynamics can be expressed as

$$\dot{\tilde{\theta}} = g - \hat{g} - \Gamma J^T(-\dot{e} + J\tilde{\theta} - \alpha e).$$

Using Assumption 4 and applying the mean value theorem, the difference $g(\cdot) - \hat{g}(\cdot)$ can be written as

$$g(z,\theta,u) - g(z,\hat{\theta},u) = \Lambda(z,\hat{\theta},u)\tilde{\theta}(t), \qquad (32)$$

and the matrix $\Lambda(z,\hat{\theta},u)$ is bounded over all time t as

$$\bar{\Lambda} = \sup_t \left\|\Lambda(z,\hat{\theta},u)\right\| . \qquad (33)$$

The $\dot{\tilde{\theta}}(t)$ dynamics can be written as

$$\begin{aligned} \dot{\tilde{\theta}} &= (\Lambda - \Gamma J^T J)\tilde{\theta} - \Gamma J^T(-r), \\ \dot{\tilde{\theta}} &= \Pi\tilde{\theta} + \Gamma Y^T r. \end{aligned} \qquad (34)$$

The nonhomogeneous differential equation given by (34), describes a linear time varying system in $\tilde{\theta}(t)$ with a vanishing nonhomogeneous part. Consider the homogeneous part of (34)

$$\dot{e}_2 = -\Gamma J^T J e_2 + \Lambda e_2 \qquad (35)$$

where $e_2(t) \in \mathbb{R}$ is the solution of (35). Let $\Phi(t,t_0)$ be a state transition matrix of $-\Gamma J(t)^T J(t)$. From Assumption 6, there exists $a, b \in \mathbb{R}^+$ such that

$$\|\Phi(t,t_0)\| \leq a e^{-b(t-t_0)}.$$

The solution of (35) can be written as

$$e_2(t) = \Phi(t,t_0)e_2(t_0) + \int_{t_0}^{t} \Phi(t,\tau)\Lambda(\tau)e_2(\tau)d\tau. \qquad (36)$$

Using (33), the expression yields

$$\begin{aligned} e^{bt}\|e_2(t)\| &\leq a e^{bt_0}\|e_2(t_0)\| \\ &\quad + \int_{t_0}^{t} a\bar{\Lambda}(e^{b\tau}\|e_2(\tau)\|)d\tau. \end{aligned}$$

Using the Gronwall-Bellman inequality [30], the following inequality can be obtained

$$\|e_2(t)\| \leq a\|e_2(t_0)\| e^{-(b-a\bar{\Lambda})(t-t_0)}. \qquad (37)$$

Thus, $e_2(t)$ is exponentially stable. Now, consider a state transition matrix $\Phi_1(t,t_0)$ for $(\Lambda - \Gamma J^T J)$. From (37), there exists $a_1, b_1 \in \mathbb{R}^+$ such that following inequality holds

$$\|\Phi_1(t,t_0)\| \leq a_1 e^{-b_1(t-t_0)}.$$

Thus, the solution to nonhomogeneous system (34) can be written as

$$\tilde{\theta}(t) = \Phi_1(t,t_0)\tilde{\theta}(t_0) + \int_{t_0}^{t} \Phi_1(t,\tau)(\Gamma Y^T(\tau)r(\tau))d\tau. \qquad (38)$$

As the nominal system (35) is exponentially stable, Lemma 9.6 of [31] can be used along with (31) to conclude that

$$\left\|\tilde{\theta}(t)\right\| \to 0 \quad \text{as} \quad t \to \infty.$$

V. Conclusion

In this research effort, a nonlinear observer to solve SaMfM problem is developed. A solution is proposed for a particular case of feature points on the object moving with constant velocities. The approach presented in this paper does not assume minimum number of views or feature points. The assumption of feature points moving with constant velocity is valid in many practical scenarios such as range and speed estimation of vehicles moving on highways, object moving on conveyer etc.

The proposed observer cannot be used if the object is not moving with constant velocity because the expression in (14) will have unknown time varying terms. Since the model for time varying terms (i.e. model of the velocities of the feature points) is generally not known, the time varying velocities will act as non-vanishing perturbations on $\tilde{\theta}(t)$ error dynamics. Hence, it is not trivial to drive $\tilde{\theta}(t)$ error to zero. Future efforts will focus on applying the proposed observer to the real data and designing an observer that can address the case of time varying feature point velocities.

REFERENCES

[1] F. Kahl and R. Hartley, "Multiple-view geometry under the L_∞ -norm," *IEEE Transactions on Patterm Analysis and Machine Intelligence*, vol. 30, no. 9, pp. 1603–1617, Sept. 2008.

[2] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2003.

[3] J. Oliensis, "Exact two-image structure from motion," *IEEE Transactions on Patterm Analysis and Machine Intelligence*, vol. 24, no. 12, pp. 1618–2002, 2002.

[4] ——, "A critique of structure-from-motion algorithms," *Computer Vision and Image Understanding*, vol. 80, pp. 172–214, 2000.

[5] P. Sturm and B. Triggs, "A factorization based algorithm for multi-image projective structure and motion," *Lecture Notes in Computer Science*, vol. 1065, pp. 709–720, 1996.

[6] K. Sim and R. Hartley, "Recovering camera motion using l_∞ minimization," in *Computer Vision and Pattern Recognition*, vol. 1, 2006, pp. 1230–1237.

[7] J. Oliensis and R. Hartley, "Iterative extensions of the strum/triggs algorithm: convergence and nonconvergence," *IEEE Transactions on Patterm Analysis and Machine Intelligence*, vol. 29, no. 12, pp. 2217–2233, 2007.

[8] S. Soatto, R. Frezza, and P. Perona, "Motion estimation via dynamic vision," *IEEE Trans. Automat. Contr.*, vol. 41, no. 3, pp. 393–413, 1996.

[9] G. Hu, D. Aiken, S. Gupta, and W. Dixon, "Lyapunov-based range identification for paracatadioptric systems," *IEEE Transactions on Automatic Control*, vol. 53, no. 7, pp. 1775–1781, 2008.

[10] S. Soatto and P. Perona, "Reducing structure from motion: A general framework for dynamic vision, part 1: Modeling," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 9, 1998.

[11] 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.

[12] W. E. Dixon, Y. Fang, D. M. Dawson, and T. J. Flynn, "Range identification for perspective vision systems," *IEEE Trans. Automat. Contr.*, vol. 48, no. 12, pp. 2232–2238, 2003.

[13] O. Dahl, F. Nyberg, and A. Heyden, "Nonlinear and adaptive observers for perspective dynamic systems," in *American Controls Conference*, New York City, USA, July 2007, pp. 966–971.

[14] A. De Luca, G. Oriolo, and P. Giordano, "On-line estimation of feature depth for image-based visual servoing schemes," in *Proc. IEEE Int. Conf. Robotics and Automation*, 2007, pp. 2823–2828.

[15] M. Jankovic and B. Ghosh, "Visually guided ranging from observations points, lines and curves via an identifier based nonlinear observer," *Systems and Control Letters*, vol. 25, no. 1, pp. 63–73, 1995.

[16] S. Avidan and A. Shashua, "Trajectory triangulation: 3d reconstruction of moving points from a monocular image sequence," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, no. 4, pp. 348–357, Apr 2000.

[17] D. Segal and A. Shashua, "3D reconstruction from tangent-of-sight measurements of a moving object seen from a moving camera," *Lecture notes in computer science*, pp. 621–631, 2000.

[18] M. Han and T. Kanade, "Reconstruction of a scene with multiple linearly moving objects," in *CVPR - International Conference on Computer Vision and Pattern Recognition,*, vol. II, 2000, pp. 542–549.

[19] P. Sturm, "Structure and motion for dynamic scenes-the case of points moving in planes," *Lecture Notes in Computer Science*, pp. 867–882, 2002.

[20] A. Bartoli, "The geometry of dynamic scenes – On coplanar and convergent linear motions embedded in 3D static scenes," *Computer Vision and Image Understanding*, vol. 98, no. 2, pp. 223–238, 2005.

[21] Y. Ma, S. Soatto, J. Kosecká, and S. Sastry, *An Invitation to 3-D Vision*. Springer, 2004.

[22] A. D. Luca, G. Oriolo, and P. R. Giordano, "Feature depth observation for image-based visual servoing: Theory and experiments," *The International Journal of Robotics Research*, vol. 27, no. 10, pp. 1093–1116, 2008.

[23] H. K. Khalil, *Nonlinear Systems*, 2nd ed. Prentice Hall, 1996.

[24] A. M. A. Kumpati S. Narendra, *Stable Adaptive Systems*. Dover, 2005.

[25] H. Khalil, "Adaptive output feedback control of nonlinear systems represented by input-output models," *IEEE Transactions on Automatic Control*, vol. 41, no. 2, p. 177, 1996.

[26] J. Pomet and L. Praly, "Adaptive nonlinear regulation: estimation from the lyapunov equation," *IEEE Trans. Automat. Contr.*, vol. 37, no. 6, pp. 729–740, 1992.

[27] B. Xian, D. Dawson, M. de Queiroz, and J. Chen, "A continuous asymptotic tracking control strategy for uncertain nonlinear systems," *IEEE Transactions on Automatic Control*, vol. 49, no. 7, pp. 1206–1211, 2004.

[28] B. Xian, M. S. De Queiroz, and D. M. Dawson, "A continuous control mechanism for uncertain nonlinear systems. in optimal controls, stabilization, and nonsmooth analysis." *Lecture Notes in Control and Information Sciences. Heidelberg, Germany: Springer*, vol. 301, pp. 251–164, 2004.

[29] J. Slotine and W. Li, *Applied Nonlinear Control*. Prentice Hall, 1991.

[30] C. Chicone, *Ordinary Differential Equations with Applications*, 2nd ed. Springer, 2006.

[31] H. K. Khalil, *Nonlinear Systems*, 3rd ed. New Jersey: Prentice Hall, 2002.