May 1998

Toward global visual servos and estimators for rigid bodies

Noah J. Cowan
University of Michigan

Daniel E. Koditschek
University of Pennsylvania, kod@seas.upenn.edu

Copyright 1998 IEEE. Reprinted from *Proceedings of the IEEE International Conference on Robotics and Automation*, Volume 3, 1998, pages 2658-2663.

NOTE: At the time of publication, author Daniel Koditschek was affiliated with the University of Michigan. Currently, he is a faculty member in the Department of Electrical and Systems Engineering at the University of Pennsylvania.

Toward global visual servos and estimators for rigid bodies

Abstract
We describe work-in-progress toward a nonlinear image-based rigid body *dynamic triangulator* which we believe tracks a moving target from "essentially all" initial conditions (all initial conditions except a set of measure zero). The dynamic triangulator depends on the goal state only through its image plane position and velocity and requires a *navigation function*, imposed directly upon image features, to serve as a regressor for a gradient-like state update law.

NOTE: At the time of publication, author Daniel Koditschek was affiliated with the University of Michigan. Currently, he is a faculty member in the Department of Electrical and Systems Engineering at the University of Pennsylvania.

Proceedings of the 1998 IEEE
International Conference on Robotics & Automation
Leuven, Belgium • May 1998

Toward Global Visual Servos and Estimators for Rigid Bodies*

Noah J. Cowan and Daniel E. Koditschek

Electrical Engineering and Computer Science, The University of Michigan
Ann Arbor, MI 48105; E-mail: {ncowan, kod}@eecs.umich.edu

Abstract

We describe work-in-progress toward a nonlinear image-based rigid body *dynamic triangulator* which we believe tracks a moving target from "essentially all" initial conditions (all initial conditions except a set of measure zero.) The dynamic triangulator depends on the goal state only through its image plane position and velocity and requires a *navigation function*, imposed directly upon image features, to serve as a regressor for a gradient-like state update law.

1 Introduction

The control and vision literature loosely define *visual servoing* and *visual estimation* as computer-vision-based closed-loop servo control and state estimation, respectively. Sanderson and Weiss [14] propose two classifications for visual servos, *position-based*, in which the objective is to minimize a positioning error defined in the robot's Cartesian task space, and *image-based* in which the controller directly minimizes the perceived error. The same taxonomy applies to visual estimators, i.e. a position based estimator minimizes the task space tracking error and an image based estimator dynamically updates the estimate to drive the internal model to visually align with the observation.

Generically, all vision-based estimators and servos are *triangulators* in the sense that they (either explicitly or implicitly) "compute" the task space coordinates of the objects observed by cameras. Position-based systems are *algebraic triangulators* since they explicitly compute task-space information from image features and parametric knowledge of the world. Image based systems are *dynamic triangulators* since they do not require the explicit inversion of perceptual models to recover task space coordinates of the goal. Consequently, they often require less computation and are thought to be more robust with respect to calibration uncertainty [7, 8, 16, 11, 5].

*This work was supported in part by the NSF under grant IRI-9510673

When the object being tracked has rotational degrees of freedom, dynamic vision is greatly complicated. Many researchers in object tracking literature have addressed this problem by using local linearizations, e.g. Extended Kalman Filters [16, 15], which provide good results for incremental tracking but do not address the issue of large initial error.

1.1 Motivation

The long term aim of our research seeks to develop a system that couples visual estimation of a dynamical rigid body with visual servoing of a robot manipulator in order to achieve a dynamical task, such as catching

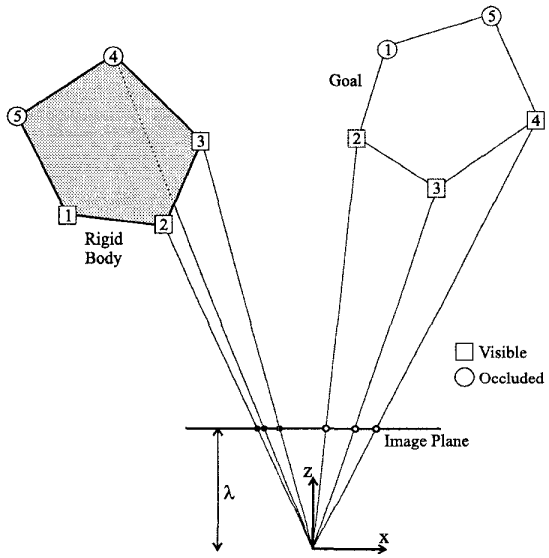

Figure 1: The objective is to drive the rigid body so that each corner aligns with the respective corner on the goal. The algorithm proposed requires three common visible corners, a condition not always satisfied. The "simple-minded" workaround is to "hallucinate" the occluded feature for the controller. We prefer to work only with the visible features available at each position, as depicted in Figure 3.

an otherwise unsensed falling body, or snatching an object from a conveyor. Our approach to such problems presupposes well designed robust "early vision" algorithms [9, 4] that track features such as corners and edges of the objects being observed. This affords the use of a growing body of signal processing algorithms designed to identify such features of an image and models the camera as a *virtual sensor* providing *image plane coordinates* for the objects being observed. Hutchinson *et. al.* provide a tutorial introduction to this approach [7].

As a rigid body moves in space, actuated or not, its corners and edges typically cycle into and out of the view of the cameras. Consequently it will often be necessary to switch the focus of attention during motion, introducing a hybrid aspect to the problem. To achieve stable systems, therefore, it is desirable to develop dynamic triangulators with very large domains of attraction in order to simplify the very challenging switching problem that inevitably results from the image-based approach.

1.2 Relation to Existing Literature

Much of the recent literature [15, 6, 16] uses local linearizations to solve tracking and servoing problems for both points and 3D objects. Some recent papers from our laboratory [13, 11] present algorithms, stability analysis and a working implementation [11] of systems with provably large domains of attraction for point positioning and estimation without local linearizations. This paper proposes extensions of that work to rigid bodies.

1.3 Organization and Contributions

The next section introduces our virtual sensor. Section 3 describes an approach to dynamic triangulation that imposes a cost function directly upon image features, and uses that cost as a regressor in a gradient-based state update. If one can show that the cost is in fact a *navigation function*,[1] then convergence to a static goal is guaranteed [12]. Similarly, using an image plane "tachometer" (Section 3.1) we might achieve asymptotic tracking of a moving target as well, subject to the extension of nonlinear time-varying stability theory to time-varying navigation functions. After presenting our triangulator we discuss the analytical properties for a specific objective function for a planar monocular camera in Section 3.2. In the planar case ($n = 2$) our nearly complete characterization of the critical points suggests that the cost function is indeed a navigation function. We also present a statistical summary of our simulation results suggesting that the servoing system is reasonably efficient. Finally we speculate on the implications of this paper and discuss future directions of our work in Section 4.

[1] A navigation function has a unique global minimum, and all other critical points are nondegenerate saddles and maximums.

2 Virtual Sensor

As stated, we wish to pose an objective function in "camera-space", $\mathcal{C}$, and therefore we must construct a virtual sensor, $c : \mathrm{SE(n)} \to \mathcal{C}$, from the camera data using knowledge of the rigid body. First we introduce some notation, and then we present the system output model.

2.1 Rigid Transformations

The group of rigid transformations, SE(n), may be embedded in $\mathrm{GL}(n+1)$ (nonsingular matrices) by writing the transformation in homogeneous representation

$$\begin{aligned} \mathrm{SE(n)} &= \left\{ H \in \mathrm{GL}(n+1) \,\middle|\, R^T R = I, |R| = 1 \right\} \\ H &= \begin{bmatrix} R & r \\ 0^T & 1 \end{bmatrix} \end{aligned} \tag{1}$$

where $R \in \mathrm{SO}(n)$ is an $n \times n$ rotation matrix and $r \in \mathbb{R}^n$ is a translation vector. Let $h : \mathbb{R}^n \to \mathbb{R}^n$ be a rigid transformation. Then, using the notation above, if $p \in \mathbb{R}^n$ then the point $b = h(p)$ is given by

$$\begin{bmatrix} b \\ 1 \end{bmatrix} = H \begin{bmatrix} p \\ 1 \end{bmatrix}$$

where H is the homogeneous representation of h.

2.2 Camera Model

The map, $\pi : \mathbb{R}^n \to \mathbb{R}^{(n-1)}$, maps a point in space to a point on the image plane. It is assumed that π takes an argument in a local camera coordinate frame. Figure 1 depicts the planar case ($n = 2$), corresponding to a "one dimensional" camera and a planar world. The specific form of π depends on the parametric camera model chosen. The pinhole camera model, reviewed in Appendix A, has lent theoretical and practical utility to previous work by our laboratory [13, 11] and we have chosen to exploit its simple structure when analyzing the specific cost function presented in Section 3.2.

Let ${}^i h$ denote the rigid transformation from world coordinates into the i^{th} camera coordinate frame. The total camera map is then given by

$$g(b) = \begin{bmatrix} \pi \circ {}^1h(b) \\ \vdots \\ \pi \circ {}^k h(b) \end{bmatrix} \tag{2}$$

where k is the total number of cameras.

2.3 System Output

Consider a rigid body and let $H \in \mathrm{SE(n)}$ denote the homogeneous representation of h, the change from body to world coordinates, i.e. if p is a point in body coordinates, then $b = h(p)$ is the same point in world coordinates. Let $P = [p_1, \ldots, p_m] \in \mathbb{R}^{n \times m}$ denote m distinguishable points ("corners") fixed on the rigid body, expressed with respect to the body reference frame. Let B denote the same set of m feature points as expressed in world coordinates, i.e.

$$\begin{bmatrix} B \\ \mathbf{1}^T \end{bmatrix} = H \begin{bmatrix} P \\ \mathbf{1}^T \end{bmatrix}$$

where $\mathbf{1} = [1, \ldots, 1]^T$. The constant matrix P is assumed known *a priori*, i.e. we know the block geometry exactly.

Let $c : \mathrm{SE(n)} \to \mathcal{C}$ denote the camera image of the m points in the rigid body, i.e.

$$c(H) := \begin{bmatrix} g(b_1) \\ \vdots \\ g(b_m) \end{bmatrix} =: \begin{bmatrix} c_1 \\ \vdots \\ c_m \end{bmatrix}, \tag{3}$$

where g is given in (2). The camera space $\mathcal{C}$ is $\mathbb{R}^{mk(n-1)}$, where k is the number of cameras.

3 Dynamic Triangulation

Suppose there is a target, whose position and orientation is given by $H^* \in \mathrm{SE(n)}$, which cannot be directly measured; instead we measure $c^* = c(H^*)$. In effect, our objective is to solve

$$c^* - c(H) = 0, \tag{4}$$

for H where the parameters of c, such as the focal length and P, are assumed known.

We distinguish two types of triangulation: dynamic and algebraic. The aim of both is to solve Equation (4). Algebraic triangulation provides a "pseudo-inverse" $c^\dagger : \mathcal{C} \to \mathrm{SE(n)}$, whereas dynamic triangulation uses an iterative method such as gradient or Newton descent to dynamically solve for the minimum of an objective function on the perceived output $c(H)$ and the perceived goal c^*. For example $\|c^* - c(H)\|$ is a candidate objective function with a global minimum at $H = H^*$. Of course, in our research agenda the recourse to dynamical triangulation is motivated in part by a real-time servo implementation wherein the descent step is executed in the physical world by the direct manipulation of the observed object. Alternatively, we might wish to obtain an asymptotic estimate of the position, orientation and velocity of an object which may be moving according to some dynamic equation.

Whether for estimation or for servoing, we posit a purely kinematic model of the form

$$\begin{aligned} \dot{H} &= u \\ y &= c(H) \end{aligned} \tag{5}$$

where $u \in T\mathrm{SE(n)}$ is the input variable.[2] In terms of local coordinates ((16) in Appendix B), we have

$$\begin{aligned} \dot{q} &= u \\ y &= c(q) \end{aligned} \tag{6}$$

where we associate with u its local coordinate representation, and by $c(q)$ it is understood that we mean $c \circ \phi_{H_0}^{-1}(q)$, although an abuse of notation. If we choose $H_0 = I$ in (16), then our local coordinates are given by $q = [\theta, r^T]^T$, the rotation and translation of the body, and $u = [\omega, v^T]^T$, the angular and translational velocity.

3.1 Generic Image-Based Tracking

We wish to triangulate a part moving on a conveyor, or a falling body, with an input dependent on the goal only through its image plane positions and velocities, and yet still guarantee convergence. To achieve this we pose the cost $\varphi : \mathrm{SE(n)} \times \mathrm{SE(n)} \to \mathbb{R}$ on image plane measurements, that is, φ admits of a factorization

$$\varphi(H, H^*) = \overline{\varphi}(c(H), c(H^*)). \tag{7}$$

Furthermore, we suppose the possibility of taking numerical derivatives of our image plane motions, and assume that we have perfect measurement of the image plane coordinates, $c^* = c(H^*)$ of the body we are tracking, and the image plane velocities, $\dot{c}^*$ (motivating the term image plane "tachometer").

Our input is given in local coordinates (16) by

$$u = -M^{-1}(q)\, D_q \varphi(q, q^*)^T - u_2 D_{c^*} \overline{\varphi}(c, c^*)\, \dot{c}^* \tag{8}$$

where

$$u_2 = D_q \varphi(q, q^*)^T (D_q \varphi(q, q^*) D_q \varphi(q, q^*)^T)^{-1}$$

and $D_x f$ denotes the Jacobian of f with respect to x. M is an arbitrary Riemannian metric.

The reader can check that u depends on $(H^*, \dot{H}^*)$ only through $(c^*, \dot{c}^*)$. Furthermore, since

$$\dot{\varphi} = D_q \varphi(q, q^*)\, u + D_{c^*} \overline{\varphi}(c, c^*)\, \dot{c}^*,$$

substituting for u from (8)

$$\dot{\varphi} = -\|\mathrm{grad}_q \varphi(q, q^*)\|_M^2 \leq 0.$$

[2] $T\mathrm{SE(n)}$ denotes the tangent space of SE(n).

If $\varphi(\cdot, H^*)$ is a navigation function and $\dot{H}^* = 0$ then we achieve asymptotic tracking for "essentially all" initial conditions [12]. When the goal is moving then $\varphi(\cdot, H^*(t))$ is time varying, and the convergence result is slightly more elusive.[3]

3.2 Navigation Function Candidate

We now investigate the critical points of a novel cost function, based on "angular error", using a planar monocular pinhole camera (see Appendix A).

Note that given the projection of two points on the image plane one can deduce the angle subtended between them (measured in the camera frame) purely from image plane coordinates, and let

$$\gamma_i = \frac{b_i^T b_i^*}{\|b_i\| \, \|b_i^*\|} = \frac{c_i^T c_i^* + \lambda^2}{\sqrt{(\|c_i\|^2 + \lambda^2)(\|c_i^*\|^2 + \lambda^2)}} \quad (9)$$

denote the cosine of the angle between the i^{th} corner and its goal, where $c = c(H)$, $c^* = c(H^*)$ and λ is the focal length. This serves as the primary building block for our cost function

$$\varphi(H, H^*) = \sum_{i=1}^{m} 1 - \gamma_i. \quad (10)$$

Note that φ has been factored according to (7). We now wish to verify that φ is a navigation function by investigating the critical points.

3.2.1 Gradient

In reporting the progress of our analysis, we restrict attention to the case in which the rigid body is a triangle ($m = 3$) on the plane ($n = 2$).

To investigate the gradient at a particular H_0, we simply compute the gradient in local coordinates (16) and evaluate at $q = 0$

$$\text{grad}_q \, \varphi(q, q^*)\big|_{q=0} = M^{-1}(0) \, (D_q \varphi(q, q^*))^T\big|_{q=0} \quad (11)$$

where M is a Riemannian metric. For convenience, let $M(0) = I$, as the choice of metric does not change the limiting behavior. Simplifying, yields

$$(D_q \varphi(q, q^*))^T_{q=0} = A \, s$$

where

$$A = \begin{bmatrix} 1 & 1 & 1 \\ \frac{Jb_1}{\|b^1\|^2} & \frac{Jb_2}{\|b^2\|^2} & \frac{Jb_3}{\|b^3\|^2} \end{bmatrix}, \quad J = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix},$$

$$s_i = \frac{\lambda(c_i^* - c_i)}{\sqrt{(\|c_i\|^2 + \lambda^2)(\|c_i^*\|^2 + \lambda^2)}}, \; i = 1, 2, 3.$$

Assuming that (4) has a unique solution in front of the camera,[4] we believe but have not yet shown formally that $\varphi(\cdot, H^*)$ has exactly two critical points. One of the critical points is the goal, which is a minimum by design. Another is the "ghost goal" behind the camera and it is a maximum, as shown in the next section. The details of this conjecture (the final proof of which is in progress) may be found in our technical report [2].

3.2.2 Hessian and Stability

The Hessian is calculated by taking the Jacobian matrix of the vector field which we may evaluate at $q = 0$ to study the stability properties at H_0. The calculations are further simplified if we evaluate the Hessian at a critical point, which implies that $s = 0$:

$$\mathbf{H} = D_q (D_q \varphi(q, q^*))^T\big|_{q=0, \, s=0}. \quad (12)$$

It is straightforward to show that

$$\mathbf{H} = A \Gamma A^T \quad (13)$$

where $\Gamma = \text{diag}\{\gamma_1, \gamma_2, \gamma_3\}$.

At the goal $\Gamma = I$ and at the "ghost goal" $\Gamma = -I$. Hence the goal is a local minimum and the remaining critical point is a maximum. Subject to the verification that there are no additional critical points, we have shown $\varphi(\cdot, H^*)$ satisfies the requirements of a navigation function. In addition to the analytical evidence, numerical simulations suggest that this is true.

3.2.3 Numerical Results

Table 1 summarizes the results of 125 simulations of (8) assuming a static goal and no occlusions, wherein 25 initial conditions were spaced evenly around each of five different initial distance balls. The maximum initial error[5] corresponds to about 100° of angular error or about three times the body radius of translational error. The camera was assumed to have unit focal length, and the rigid body is an equilateral triangle inscribed in a circle of radius 1. The goal is centered at $(0, 5)$ and was chosen as the local coordinates for the gradient calculation, i.e. $H_0 = H^*$. Of course practical settings will vary greatly in detail and the table is merely qualitative.

[3] We are presently developing a time-varying extension to navigation functions (similar to the time-varying extension to Lyapunov theory presented, for example, in Khalil [10]) that will guarantee convergence under (presumably reasonable) restrictions on H^*.

[4] Such a solution exists provided the circle which passes through the three features of the goal does not pass through the camera's focus.

[5] We have arbitrarily scaled orientation angles by body radius to fix a unique metric for SE(2).

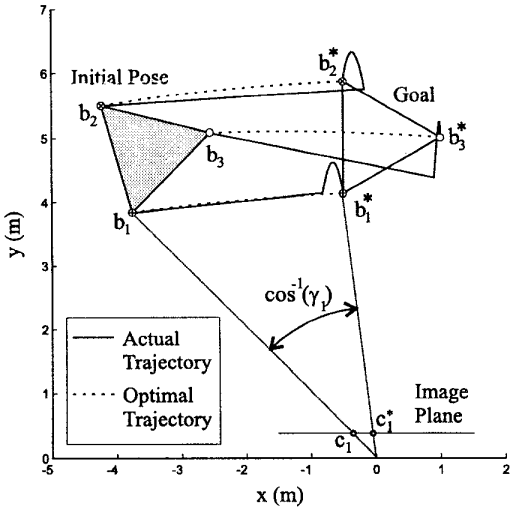

Figure 2: Typical simulated trajectory.

Distance to Goal ($\|\|q(0)\|\|$)	Avg Distance Traveled	Average Inefficiency	Standard Dev
0.25	0.37	48%	25%
0.50	0.73	46%	25%
1.00	1.44	44%	25%
1.25	1.80	41%	25%
1.50	1.39	39%	25%

Table 1: Summary of 125 simulations, wherein 25 initial conditions were spaced evenly around each of 5 different initial distances to the goal. "Average Inefficiency" is the extra distance traveled by the vision-based algorithm compared to the "straight-line" trajectory.

The algorithm seems to be reasonably "efficient" in the sense that the distance traveled is generally about one and a half times the shortest possible distance. Of course gradient algorithms do not guarantee optimality. A typical trajectory is shown in Figure 2.

4 Conclusions

4.1 Summary

The two central contributions of this paper to the field of visual servoing estimation are the application of gradient techniques to the dynamic triangulation problem for rigid bodies, including the introduction of a novel angle-based candidate navigation function, and the use of an image-based velocity estimate – image plane "tachometer" – to track nonstationary objects.

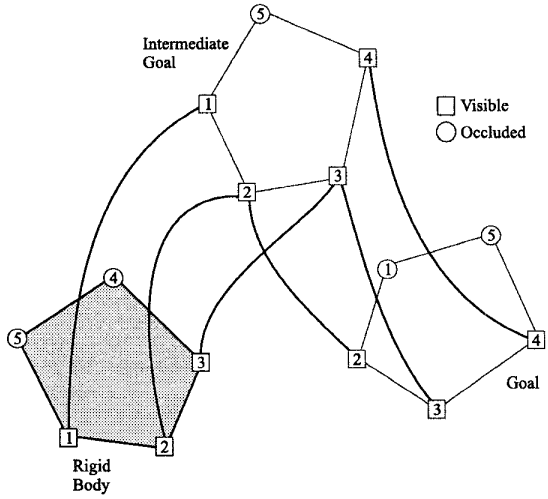

Figure 3: The camera (not shown), rigid body and goal are in the same positions as in Figure 1, and, as before, the objective is to drive the rigid body corner-wise to the goal. Here, there is a hierarchy of two controllers for backchaining. The first controller "prepares" the second.

4.2 Work-in-Progress

Our near-term research agenda addresses dynamic estimation and control in which the accelerations of the bodies being triangulated change in some known way, e.g. according to Newton's laws or a control law, so we can advance from the quasi-static approach we have used thus far.

We have also begun to address the problem of "switching." As the rigid body is servoed, it will generally be necessary to switch the focus of attention (due to occlusions), as depicted in Figure 3. Consequently a systematic switching method must be used to guarantee convergence to the goal. Back-chaining [1], wherein one constructs a set of intermediate goals which satisfy an appropriate "prepares" relationship, is one such approach. A "cartoon" version of this is shown in Figure 3. The initial condition and goal are the same as in Figure 1, however there is an intermediate goal, H' constructed *a priori* that shows corners $1, 2, 3, 4$. One first navigates according to $\varphi(\cdot, H')$, to the *intermediate* goal until corner 4 is visible, at which time the controller switches to the navigation function $\varphi(\cdot, H^*)$. In this sense, the first navigation function "prepares" the second. Of course one seeks to plan out a complete set of these intermediate "preparatory" subgoals in such a fashion that the entire configuration space, SE(2), is "covered".

Finally, and most importantly, we have begun constructing research apparatus for validating the empir-

ical utility of the analytical results and conjectures set forth in this paper.

A Pinhole Camera

The spatial pinhole camera transformation $\pi_\lambda : \mathbb{R}^3 \to \mathbb{R}^2$ is given by

$$\pi_\lambda(b) := \frac{\lambda}{b_z} \begin{bmatrix} b_x \\ b_y \end{bmatrix}, \tag{14}$$

where λ is the focal length (assumed known) of the camera and $b = [b_x, b_y, b_z]^T$ is a point specified in camera coordinates. Note that the z-axis is perpendicular to the image plane.

For the planar case, SE(2), there are only two frame axes which we denote $\{x, z\}$. The z-axis is chosen orthogonal to the image line, as before.[6] See Figure 1. The planar pinhole camera map $\pi_\lambda : \mathbb{R}^2 \to \mathbb{R}$ is simply

$$\pi_\lambda(b) = \frac{\lambda}{b_z}[b_x]. \tag{15}$$

Note that the same symbol, π_λ is used for both the planar and spatial camera. It will be clear from context which is being used.

B Local Coordinates on SE(n)

The local coordinate charts used in this paper were chosen to simplify the analysis, and are defined below.

The skew symmetric operator $J : \mathbb{R}^m \to \text{Skew}(m)$ is given by

$$J(v) := \begin{bmatrix} 0 & -v_3 & v_2 \\ v_3 & 0 & -v_1 \\ -v_2 & v_1 & 0 \end{bmatrix}.$$

for $m = 3$ and

$$J(v) := \begin{bmatrix} 0 & -v \\ v & 0 \end{bmatrix}$$

for $m = 1$. Let $q_1 \in \mathbb{R}^{n(n-1)/2}$ and $q_2 \in \mathbb{R}^n$. Define the map $\phi_{H_0}^{-1} : \mathbb{R}^{n(n+1)/2} \to \text{SE(n)}$ by

$$\phi_{H_0}^{-1}(q) := \begin{bmatrix} \exp(J(q_1)) & q_2 \\ 0^T & 1 \end{bmatrix} H_0. \tag{16}$$

In a neighborhood of H_0, $\phi_{H_0}^{-1}$ is invertible. In particular, $\phi_{H_0} \circ \phi_{H_0}^{-1}$ is the identity on $\mathbb{R}^{n(n+1)/2}$ in a neighborhood of 0, and $\phi_{H_0}^{-1} \circ \phi_{H_0}$ is the identity on SE(n) in a neighborhood of H_0 [3].

[6] Since the "world" is chosen to be a plane, the camera image is one dimensional.

References

[1] R. R. Burridge, A. A. Rizzi, and D. E. Koditschek. Sequential composition of dynamically dexterous robot behaviors. *Int. J. Rob. Res.*, (to appear).

[2] N. J. Cowan and D. E. Koditschek. Visual servos and estimators for rigid bodies. Technical Report CGR 98-07, University of Michigan, 1998.

[3] M. Curtis. *Matrix Groups.* Springer Verlag, New York, 1970.

[4] G. D. Hagar. Xvision visual tracking software, 1996.

[5] G. D. Hager. Calibration-free visual control using projective invariance. In *Proceedings of 5th ICCV*, 1995.

[6] K. Hashimoto, T. Elbine, and H. Kimura. Visual servoing with hand-eye manipulator- optimal control approach. *IEEE Transactions on Robotics and Automation*, pages 651–670, October 1996.

[7] S. Hutchinson, G. D. Hager, and P. I. Corke. A tutorial on visual servo control. *IEEE Transactions on Robotics and Automation*, pages 651–670, October 1996.

[8] M. Jägersand, O. Fuentes, and R. Nelson. Experimental evaluation of uncalibrated visual servoing for precision manipulation. In *International Conference on Robotics and Automation*, Albuquerque, NM, April 1997. IEEE.

[9] R. Jain, R. Kasturi, and B. Schunck. *Machine Vision.* McGraw-Hill, Inc., 1995.

[10] H. K. Khalil. *Nonlinear Systems.* Prentice Hall, 1996.

[11] D. Kim, A. A. Rizzi, G. D. Hager, and D. E. Koditschek. A "robust" convergent visual servoing system. In *International Conf. on Intelligent Robots and Systems*, Pittsburgh, PA, 1995. IEEE/RSJ.

[12] E. Rimon and D. E. Koditschek. Exact robot navigation using artificial potential fields. *IEEE Transactions on Robotics and Automation*, 8(5):501–518, Oct 1992.

[13] A. A. Rizzi and D. E. Koditsckek. An active visual estimator for dexterous manipulation. *IEEE Transactions on Robotics and Automation*, pages 697–713, October 1996.

[14] A. C. Sanderson and L. E. Weiss. Image-based visual servo control using relational graph error signals. In *Proceedings of the IEEE*, pages 1074–1077. IEEE, 1980.

[15] S. Soatto, R. Frezza, and P. Perona. Motion estimation via dynamic vision. *IEEE Transactions on Automatic Control*, 41(3):393–413, March 1996.

[16] P. Wunsch and G. Hirzinger. Real-time visual tracking of 3-d objects with dynamic handling of occlusions. In *International Conference on Robotics and Automation*, pages 2868–2873, Albuquerque, NM, 1997. IEEE.