Binocular Visual Servoing with a Limited Field of View

Noah J. Cowan

University of California at Berkeley

3060 VLSB, #3140

Berkeley, California 94720-3140

`http://www.cs.berkeley.edu/~ncowan/`

Abstract

Geometrically characterizing a set of visible configurations — those for which a suitable set of features can be seen by a camera system — and its image under perspective projection provides an effective means by which to design global visual servoing systems. This paper presents a simple, globally convergent, binocular image-based visual servoing algorithm which accounts for the limited field-of-view of a CCD camera system. Recourse to Navigation Functions, a refined notion of artificial potential functions, leads to controllers for first order, quasi-static as well as second order, dynamic plants. Underlying these constructions is a global account of the transformation between the task and image space.

1 Introduction

The field of visual servoing, robot control in which a computer vision system serves as the primary feedback sensor, has enjoyed relatively slow but steady progress over the last fifteen or so years. Early ideas in vision based robotics and control predate the mid-eighties, of course, but advances in computing technology and vision equipment during recent years has lead to an increased interest in visual servoing. Simple visual position regulation, for which a suitable set of visible features contrast well their surroundings, has more-or-less been solved: a wide body of literature provides the necessary tools to design high performance, dynamic (albeit local) vision-based controllers [2]. This complete characterization is owed primarily to the local nature of such regulation problems. However, the creation of a richer set of behaviors whose *global* (and hence highly nonlinear) properties are well characterized remains a significant challenge to visual servoing.

The classical approach to visual servoing (for a tutorial, see [10]) attempts to impose straight-line trajectories on image feature points. To illustrate, suppose a camera observes four feature points affixed to a rigid body moving in space. The configuration space of

1

the rigid body may be parameterized locally by six numbers, $q \in \mathbb{R}^6$. These numbers may represent, for example, the joint angles of a six DOF robot, or the three translational degrees of freedom and the three Euler angles. The image space may be defined in terms of the four image plane pairs, $y \in \mathbb{R}^8$. Locally, then, the camera's image may be modeled by a map $c : \mathbb{R}^6 \to \mathbb{R}^8$

$$y = c(q) = \begin{bmatrix} u_1 & v_1 & \cdots & u_4 & v_4 \end{bmatrix}^T, \quad \text{where} \quad \begin{bmatrix} u_i \\ v_i \end{bmatrix}, \quad i = 1, \ldots, 4,$$

are the image plane feature locations of the four features being observed. The traditional (kinematic) visual servoing law is then

$$\dot{q} = -J^\dagger (y - y^*), \quad \text{where} \quad J^\dagger = (J^T J)^{-1} J^T \in \mathbb{R}^{6 \times 8} \tag{1.1}$$

is a pseudo inverse of the Jacobian matrix, $[J]_{i,j} = \frac{\partial y_i}{\partial q_j}$.

The classical approach to visual servoing works fine if the initial configuration is sufficiently close to the goal. However, it is well known that the above approach can lead to local minima and often exhibits very poor task-space trajectories. A particularly curious such behavior arises if the body starts 'upside-down' relative to its visual goal. In this case, the straight-line image trajectories force the body out along the camera z-axis to infinity! This observation was first reported by Chaumette [1], and has since been dubbed the 'Chaumette Conundrum' [3]. Indeed such examples are not mysterious at all, but a mere consequence of failing to characterize the relationship between the domain and range of the camera transformation from which they arise.

1.1 Three basic challenges to visual servoing

Several challenges inherent to visual servoing are described in this section. These challenges leave many opportunities to advance the field.

Challenge #1: Global convergence via smooth feedback is technically *impossible*

Designing feedback systems with large domains of attraction represents a challenge. The reason for this in the context of visual servoing is that the geometry of rigid motion is complicated by perspective projection, and many subtle and counter intuitive problems arise.

Recent visual servoing literature, reviewed in Section 1.4, has generated successful servoing strategies which exhibit large domains of attraction, avoid FOV boundaries and confer robustness to parametric uncertainty. Why have these innovations thus far been few and far between? What ties together the successful solutions, and how can we extend the constructions to new circumstances? Is there a way to demystify the construction of visual servoing systems?

Issues of convergence may be greatly illuminated by understanding the geometry of and transformations between the underlying spaces involved. Consider again the traditional

visual feedback given by (1.1), and observe that it can be cast into the framework of gradient decent on the potential function $\varphi : SE(3) \to \mathbb{R}$, given by

$$\varphi(y) = \frac{1}{2} \sum_{i=1}^{4} \|y_i - y_i^*\|^2, \tag{1.2}$$

Now, let $\dot{q} = -M\nabla_q(\varphi \circ c)$, where $M(q) = (J^T J)^{-1}$ is presumed to exist and be positive definite (at least in some neighborhood of the goal). It is a well known fact from Morse theory that a smooth, scalar valued function on SE(3) must have at least four critical points, hence it is now clear that global visual servoing based on smooth potential functions is actually *impossible.* In fact, any smooth vector field on SE(3) must have extraneous critical points. These facts hold for first order, kinematic systems and second order Lagrangian systems. Hence *global rigid body control via smooth feedback is not possible!* However, we may be able to insure that almost every initial configuration converges – that is, all but a set of measure zero.

Chaumette Conundrum revisited. The camera transformation, c, can be decomposed into the rigid kinematics which moves feature points around in the FOV, followed by the perspective projection of the feature points onto the compact image. Notice that perspective projection compactifies the entire FOV: all points along a ray through the camera pinhole are projected to the same point on the closed and bounded image plane. By moving a rigid body along one of these rays to infinity, two things happen. First, the feature point projections converge to a single point. Second, the Jacobian becomes increasingly singular due to the well known dependence of the Jabobian on inverse depth. The author conjectures that these two facts lead to a 'saddle at infinity' in the cost function (1.2). Empirically, this has been supported by the literature – only perfectly 'upside-down' configurations (a set of measure zero) diverge to this saddle point – though it remains to be verified analyitically. Initial conditions that are 179° — i.e. almost upside-down — will converge, though they exhibit qualitatively bad trajectories, possibly because they begin near the stable manifold of the saddle point.

Challenge #2: Occlusions are virtual obstacles to visual servoing

The use of cameras for sensing imposes certain physical limitations. For example, off-the-shelf cameras have a limited field of view (FOV). Occlusions and FOV limitations represent virtual obstacles which should be accounted for to create successful controllers.

The challenges bestowed upon us due to occlusions and FOV limitations are fairly intuitively. The question is, how do we create controllers which help us to deal with them? This paper, and previous work by the author and colleagues [4, 5, 6, 7], aims to formally define a set of configurations that are 'visible', and then understand how that set projects to image coordinates so that it can be appropriately navigated.

Challenge #3: Dynamics limit performance

In real robot control systems, kinematic trajectories must be enforced by robot controllers that can manage a robot's kinetic energy. Treating the visual servoing law as a 'reference generator' enables us to separate the planning problem from the control problem, as is standard in the robotics community. However, sweeping the issue of dynamics under the carpet requires the use of sophisticated robot control techniques that require precise parametric knowledge of the robot's kinematics and dynamics, such as adaptive inverse dynamics based control strategies [22]. This extra complexity seems superfluous given the simple end-point convergence objective of most visual servoing algorithms. In other words, the specific reference trajectories generated by kinematic controllers are merely a means to achieve an end-point task; tracking those trajectories exactly may not be necessary. Such heavy reliance on robot kinematic and dynamic parameters may be undesirable, especially when manipulating objects of imprecisely known size and weight. Can we design *dynamical* vision-based controllers that achieve an end-point objective?

1.2 Overcoming the challenges: safe dynamical convergence

Although the geometry may require multiple critical points, we can design our potential function with only one attracting critical point, thus insuring all initial configurations except a set of zero measure will converge to the global minimum. Moreover, we can avoid the poor trajectories that might arise from an incomplete characterization of the critical points as in Chaumette's example. Finally, one probably would like to keep configurations within the FOV, and within a reasonable range of distances to the camera system, cutting out a significant portion of the configuration space. This may change the topology of our free configuration space, and thus change the constraints on the number of critical points that must be present.

By restricting the subset of configurations to a compact subset of the configuration space, $\mathcal{D} \subset \mathcal{Q}$, and designing a so-called Navigation Function (NF) $\varphi : \mathcal{D} \to \mathbb{R}$ on this domain, as the author and collegues have shown [4, 5, 6, 7], essentially global convergence can be achieved — that is, all initial conditions except a set of measure zero converge. Moreover, the approach generates controllers capable of extremely high performance, which exhibit dynamical convergence to the end-point goal, without the burden of precisely tracking a (clearly somewhat arbitrary) kinematic reference trajectory. Without prescribing specific reference trajectories, the proposed methodology nevertheless affords certain guarantees on the trajectories that result. For example, features are guaranteed to remain visible throughout transients (even in the presence of Newtonian dynamics).

The prior work thus far applies only to *monocular* visual servoing systems, yielding feedback controllers that are

1. *dynamic:* applicable to second order (Lagrangian) as well as first order (kinematic) actuation models;

2. *global:* guaranteeing a basin of attraction encompassing essentially the entire set of "safe" initial configurations that maintain feature visibility;

3. *visibility-obstacle free:* avoiding configurations that lose features due to either self occlusion or departure from the FOV.

One of the key contributions in the prior work is that visual servoing systems need not appear mystical: it is possible to model the set of visible configurations of a robot (those for which a suitable set of features are visible to the camera system) and formally describe how this visible set is distorted by perspective projection into a subspace of image-feature locations. Then, a set of obstacles and goals can be designed that, in a sense, respect the transformation between task and image coordinates. It is then our challenge to define controllers stay within the free space while converging to the image-encoded goal location.

1.3 Contribution and Organization

This paper adheres to the same general design philosophy of the prior work: by characterizing the space of visible configurations, and its image on a camera system, one can create a navigation function (NF) based controller that ensures that a visual goal is achieved while maintaining feature visibility. (The method of NFs is very briefly sketched in Appendix 4.)

The present paper seeks to extend the methodology of earlier work to address *binocular* servoing. In particular, Section 2 presents the construction of NF's for two very specific and simple examples of binocular systems in a 2D world: simple "base-line disparity" (Section 2.1) and a case in which the optical axes are mutually orthogonal (Section 2.2). In Section 2.3 we discuss how to generalize these special cases to more general camera placement. First, we review prior contributions to the field of global visual servoing.

1.4 Relation to existing literature

A first step in visual servoing is understanding what can be verified with a camera system, i.e. what tasks can be encoded in image coordinates? In the context of binocular vision, this problem has largely been answered [9]. In the present paper, the task is geometrically trivial — move a point to a desired position in image coordinates — and hence encoding the task in image coordinates is trival. The present challenge lies in achieving dynamical convergence while avoiding visibility obstacles, which we achieve by recourse to NF's. It is worth noting that an NF is a functional that encodes the task. Of course, simpler functionals could be designed to encode the task, but the NF is employed explicitly to create a global, dynamical controller.

Zhang and Ostrowski [23] developed a controller for an Unpiloted Aerial Vehicle (UAV) equipped with a camera for which they exhibit a diffeomorphism between the centroid and radius of a circle in the image plane and the position and orientation of the UAV relative to a

sphere in the work space. Formulating the dynamics in generalized image-plane coordinates leads to a feedback linearized controller that accounts for the mechanical system dynamics.

In addition to our preliminary results [4, 5, 6, 7], there have been some recent efforts to address the problems of convergence and feature visibility, albeit in a quasi-static setting. For 6DOF visual servoing, Malis *et. al.* [16] guarantee that a single base point remains within the FOV while, as noted above, guaranteeing convergence for a large basin of attraction. Morel *et. al.* [18] extend this idea by decoupling the image-plane motion of a cleverly chosen feature vector – a circle containing all the feature points – from the rotational motion of the camera; by keeping this conservative feature vector within the image plane, one is guaranteed that all feature points remain within the FOV (though self-occlusion avoidance was not guaranteed). Corke and Hutchinson [3] have designed a clever "partitioned" kinematic visual servo strategy for which simulations suggest a large basin of attraction while maintaining all features within the FOV boundary. Mezouar *et. al.* [17] have adopted the approach of image-based path planning and local visual servoing along those paths to avoid mechanical limits and visibility obstacles.

Potential field methods have been employed for a wide variety of robot navigation problems (for a survey, see [15], Chapter 7). Traditionally, gradient fields have been used to generate reference trajectories which are then tracked by a lower-level robot controller. Potential functions often encode obstacles as local maxima, or at least ensure that the gradient flow does not intersect obstacles. The refinement to NF's, first articulated by Koditschek and Rimon [13, 14, 20], provides machinery to "lift" potential functions to second order plants, while still ensuring obstacle avoidance with convergence guarantees, and no need for intermediate trajectories.

"Natural" coordinates for visual servoing have not yet emerged in any generality. In the example applications to follow, we identify in each case a model space for the unoccluded configurations, construct a solution for the model, and furnish a change of coordinates that pulls the model solution back into terms of the physically measured features. However, we are not yet able to characterize the general properties of the features that would make such constructions possible. A promising approach relies on projective kinematics [21], although it has not been validated empirically and currently lacks the machinery required to lift it to the second order setting or maintain feature visiblity.

2   Planar Binocular Visual Servoing

The world is taken to be two dimensional Euclidean space, $\mathbb{E}^2$, and the cameras are represented by pinhole projection, $\pi : \{\mathbf{x} \in \mathbb{E}^2 : x_2 > 0\} \to \mathbb{R}$,

$$\pi(\mathbf{x}) = \frac{1}{\lambda} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad \text{where} \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ 1 \end{bmatrix} \quad \text{and} \quad \mathbf{0} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}. \tag{2.3}$$

6

We assume for convenience that the focal length, $\lambda = 1$, for the remainder of the paper. The 1D camera model (2.3) is applicable, for example, to mobile robots translating and rotating in the horizontal plane in an indoor structured environment for which the horizontal location of vertical edge features is taken as the image measurement from a standard 2D CCD camera. Points in the base coordinate system are transformed into the right and left camera frames by

$$^{L}H_0 = \begin{bmatrix} ^{L}R_0 & ^{L}T_0 \\ 0 \quad 0 & 1 \end{bmatrix} \quad \text{and} \quad ^{R}H_0 = \begin{bmatrix} ^{R}R_0 & ^{R}T_0 \\ 0 \quad 0 & 1 \end{bmatrix} \tag{2.4}$$

respectively. The total camera map, defined only for points in front of both cameras, is given by

$$c : \{\mathbf{x} \in \mathbb{E}^2 : {}^{L}x_2, {}^{R}x_2 > 0\} \rightarrow \mathbb{R}^2 \tag{2.5}$$

$$c : \mathbf{x} \qquad\qquad \mapsto \begin{bmatrix} \pi({}^{L}\mathbf{x}) \\ \pi({}^{R}\mathbf{x}) \end{bmatrix} \tag{2.6}$$

$$\text{where} \quad {}^{L}\mathbf{x} := {}^{L}H_0{}^{0}\mathbf{x} \quad \text{and} {}^{R}\mathbf{x} := {}^{R}H_0{}^{0}\mathbf{x} \tag{2.7}$$

where $^{0}\mathbf{x}$ is $\mathbf{x}$ expressed in base coordinates.

This paper considers a camera model which accounts for the limited FOV present in CCD cameras, modeled by a minimum and maximum value on the image plane between which our measurements must be bounded to ensure visibility. The *visible set*, $\mathcal{V}$, is thus defined for binocular vision to be all those configurations in front of both cameras such that the projection is on both image planes. Formally,

$$\mathcal{V} = \left\{ \mathbf{x} \in \mathbb{E}^2 : {}^{L}x_2, {}^{R}x_2 > 0 \text{ and } y_1, y_2 \in [y_{\min}, y_{\max}] \right\}, \tag{2.8}$$

$$\text{where} \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = c(\mathbf{x}). \tag{2.9}$$

Denote the image of the visible set

$$\mathcal{J} := c(\mathcal{V}) \subset [y_{\min}, y_{\max}] \times [y_{\min}, y_{\max}]. \tag{2.10}$$

The objective of our visual servoing problem is to find a control law for which $\mathbf{x} \rightarrow \mathbf{x}^*$ as $t \rightarrow \infty$, using the image measurement, $\mathbf{y}^* = c(\mathbf{x}^*)$, of our goal.

We now consider two special cases of binocular visual servoing. For case 1 (Section 2.1), we assume a simple "base-line disparity" between the cameras as shown in Figure 1 so the camera optical axes are parallel. In the second case (Section 2.2), the optical axes are perpendicular, as shown in Figure 2, which is generalized in Section 2.3.
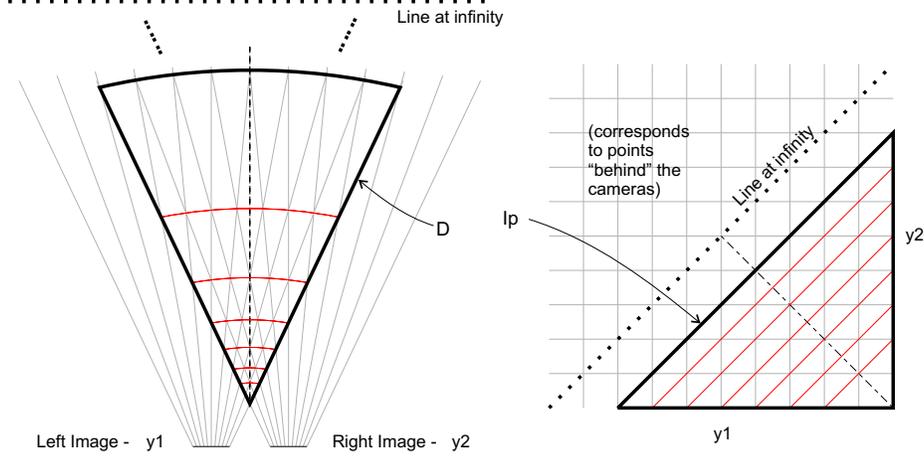
Figure 1: **Left:** Two cameras with a base-line disparity. The cameras are assumed to have a finite FOV, depicted by the extent of the gridlines radiating from each image plane. **Right:** The image space $[y_{\min}, y_{\max}] \times [y_{\min}, y_{\max}]$ is the cross product of the two image plane measurements. The line $y_1 = y_2$ is the projection of the "line at infinity". The *visible set*, $\mathcal{V}$, of configurations consists of those points which may be seen simultaneously by both cameras, and $\mathcal{D} \subset \mathcal{V}$ is a compact subset. The image $\mathcal{I}_\rho = c(\mathcal{D})$ under perspective projection is a compact solid triangle.

2.1   Case 1: Base-line disparity

For this case, depicted in Figure 1, both cameras face the same direction, seperated by a distance δ, *i.e.*

$$^{L}H_0 = \begin{bmatrix} 1 & 0 & \delta/2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad \text{and} \quad {}^{R}H_0 = \begin{bmatrix} 1 & 0 & -\delta/2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{2.11}$$

with the base frame taken to be between the cameras. The visible set is an unbounded cone, and its image is the "triangle" $c(\mathcal{V}) = \{y_{\min} \leq y_1 < y_2 \leq y_{\max}\}$. It is interesting to note that the set $\{y_{\min} \leq y_2 < y_1 \leq y_{\max}\}$ corresponds to Euclidean point locations behind the image plane pair, and the line $y_2 = y_1$ corresponds to the "line at infinity" depicted in Figure 1. Neither the line at infinity nor the points behind the camerea are part of the visible set. The camera map reduces to

$$c(\mathbf{x}) = \begin{bmatrix} \frac{x_1 - \delta/2}{x_2} & \frac{x_1 + \delta/2}{x_2} \end{bmatrix}^T. \tag{2.12}$$

It is well known that given the projection $\mathbf{y}$ of a point on a binocular image plane pair, one can compute the unique location $\mathbf{x} = c^{-1}(\mathbf{y})$ of the point in front of the camera which corresponds to that projection. Moreover, the mapping c and its inverse c^{-1} are smooth, *i.e.*

$$D_{\mathbf{x}}c(\mathbf{x}) = \frac{1}{x_2^2} \begin{bmatrix} x_2 & -x_1 + \delta/2 \\ x_2 & -x_1 - \delta/2 \end{bmatrix} \quad \Longrightarrow \quad |D_{\mathbf{x}}c(\mathbf{x})| = \delta/(x_2^3) \tag{2.13}$$

and hence the Jacobian matrix and its inverse are well defined over the visible set. The camera map is, therefore, a diffeomorphism from $\mathcal{V}$ to its image.

Because c is a diffeomorphism, we may treat the measurement vector $\mathbf{y} = c(\mathbf{x})$ as a set of generalized coordinates for the motion of the point $\mathbf{x}$. Therefore, an NF, $\widetilde{\varphi}$, imposed on $\mathcal{J}$ may be pulled back into the task space through c and, more importantly, the gradient may be pulled back through the transposed Jacobian of c, *i.e.*

$$\varphi = \widetilde{\varphi} \circ c \quad \implies \quad \nabla \varphi = (Dc)^T \nabla \widetilde{\varphi}. \tag{2.14}$$

This provides a great simplification to the problem of remaining in the FOV of both cameras, by reducing the problem to that of designing an NF on image coordinates which renders the FOV boundary as an obstacle. As shown in Appendix 4, the NF can then be used to generate convergent first and second order controllers.

Before designing an NF, a few subtle issues remain. First, notice that $\mathcal{V}$ is not compact. However our domain must be compact (see Appendix 4). Fortunately, we may provide an additional border in the image plane to construct a compact domain $\mathcal{D} \subset \mathcal{V}$. In other words, we let $\mathcal{D} = c^{-1}(\mathcal{I}_\rho)$, where $\mathcal{I}_\rho$ is the compact set

$$
\begin{aligned}
\mathcal{I}_\rho &= \overbrace{\{y_{\min} \leq y_1 < y_2 \leq y_{\max}\}}^{\mathcal{J}} \cap \overbrace{\{y_2 - y_1 \geq \rho\}}^{\text{safety collar}} \\
&= \{y_{\min} \leq y_1, y_2 - y_1 \geq \rho, y_2 \leq y_{\max}\} \subset \mathcal{J}, \quad \text{where } \rho > 0. \tag{2.15}
\end{aligned}
$$

The set $\mathcal{D}$ and its image $\mathcal{I}_\rho = c(\mathcal{D})$, are depicted in Figure 1. A natural consequence of using image coordinates is that the FOV shows up in a simple way.

Proposition 2.1. *(Adapted from [12]) The objective function*

$$\overline{\varphi}(\mathbf{y}) = \frac{\|\mathbf{y} - \mathbf{y}^*\|^{2k}}{\prod_{i=0}^n (y_{i+1} - y_i)^2 - \rho_i^2}$$

is convex on

$$\mathcal{I}_\rho = \{y \in \mathbb{R}^n : y_{i+1} - y_i \geq \rho_i, \; i = 0, \ldots, n\}$$

where y_0, y_{n+1} and $\rho_i, i = 0, \ldots, n$ are suitable constants and $k > (2n+3)/2$ in which case

$$\widetilde{\varphi} := \frac{\overline{\varphi}^{1/k}}{(1 + \overline{\varphi})^{1/k}} \tag{2.16}$$

is a navigation function on $\mathcal{I}_\rho$.

For planar servoing, $n = 2$ and so we require $k > 7/2$, hence

$$\overline{\varphi}(y) = \frac{\|\mathbf{y} - \mathbf{y}^*\|^{2k}}{(y_1 - y_{\min})^2 \left[(y_2 - y_1)^2 - \rho^2\right] (y_{\max} - y_2)^2} \tag{2.17}$$

is convex. Hence $\widetilde{\varphi}$ given by (2.16) is an NF on $\mathcal{I}_\rho$ and, since NF's are invariant under diffeomorphism, $\varphi := \widetilde{\varphi} \circ c$ is an NF on $\mathcal{D}$.
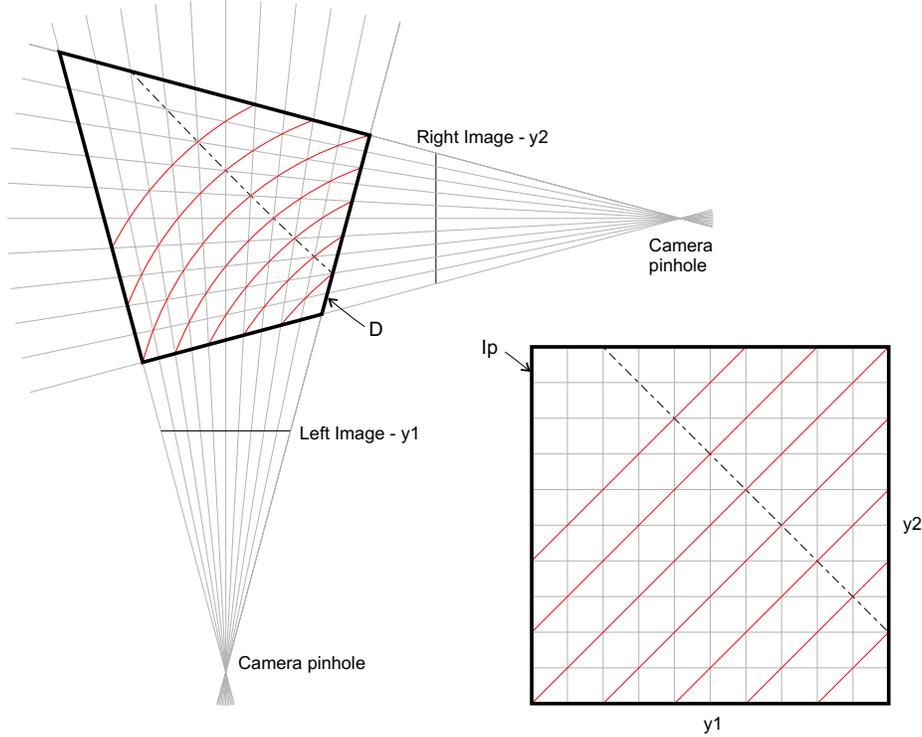
Figure 2: **Left:** Two cameras with orthogonal optical axes. It is assumed that the two FOV's intersect in a parallelogram as depicted. **Right:** The image space is the cross product of the two image plane measurements. In this case, $\mathcal{D} = \mathcal{V}$ is compact, and $\mathcal{I} = c(\mathcal{D}) = [y_{\min}, y_{\max}] \times [y_{\min}, y_{\max}]$.

2.2 Case 2: Orthogonal optical axes

In this section we consider the camera configuration depicted in Figure 2. For simplicity, we position the world frame coincident with the left camera, and

$$
{}^{L}H_0 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad {}^{R}H_0 = \begin{bmatrix} 0 & 1 & -\delta \\ -1 & 0 & \delta \\ 0 & 0 & 1 \end{bmatrix} \tag{2.18}
$$

and hence

$$
c(\mathbf{x}) = \begin{bmatrix} \frac{x_1}{x_2} \\ \frac{-\delta + x_2}{+\delta - x_1} \end{bmatrix} \implies D_{\mathbf{x}} c(\mathbf{x}) = \begin{bmatrix} \frac{1}{x_2} & -\frac{x_1}{x_2^2} \\ \frac{x_2 - \delta}{(\delta - x_1)^2} & \frac{1}{\delta - x_1} \end{bmatrix} \tag{2.19}
$$

$$
\implies |D_{\mathbf{x}} c(\mathbf{x})| = \frac{\delta(x_2 - x_1)}{x_2^2 (x_1 - \delta)^2}. \tag{2.20}
$$

For this camera arrangement, there are some possible singularities to the camera map. If the FOV of each camera subtends an angle greater than $\pm 45^o$, then the line $x_2 = x_1$ will lie within the FOV of both cameras, and on that line the Jacobian becomes singular. Points along

the line $x_2 = x_1$ cannot be triangulated, as they all correspond to an image measurement of $\mathbf{y} = [1, -1]^T$. (In a more general camera placement, points along the line connecting the two camera pinholes are not resolvable). We make the assumption for this simple example that $y_{\min} > -1$ and $y_{\max} < 1$. In this case, the FOV's intersect as shown in Figure 2, and there are no singularities in $\mathcal{V}$.

In the orthogonal case, the geometry is a bit simpler. We may take $\mathcal{D} = \mathcal{V}$, and hence $\mathcal{I} = \mathcal{J}$. For simplicity, assume that we have rescaled the image plane so that $-y_{\min} = y_{\max} = \frac{1}{2}$. A navigation function on $\mathcal{I}$ is given as follows. Consider the function

$$f(\mathbf{y}) = \left[\frac{y_1 - y_1^*}{\left(\left(\frac{1}{2} \right)^2 - y_1^2 \right)^{\frac{1}{2}}} \quad \frac{y_2 - y_2^*}{\left(\left(\frac{1}{2} \right)^2 - y_2^2 \right)^{\frac{1}{2}}} \right]^T \quad \text{and} \quad \overline{\varphi} := \tfrac{1}{2} f^T K f$$

where $K \in \mathbb{R}^{2 \times 2}$ is a positive definite symmetric matrix. It is routine to verify that the function

$$\widetilde{\varphi} := \frac{\overline{\varphi}}{1 + \overline{\varphi}} \tag{2.21}$$

is a navigation function on $\mathcal{I}$, and hence $\varphi = \widetilde{\varphi} \circ c$ is a navigation function on $\mathcal{D}$.

2.3 Toward General Camera Placement

The two example systems discussed above, parallel cameras and orthogonal cameras, are special arrangements of the camera pair, designed to simplify the mathematics. However, it would be comforting to know that the results may be applied to a more general setting, lest they serve no practical value.

Fortunately, for very little additional effort we may generalize the orthogonal configuration of Section 2.2 to an open subset of the configuration space, by observing that sufficiently small changes in the location of the right camera w.r.t. the left camera, ${}^R H_0$, and in the FOV limits $(y_{\min}, y_{\max})$ for each camera, still results in a quadrilateral visible set, $\mathcal{V}$, and, more importantly $\mathcal{J} = c(\mathcal{V}) = [y_{\min}, y_{\max}] \times [y_{\min}, y_{\max}]$ does not change. Moreover, the camera map remains a diffeomorphism onto its image for small analytic perturbations of the parameters. Hence the NF in (2.21) applies by simply rescaling the image plane to $[-\frac{1}{2}, \frac{1}{2}] \times [-\frac{1}{2}, \frac{1}{2}]$. We conclude with the following conjecture, proof of which is in progress.

Conjecture 2.1. *If the FOV's intersect in a compact quadrilateral with boundary to form the visible set $\mathcal{V}$, then its image, $\mathcal{J} = c(\mathcal{V}) = [y_{\min}, y_{\max}] \times [y_{\min}, y_{\max}]$. Moreover, $\widetilde{\varphi}$ (2.21) is an NF on $\mathcal{J}$, and hence $\varphi = \widetilde{\varphi} \circ c$ is an NF on $\mathcal{V}$.*

3 Conclusions

This paper presents a very simple approach to binocular image-based visual servoing. The approach yields dynamic, convergent servoing algorithms which respect the boundaries of the FOV. To the best of the author's knowledge, there are no prior visual servoing algorithms

for a binocular rig which guarantee convergence to a goal while avoiding FOV boundaries, much less in a dynamic setting.

Notwithstanding the improvements that this paper suggests may be possible over traditional algorithms, there is much left to do. Generalization to 3D settings seems is a high priority, as is handing of rotational degrees of freedom. A challenge with two spatial cameras lies in the fact that each image provides two numbers, so that a 3 DOF point in space is represented by four numbers. The author believes that incorporating the epipolar geometry will lead to a 3 DOF image based coordinate system, so that the algorithms present in this paper may be easily generalized to 3D.

Application to new settings requires the construction of specific transformations and coordinate systems for each new setup. Nevertheless, the basic approach of geometrically characterizing the visible configurations and their image under projection seems an effective way generate formally valid, globally convergent, dynamical visual servoing systems.

4 Robot control via navigation functions (NF's)

Assume we have a holonomically constrained, fully actuated robot with known kinematics and configuration space $\mathcal{Q}$. The system dynamics may be found using Lagrange's equations (see, for example, [8, 19]). Assume exact knowledge of the kinematics and that the effect of gravity is exactly known and may be cancelled. Also, assume that one may cancel any external forces such as friction. The robot moves a set of feature points via the kinematics in view of a camera system, which acts as an output, $c : \mathcal{Q} \to \mathcal{Y}$.

4.1 Task specification

The state space is constrained by the presence of forbidden configurations, the *obstacle set* $\mathcal{O} \subset \mathcal{Q}$. The *free space* is defined as the obstacle-free configuration space $\mathcal{V} = \mathcal{Q} - \mathcal{O}$, and *safe configurations* $\mathcal{D} \subset \mathcal{V}$ form a compact connected differentiable manifold with boundary. The positioning objective is described in terms of a *goal* $q^* \in \mathcal{D}$. The task is to drive q to q^* asymptotically, subject to the system dynamics, by an appropriate choice of torque input. Moreover, we wish to maintain our configuration within $\mathcal{D}$ to avoid obstacles. We require that the basin of attraction $\mathcal{E} \subset T\mathcal{D}$ includes a dense subset of the zero velocity section of $T\mathcal{D}$, so that we may guarantee convergence from the entire configuration space. Obstacle avoidance requires that the trajectories avoid crossing the *boundary set* $\mathcal{B} = \partial\mathcal{D}$, *i.e.* $q(t) \in \mathcal{D}$, for all $t \geq 0$.

4.2 Navigation functions

The task of moving to a goal while avoiding obstacles along the way can be achieved via a nonlinear generalization of proportional-derivative (PD) control deriving from Lord Kelvin's century old observation that total energy always decreases in damped mechanical systems

[13]. Formally, this entails the introduction of a gradient vector field from a "navigation function," a refined notion of an artificial potential function, together with damping to flush out unwanted kinetic energy. A good survey of potential field methods for robot navigation is given by ([15], Chapter 7). The refinement to *navigation functions*, first articulated by Koditschek and Rimon [13, 14, 20], is only very briefly sketched here, through a very simplified example. The application of navigation functions to visual servoing first appeared in [4], and is most clearly articulated in [7].

4.3 Simple Example

Consider a system with configuration space $\mathcal{Q} = \mathbb{R}$, with two point obstacles $\mathcal{O} = \{-1, 1\}$, and goal $q^* = 0$. The free space is given by $\mathcal{V} = (\infty, -1) \cup (-1, 1) \cup (1, \infty)$. There are three connected components to $\mathcal{V}$, so we choose the compact safe domain $\mathcal{D} = [-1 + \rho, 1 - \rho]$, where $\rho > 0$ is an arbitrarily small safety collar, to contain the goal point. Suppose the dynamics are given by

$$m\ddot{q} = u$$

and apply the NF $\varphi(q) = q^2/(1 - \rho)^2$. Note that φ evalutes to 0 at the goal and 1 on the boundary of $\mathcal{D}$. Letting

$$u = -\alpha \nabla \varphi(q) - b\dot{q} \quad \text{where} \quad \alpha := \frac{k}{2}(1 - \rho)^2 \quad \implies \quad m\ddot{q} + b\dot{q} + kq = 0$$

which renders 0 globally asymptotically stable, and guarantees never to cross the boundary set $\mathcal{B} = \{\pm(1 - \rho)\}$ for all initial positions on $\mathcal{D}$.[1]

   For the purposes of this paper, one may simply consider the first order gradient system given by

$$\dot{q} = -\nabla \varphi(q)$$

when considering the convergence properties of the controller. This lifts naturally to the second order setting as shown above by simply adding a suitable damping term (this is distinctly different from trajectory planning and trajectory tracking).

4.4 Invariance under diffeomorphism

One last key ingredient in the mix of geometry and dynamics underlying the results presented in this paper revolves around the realization that a navigation function in one coordinate system is a navigation function in another coordinate system, if the two coordinate systems are diffeomorphic [13]. This affords the introduction of geometrically simple model spaces, such as the image space, and their correspondingly simple model navigation functions.

---

[1]To be precise, one must put bounds on the initial velocities to compute the domain of attraction on $T\mathcal{D}$ as a function of damping [11].

References

[1] Francois Chaumette. *The Confluence of Vision and Control*, chapter Potential problems of stability and convergence in image-based and position-based visual servoing. Springer-Verlag, 1999.

[2] P. I. Corke and M. C. Good. Dynamic effects in visual closed-loop systems. *IEEE Transactions on Robotics and Automation*, 12(5):651–670, October 1996.

[3] Peter I. Corke and Seth A. Hutchinson. A new partitioned approach to image-based visual servo control. *IEEE Transactions on Robotics and Automation*, 17(4):507–515, 8 2001.

[4] Noah J. Cowan and Daniel E. Koditschek. Planar image based visual servoing as a navigation problem. In *International Conference on Robotics and Automation*, volume 1, pages 611–617, Detroit, MI, 1999. IEEE.

[5] Noah J. Cowan, Gabriel A. D. Lopes, and Daniel E. Koditschek. Rigid body visual servoing using navigation functions. In *Conference on Decision and Control*, pages 3920–3926, Sydney, Australia, 2000. IEEE.

[6] Noah J. Cowan, Joel D. Weingarten, and Daniel E. Koditschek. Empirical validation of a new visual servoing stragegy. In *Conference on Control Applications*, Mexico City, September 2001. IEEE, Omnipress.

[7] Noah J. Cowan, Joel D. Weingarten, and Daniel E. Koditschek. Visual servoing using navigation functions. *IEEE Transactions on Robotics and Automation*, 2002. in review.

[8] J. Craig. *Introduction to Robotics*. Addison-Wesley, Reading, Mass., 1986.

[9] J. Hespanha, Z.Dodds, G. D. Hager, and A.S. Morse. What tasks can be performed with an uncalibrated stereo vision system? *The International Journal of Computer Vision*, 35(1), 1999.

[10] 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.

[11] D. E. Koditschek. The control of natural motion in mechanical systems. *ASME Journal of Dynamic Systems, Measurement, and Control*, 113(4):547–551, Dec 1991.

[12] D. E. Koditschek. An approach to autonomous robot assembly. *Robotica*, 12:137–155, 1994.

[13] Daniel E. Koditschek. The application of total energy as a Lyapunov function for mechanical control systems. In *Dynamics and control of multibody systems (Brunswick, ME, 1988)*, pages 131–157. Amer. Math. Soc., Providence, RI, 1989.

[14] Daniel E. Koditschek and Elon Rimon. Robot navigation functions on manifolds with boundary. *Advances in Applied Mathematics*, 11:412–442, 1990.

[15] Jean-Claude Latombe. *Robot Motion Planning*. Kluwer Academic Publishers, Boston, 1991.

[16] Ezio Malis, Francois Chaumette, and Sylvie Boudet. 2-1/2-d visual servoing. *IEEE Transactions on Robotics and Automation*, pages 238–250, 1999.

[17] Youcef Mezouar and Francois Chaumette. Design and tracking of deserable trajectories in the image space by integrating mecahnical and visibility constraints. In *International Conference on Robotics and Automation*, volume 1, pages 731–6, Seoul, Korea, 2001. IEEE.

[18] G. Morel, T. Leibezeit, J. Szewczyk, S. Boudet, and J. Pot. Explicit incorporation of 2d constraints in vision based control of robot manipulators. In Peter Corke and James Trevelyan, editors, *Experimental Robotics VI*, volume 250 of *Lecture Notes in Control and Information Sciences*. Springer-Verlag, 2000. ISBN: 1 85233 210 7.

[19] Richard M. Murray, Zexiang Li, and S. Shankar Sastry. *A Mathematical Introduction to Robotic Manipulation*. CRC Press, Reading, Mass., 1994.

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

[21] Andreas Ruf and Radu Horaud. Visual servoing of robot manipulators, part "i" : Projective kinematics. *International Journal on Robotics Research*, 18(11):1101 – 1118, November 1999.

[22] Louis L. Whitcomb, Alfred A. Rizzi, and Daniel E. Koditschek. Comparative experiments with a new adaptive contoller for robot arms. *IEEE Transactions on Robotics and Automation*, 9(1):59–70, Feb 1993.

[23] Hong Zhang and James P. Ostrowski. Visual servoing with dynamics: Control of an unmanned blimp. In *International Conf. on Robotics and Automation*, 1999.