What Can Be Done with an Uncalibrated Stereo System?

João Hespanha, Zachary Dodds, Gregory D. Hager, and A.S. Morse
Center for Computational Vision and Control
Yale University
New Haven, CT, 06520
E-mail: (gregory.hager,joao.hespanha,zachary.dodds,as.morse)@yale.edu

Abstract

Over the last several years, there has been an increasing appreciation of the impact of control architecture on the accuracy of visual servoing systems. In particular, it is generally acknowledged that so-called image-based methods provide the highest guarantees of accuracy on inaccurately calibrated hand-eye systems.

Less clear is the impact of the control architecture on the set of tasks which the system can perform. In this article, we present a formal analysis of control architectures for hand-eye coordination. Specifically, we first state a formal characterization of what makes a task performable under three possible encoding methods. Then, for the specific case of cameras modeled using projective geometry, we relate this characterization to notions of projective invariance and demonstrate the limits of achievable performance in this regard.

1 Introduction

In the design of vision-based control systems, two families of control architectures dominate [4, 11]. The first family, characterized by the *position-based* approach, computes feedback from estimates of geometric quantities (pose, feature position, etc.) in the Cartesian space of the robot. The second family, the *image-based* approach, computes feedback directly from image features.

We now restrict our attention to a two-camera system observing both controlled features (e.g. those on a robot manipulator) and a (static) target. It has been observed that there exist approaches which can achieve precise task positioning, even though the relationship between the manipulator and the observing cameras may not be precisely known. [11, 8, 12]. In particular, prior work has shown that one route toward defining tasks performable with absolute precision is to take advantage of the invariance of certain image-level constructions to perspective projection ([7, 9, 2]). It is also known that there are tasks which, under certain conditions, are performable using image-based methods and not performable using position-based methods [8].

In [6] Faugeras specifies the extent to which the world is *reconstructable* using a pair of uncalibrated cameras. It is shown that with a "weakly calibrated" stereo rig, i.e., one in which the epipolar geometry of the cameras is known, reconstruction can be performed up to a projective transformation. This result suggests an additional line of investigation — namely, to what extent does a weakly calibrated rig affect task performability?[1]

Such observations motivate the central question of this article: *what tasks are precisely performable using an imprecisely calibrated pair of cameras?* More specifically, is the general set of tasks performable under image-based control strictly larger than for position-based control? Can one precisely characterize the limits of performability? How is the notion of performability related to projective invariance?

In this article we provide the following answers to the above questions. First, we establish a formal characterization of the set of performable, or *testable*, tasks and show that this is exactly the set of tasks performable by image-based methods for both uncalibrated and weakly calibrated systems. We also provide necessary and sufficient conditions for a task to be testable with a weakly calibrated vision system. In particular, the assertion that a task is testable under weak calibration is equivalent to the statement that the task is invariant under projective transformations. Further, projective invariance is necessary, but not sufficient, to ensure that a task is testable with an

[1] Radu Horaud independently suggested this question to one of the authors at the 1995 ICCV conference.

uncalibrated vision system. Examples of projectively invariant tasks which are not testable in the uncalibrated case are demonstrated. These results bridge the classical computer-vision notions of projective geometry with sensor-based strategies from control theory; they provide a concise characterization of the limits of performability for vision-based motion control architectures.

The remainder of this paper is organized as follows. In the next section, we define and discuss the notion of task encoding and derive results regarding the performability of tasks with absolute precision. Following that, we focus on the class of image-based tasks and show that this set is maximal. In Section 4 we relate performability to invariance over projective transformations. Section 5 discusses some of the limitations of this formalization. We conclude with some observations and possible directions for future research.

Notation: Throughout this article $\mathbb{P}^3$ denotes the 3-dimensional projective space over the field of real numbers $\mathbb{R}$. We recall that an element or *projective point* of $\mathbb{P}^3$ is a one dimensional vector subspace of $\mathbb{R}^4$ and that a *projective line* in $\mathbb{P}^3$ is a two dimensional vector subspace of $\mathbb{R}^4$. For each vector $x \in \mathbb{R}^4 \setminus \{0\}$ we denote by $\mathbb{R}x$ the corresponding projective point. Given two distinct projective points $f_1, f_2 \in \mathbb{P}^3$ we denote by $f_1 \oplus f_2$ the projective line defined by f_1 and f_2. $Ker(T)$, the kernel of the function T, is the set of points x in the domain of T such that $T(x) = 0$. For a general sets, we distinguish strict set inclusion from nonstrict inclusion by using $\subset$ and $\subseteq$, respectively. We also denote the Cartesian product of a set $\mathcal{S}$, as $\mathcal{S}^n \triangleq \underbrace{\mathcal{S} \times \mathcal{S} \times \cdots \times \mathcal{S}}_{n \text{ times}}$.

2 Task Encodings

We consider specifically the problem of controlling the pose x of a robot that moves in a prescribed *workspace* $\mathcal{W} \subset SE(3)$ using two cameras. The data to accomplish this task consists of the projections of point-features attached to the robot as well as point-features attached to the environment that appear in the *two-camera field of view* $\mathcal{V}$. Typically $\mathcal{V}$ is a subset of $\mathbb{R}^3$ or $\mathbb{P}^3$.

Point-features are mapped into the *two-camera image space* $\mathcal{Y}$ through a fixed but imprecisely known *two-camera model* $C_{\text{actual}} : \mathcal{V} \to \mathcal{Y}$, where $\mathcal{Y}$ is a subset of $\mathbb{R}^2 \times \mathbb{R}^2$ or $\mathbb{P}^2 \times \mathbb{P}^2$. If f denotes a vector in $\mathcal{V}$, then the feature position in image space is a measured output vector $y \in \mathcal{Y}$ related to f by the readout formula $y = C_{\text{actual}}(f)$. The two-camera model C_{actual} is a fixed but unknown element of a prescribed set of functions $\mathcal{C}$ which map $\mathcal{V}$ into $\mathcal{Y}$. In the sequel $\mathcal{C}$ is called the *set of admissible two-camera models*. Throughout this article, we assume that all elements of $\mathcal{C}$ are injective, and therefore left-invertible, over $\mathcal{V}$.

2.1 Tasks

A typical problem or task consists of driving the pose of a robot to a "target" (or set) in $\mathcal{W}$. Both the pose of the robot and the target are determined by a list of simultaneously observed point-features in $\mathcal{V}$. Following [11, 8], we opt for an implicit representation of a task in the spirit of [14] and define a *task function* to be any function T which maps an ordered set (or list) of n simultaneously observed point-features $f \triangleq \{f_1, f_2, \ldots, f_n\}$ $\{f_i \in \mathcal{V}\}$ into a real scalar. In some cases only certain sets of features are allowed (e.g., for $n = 3$ the three point-features may be required to be collinear). The set $\mathcal{F} \subseteq \mathcal{V}^n$ of all such lists of interest is called the *feature space*. The *task* is to control the pose of a robot in $\mathcal{W}$ so as to achieve the condition

$$T(f) = 0. \tag{1}$$

Examples of tasks expressed in this form can be found in [14, 5, 3, 2, 11, 8, 13].

The problem addressed in this paper is to determine conditions under which it is possible to verify that (1) holds using only the data that can be sensed by the two cameras. These data are $y \triangleq \{y_1, y_2, \ldots, y_n\}$ with $y_i = C_{\text{actual}}(f_i)$, subject to the assumption that C_{actual} is in $\mathcal{C}$ but it is otherwise unknown.

For compactness of notation, in the remainder of this article we define for each C in $\mathcal{C}$ the function $\bar{C}$ from $\mathcal{F}$ to the set $\mathcal{Y}^n$ by the rule

$$\{f_1, f_2, \ldots, f_n\} \mapsto \{C(f_1), C(f_2), \ldots, C(f_n)\}$$

Therefore the data that can be sensed by the two cameras is simply

$$y = \bar{C}_{\text{actual}}(f) \tag{2}$$

2.2 Task Encoding

Since we presume that C_{actual} is not known with precision, the exact locations of the points f_i cannot be determined from observed data. Therefore it is unclear whether a given task expressed in the form of (1) can be accomplished. In this section we discuss three "task encodings" for implementing tasks in the form of (1). Each task encoding is a paradigm for expressing task constraints based on available information, i.e., a task encoding explicitly connects a system's goal, the task, with its resources, the available visual data. We then state necessary and sufficient conditions on the admissible set of camera models such that satisfying each of the encodings is equivalent to accomplishing the original task (1).

Cartesian-Based Approach This approach, often referred to as "position-based" in the visual servoing literature, is motivated by the intuitive idea of "certainty equivalence." In the present context, certainty equivalence advocates that one should use estimates of f to accomplish task (1), with the understanding that these estimates are to be viewed as correct even though they may not be. The construction of such estimates starts with the selection (by some means) of a two-camera model C_{estim} in $\mathcal{C}$ which is considered to be an approximate model of C_{actual}. Assume that such a C_{estim} has been chosen. Using C_{estim}, we estimate f using a left inverse $\bar{C}_{\text{estim}}^{-1}$ of $\bar{C}_{\text{estim}}$. Such a left inverse exists because $\bar{C}_{\text{estim}}$ is injective. In view of (2), it is natural to define

$$\hat{f} \triangleq \bar{C}_{\text{estim}}^{-1}(y) \tag{3}$$

as the estimate of f to be considered. In accordance with certainty equivalence, we can then seek to accomplish the re-encoded task

$$T\big(\bar{C}_{\text{estim}}^{-1}(y)\big) = 0. \tag{4}$$

Condition (4) expresses the Cartesian-based encoding of the original task (1). We would like to determine those cases when accomplishing the re-encoded task (4) is equivalent to accomplishing (1). The following lemma provides an answer.

Lemma 1. *For a fixed set of admissible models $\mathcal{C}$, define the set of* Cartesian-based tasks, $\mathbf{CB}_{\mathcal{C}}$, *to be the set of all task functions T for which there exists a left inverse $\bar{C}_{\text{estim}}^{-1}$ such that for all $C_{\text{actual}} \in \mathcal{C}$,*

$$\operatorname{Ker} T = \operatorname{Ker}(T \circ \bar{C}_{\text{estim}}^{-1} \circ \bar{C}_{\text{actual}}). \tag{5}$$

Then, for each $f \in \mathcal{F}$ and $C_{\text{actual}} \in \mathcal{C}$, equations (4) and (2) are equivalent to (1) if and only if $T \in \mathbf{CB}_{\mathcal{C}}$.

The details of the proofs of Lemma 1 and subsequent lemmas are provided in [10].

2.2.1 Image-Based Approach

Whereas the Cartesian-based encoding uses reconstructed estimates of points in $\mathcal{V}$ to implement tasks, a task can also be expressed solely in the space of the sensor. Such an encoding is known as *image-based.* This well-known approach requires that there exist a function $\tilde{T} : \mathcal{Y}^n \to \mathbb{R}$ such that

$$\operatorname{Ker} T = \operatorname{Ker}(\tilde{T} \circ \bar{C}), \qquad \forall C \in \mathcal{C} \tag{6}$$

We assume that (6) is satisfied by some computable function $\tilde{T}$. The image-based approach seeks to accomplish

$$\tilde{T}(y) = 0. \tag{7}$$

The approach is justified by the following lemma.

Lemma 2. *For a fixed set of camera models $\mathcal{C}$, define the set of* image-based tasks, $\mathbf{IB}_{\mathcal{C}}$, *to be the set of all tasks T such that (6) holds for some $\tilde{T}$. Then for each $f \in \mathcal{F}$ and $C_{\text{actual}} \in \mathcal{C}$, equations (7) and (2) are equivalent to (1) if and only if $T \in \mathbf{IB}_{\mathcal{C}}$.*

As we will see, e.g., for projective camera models, demanding that (6) holds turns out to be much less severe than requiring that the property (5) holds. For example, for a suitably restricted set of camera models $\mathcal{C}'$ it is possible to demonstrate that $T_{\text{col}} \in \mathbf{IB}_{\mathcal{C}'}$. In a sense to be specified in the next section (6) is actually a necessary condition for a problem to be solvable, and hence $\mathbf{IB}_{\mathcal{C}}$ is in this sense the *largest* set of tasks that can be performed (relative to a given set of admissible camera models $\mathcal{C}$) using cameras as sensors.

Note that, although the image-based approach does not require the selection of a candidate two-camera model C_{estim} in $\mathcal{C}$ to serve as an estimate of C_{actual}, in practice designing a controller that accomplishes (7) will typically require such an estimate. It is important to realize, however, that the choice of camera model affects the dynamics (in particular the stability) of the closed loop system, *but not its accuracy.* [8]

2.2.2 Modified Cartesian-Based Approach

An image-based task encoding, as shown in the next section, is a maximally powerful encoding. The Cartesian-based offers the advantage of expressing system control in three-dimensional Euclidean space, the natural choice for specifying robot tasks. The modified Cartesian-based encoding, introduced in [1], is essentially a compromise between these two approaches.

A modified Cartesian-based encoding resembles Cartesian-based in that it uses the left inverse of a camera model estimate, $\bar{C}_{\text{estim}}^{-1}$, to obtain approximated positions of image points. Similar to image-based encoding, the modified Cartesian-based paradigm also depends on the existence of two functions, $T' : \mathcal{F} \to \mathbb{R}$ and $R : \mathcal{Y}^n \to \mathcal{Y}^n$, such that

$$\operatorname{Ker} T = \operatorname{Ker}(T' \circ \bar{C}_{\text{estim}}^{-1} \circ R \circ \bar{C}), \quad \forall C \in \mathcal{C}. \tag{8}$$

Informally, R can be thought of as an arbitrary image-based construction of some or all of the observed points. That construction, perhaps along with the rest of the observed points, is reconstructed in Cartesian space. There, the resulting points are considered correctly configured when in the kernel of T'.

With the necessary functions in place, the modified Cartesian-based encoding seeks to achieve

$$T'(\bar{C}_{\text{estim}}^{-1}(R(y))) = 0. \tag{9}$$

As before, the following lemma defines the set of tasks performable with a modified Cartesian-based task encoding.

Lemma 3. *For a fixed set of camera models $\mathcal{C}$, define the set of* modified Cartesian-based tasks, $\mathbf{MCB}_{\mathcal{C}}$, *to be the set of all tasks T for which there exists a left inverse $\bar{C}_{\text{estim}}^{-1}$ such that* (8) *holds for some T' and R. Then for each $f \in \mathcal{F}$ and $C_{\text{actual}} \in \mathcal{C}$, equations* (9) *and* (2) *are equivalent to* (1) *if and only if $T \in \mathbf{MCB}_{\mathcal{C}}$.*

With the definitions above we can immediately state some relationships among sets of performable tasks:

Lemma 4. *For fixed $\mathcal{C}$, $\mathbf{CB}_{\mathcal{C}} \subseteq \mathbf{MCB}_{\mathcal{C}} \subseteq \mathbf{IB}_{\mathcal{C}}$.*

Setting R to the identity and T' to T in (9) suffices to show the first set inclusion. The second follows by making $\tilde{T}$ equal to $T' \circ \bar{C}_{\text{estim}}^{-1} \circ R$ in (6).

3 Testable Task Functions

In the previous section we discussed methods to re-encode a task specified in the form of (1) using only the data that can be sensed by two cameras. In this section we derive a minimum set of assumptions under which it is possible to verify if condition (1) holds using the sensed data $y \triangleq \bar{C}_{\text{actual}}(f)$, even though C_{actual} is an unknown element of $\mathcal{C}$.

In the sequel we define a task function $T : \mathcal{F} \to \mathbb{R}$ to be *testable on* $\mathcal{C}$ if there exists an algorithm $\mathfrak{A}$ which, for each two-camera model $C \in \mathcal{C}$ and each $f \in \mathcal{F}$ is able to determine whether or not $T(f) = 0$ from the measured data $y = \bar{C}(f)$. This notion of testability is meant to capture the idea of task performability without regard to the details of a control procedure. We proceed by deriving necessary and sufficient conditions characterizing those task functions which are testable.

Lemma 5 (Testability). *The following statements are equivalent:*

1. *The task function $T : \mathcal{F} \to \mathbb{R}$ is testable on $\mathcal{C}$.*
2. *There exists a map $\tilde{T} : \mathcal{Y}^n \to \mathbb{R}$ such that*

$$\operatorname{Ker} T = \operatorname{Ker}(\tilde{T} \circ \bar{C}), \qquad \forall C \in \mathcal{C}. \tag{10}$$

If T is testable on $\mathcal{C}$, and $\tilde{T}$ is as in (10), an algorithm that can be used to determine if $T(f) = 0$ is

`If` $\tilde{T}(y) = 0$ `then` $T(f) = 0$ `else` $T(f) \neq 0$.

Because of this we follow [1] to say that $\tilde{T}(y) = 0$ is an *encoding* of the task $T(f) = 0$. Note that the equivalence of conditions 1 and 2 implies that the set of testable tasks is exactly the set of image-based tasks and hence $\mathbf{IB}_{\mathcal{C}}$ contains all tasks which can be performed for a given set of admissible camera models.

We also note the following about the continuity of the encoding function, $\tilde{T}$. Suppose that $\mathcal{V}$ is a compact subset of a finite dimensional space, T is a continuous function, and $\mathcal{C}$ is of the form $\mathcal{C} = \{C_p : p \in \mathcal{P}\}$ with $\mathcal{P}$ a compact subset of a finite dimensional space. If $\{p, f\} \mapsto C_p(f)$ is a function continuous on $\mathcal{P} \times \mathcal{V}$, then, without loss of generality, the function $\tilde{T}$ in statement 2 of Lemma 5 can be assumed continuous [10].

4 Performability for Projective Systems

Until now, all results pertaining to the performability of positioning tasks were true for *any* set of admissible camera models with the appropriate structure. In this section, we focus specifically on cameras described by projective models. For these cameras, we assume a field of view $\mathcal{V}$ which is a subset of $\mathbb{P}^3$ of the form

$$\mathcal{V} = \left\{ \mathbb{R} \begin{bmatrix} x \\ 1 \end{bmatrix} : x \in \mathcal{V}_{\mathbb{R}} \right\},$$

where $\mathcal{V}_{\mathbb{R}}$ is a bounded subset of $\mathbb{R}^3$ and the image space $\mathcal{Y}$ is $\mathbb{P}^2 \times \mathbb{P}^2$.

Notation: We denote by $GL(4)$ the general linear group of nonsingular 4×4 real matrices. Given a matrix $A \in GL(4)$ we denote by $\langle A \rangle$ the action of A on $\mathbb{P}^3$ which is a function from $\mathbb{P}^3$ to $\mathbb{P}^3$ defined by the rule $\mathbb{R}x \mapsto \mathbb{R}Ax$. The function $\langle A \rangle$ is called a projective transformation. Similarly, given a 3×4 full rank matrix A, we denote by $\langle A \rangle$ the function from $\mathbb{P}^3 \setminus \operatorname{Ker} A$ to $\mathbb{P}^2$ defined by the rule $\mathbb{R}x \mapsto \mathbb{R}Ax$.

4.1 Projective Camera Models

Given two 3×4 full rank matrices M, N such that $\operatorname{Ker} M$ and $\operatorname{Ker} N$ are not projective points in $\mathcal{V}$ we say that C is the *projective two-camera model defined by the pair* $\{M, N\}$ and write $C = \{\langle M \rangle, \langle N \rangle\}$ if C is the map from $\mathcal{V}$ to $\mathcal{Y}$ defined by the rule

$$f \mapsto \{\langle M \rangle(f), \langle N \rangle(f)\}$$

The projective points defined by $\operatorname{Ker} M$ and $\operatorname{Ker} N$ are the *optical centers of the cameras*, and the projective line defined by these projective points, i.e. $\operatorname{Ker} M \oplus \operatorname{Ker} N$, is the *baseline of the cameras*. Our requirement of injectivity is that the optical centers of the two cameras must be distinct and that the cameras' baseline does not intersect $\mathcal{V}$.

In particular, we consider two sets of admissible two-camera models. The first contains uncalibrated cameras; the second, weakly calibrated rigs.

Uncalibrated projective two-camera system. Let $\mathcal{C}_{\rm uc}$ denote the set of every projective two-camera model $C = \{\langle M \rangle, \langle N \rangle\}$ defined by any pair of full rank matrices $M, N \in \mathbb{R}^{3\times 4}$ such that the cameras' centers are distinct and their baseline does not intersect $\mathcal{V}$. This situation is often referred to as an *uncalibrated projective two-camera system.*

Weakly calibrated projective two-camera system. Let $C_0 = \{\langle M_0 \rangle, \langle N_0 \rangle\}$ be some element of $\mathcal{C}_{\rm uc}$. We denote by $\mathcal{C}_{\rm wk}(C_0)$ the subset of $\mathcal{C}_{\rm uc}$ consisting of all those camera models in $\mathcal{C}_{\rm uc}$ which have the same epipolar geometry as C_0. The elements of $\mathcal{C}_{\rm wk}(C_0)$ are projective two-camera models $C = \{\langle M_0 A \rangle, \langle N_0 A \rangle\}$ where A can be any nonsingular matrix such that the cameras' centers are distinct and their baseline does not intersect $\mathcal{V}$. This situation is often referred to as a *weakly calibrated projective two-camera system.* Clearly, $\mathcal{C}_{\rm wk}(C_0) \subset \mathcal{C}_{\rm uc}$ for each $C_0 \in \mathcal{C}_{\rm uc}$. For the rest of the paper we assume that a suitable epipolar geometry is given and fixed and write $\mathcal{C}_{\rm wk}$ without the argument C_0.

4.2 Task Invariance

We further characterize the set of testable tasks by introducing the notion of *invariant tasks:*

Definition 1. We say a task function $T : \mathcal{F} \to \mathbb{R}$ is *invariant under projective transformations* if for every nonsingular matrix $A \in GL(4)$ and all length-n lists of features $f, g \in \mathcal{F}$ such that

$$\langle A \rangle(f_i) = g_i, \qquad i = 1, 2, \ldots, n$$

it is true that $T(f) = 0$ if and only if $T(g) = 0$.

Let **PI** denote the set of all task functions which are invariant under projective transformations. With these definitions in place, it is possible to prove the following statements:

Lemma 6 (Testability on $\mathcal{C}_{\rm wk}$). *A task function T is testable on $\mathcal{C}_{\rm wk}$ if and only if $T \in \mathbf{PI}$, i.e., T is invariant under projective transformations.*

Lemma 7 (Testability on $\mathcal{C}_{\rm uc}$). *A necessary condition for a task function T to be testable on $\mathcal{C}_{\rm uc}$ is that $T \in \mathbf{PI}$.*

Taken together, these results can be summarized as follows:

Theorem 1. *For sets of camera models and tasks as defined above:*

$$\mathbf{IB}_{\mathcal{C}_{\rm uc}} \subset \mathbf{PI} = \mathbf{IB}_{\mathcal{C}_{\rm wk}}$$

Strictness of the first inclusion follows from the example task of making four point features coplanar: that is, consider a task function $T_{\rm cp}$ which takes four points as input and which equals zero only when those four points are coplanar. The fact that $T_{\rm cp} \in \mathbf{IB}_{\mathcal{C}_{\rm wk}}$ is shown in [6]. The fact that $T_{\rm cp} \notin \mathbf{IB}_{\mathcal{C}_{\rm uc}}$ is shown by constructing an example of two admissible camera models C and C' and two sets of four points, f coplanar and f' not coplanar, such that $\bar{C}(f) = \bar{C}'(f')$ [10].

In short, a weakly calibrated camera system can perform all (point-based) tasks which are invariant to projective transformations. As noted above, this set also defines the upper limit of performable tasks. Uncalibrated cameras can perform a strictly smaller set of tasks. In light of Lemma 6, we know that the previously mentioned coplanarity task, $T_{\rm cp}$, is an example of a projectively invariant task which is not in $\mathbf{IB}_{\mathcal{C}_{\rm uc}}$.

5 Some Limitations of the Formulation

Surprisingly, to date we know of no example which differentiates $\mathbf{IB}_{\mathcal{C}_{\mathrm{uc}}}$ and $\mathbf{CB}_{\mathcal{C}_{\mathrm{uc}}}$ within the framework presented in this paper, although in practice we have implemented several tasks which can be performed with absolute accuracy only by image-based methods [8]. As hinted above, this incongruity has to do with the richness of the set of camera models and feature lists. In practice, the set of camera models is smaller than allowed here – in particular, visual singularities are removed from the set of admissible camera models. However, in the general case presented here such singularities make tasks such as T_{col} impossible to reconcile with the requirement that an image-based encoding must be equivalent to its underlying task *for every possible camera pair and feature list.*

The distinctions among the three task encodings also need to be tightened in two ways: both as defined here and for more specific cameras and feature lists. If we suitably restrict the viewspace $\mathcal{V}$ and the set of camera models to remove "visual singularities", it is possible to separate $\mathbf{IB}_{\mathcal{C}_{\mathrm{uc}}}$ from $\mathbf{CB}_{\mathcal{C}_{\mathrm{uc}}}$, (e.g., T_{col}) as well as $\mathbf{MCB}_{\mathcal{C}_{\mathrm{uc}}}$ from $\mathbf{CB}_{\mathcal{C}_{\mathrm{uc}}}$. A task T_{ql} which separates the latter two sets is one in which a point is to be positioned at the center (i.e., the intersection of the diagonals) of a planar quadrilateral whose corners are observed. For this task the ability to construct the unseen goal point in the image allows precise positioning via **MCB** but differing reconstructive errors do not allow precision with **CB** [1].

In all of these cases, however, the additional restrictions on admissible cameras and usable feature points expand the set of performable tasks. This expansion may introduce performable tasks which are not in **PI**. The formalism currently lacks a systematic way to reduce the features and cameras under consideration. Such a system would allow us to specify if and when the links in the chains of Lemma 4 merge into equivalent sets. Most fundamentally, we do not, to date, know of a general way of suitably weakening the definition of performability in order to explore the structure of the task space framed in this work.

6 Conclusions

The main results of this paper can be briefly summarized as follows:

- A task is performable with absolute accuracy with weakly calibrated cameras if and only if it is invariant to projective transformations. This set of tasks is therefore maximal for weakly calibrated cameras.

- If a task is performable with absolute accuracy with uncalibrated cameras for which no epipolar information can be computed, then it is invariant to projective transformations. This set of tasks is strictly smaller than that for weakly calibrated cameras.

- The set of tasks which an image-based encoding can perform is maximal among the three encoding approaches presented here. In addition, image-based tasks are, in fact, everything that can be done only using measured data.

  It is unclear at this point if in general $\mathbf{IB}_{\mathcal{C}}$ is strictly larger than the set of tasks performable using Cartesian (position-based) or modified Cartesian methods.

Section 5 provides a rough outline of the future direction we plan with this work. Up to now, the set of cameras considered admissible is, in some sense, too large. It covers many camera configurations that can be safely assumed not to occur in practice. We are working to extend the results to draw finer distinctions among the hierarchy of task spaces by diminishing the size of $\mathcal{C}$. Additional complications arise on this path, as the sets of camera configurations which we would like to exclude often depend on the positions of the feature points used or the specification of the task itself.

The use of points as the basis for task specification simplifies the notation of encodings, but more general features have been used in [1]. The set of point-to-point tasks is a rich one, as it encompasses many of the examples found in the literature. In addition, tasks involving lines and conics can often be constructed from observed point features, e.g., T_{col} and T_{ql}.

Our results seek to provide insight into what tasks are performable using the information provided by uncalibrated stereo. In reality, tasks lie along a spectrum of performability, where errors due to noise, the control algorithm, and sensor singularities contribute to imprecise execution. This paper's framework classifies tasks as belonging (or not belonging) to a particular set. The notion of "distance" from a particular set, for example, would provide a compelling extension to the binary relationship of inclusion. Such taxonomies, whether continuous or discrete, assist human or automated system executives by guiding the choice of an approach for accomplishing a certain set of tasks.

Acknowledgments This work was supported by National Science Foundation grant IRI-9420982, and by funds provided by Yale University. Thanks to Radu Horaud and David Kriegman for providing valuble encouragement to pursue this problem.

References

[1] W-C. Chang, J. P. Hespanha, A.S. Morse, and G.D. Hager. Task re-encoding in vision-based control systems. In *Proceedings, Conference on Design and Control*, volume to appear, San Diego, CA, December 1997.

[2] F. Chaumette, E. Malis, and S. Boudet. 2d 1/2 visual servoing with respect to a planar object. In *Proc. IROS Workshop on New Trends in Image-based Robot Servoing*, pages 43–52, 1997.

[3] F. Chaumette, P. Rives, and B. Espiau. Classification and realization of the different vision-based tasks. In K. Hashimoto, editor, *Visual Servoing*, pages 199–228. World Scientific, 1994.

[4] P. I. Corke. Visual control of robot manipulators—a review. In K. Hashimoto, editor, *Visual Servoing*, pages 1–32. World Scientific, 1994.

[5] B. Espiau, F. Chaumette, and P. Rives. A New Approach to Visual Servoing in Robotics. *IEEE Trans. on Robotics and Automation*, 8:313–326, 1992.

[6] Olivier D. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig? In *Proc., ECCV*, pages 563–578, 1992.

[7] G. D. Hager. Real-time feature tracking and projective invariance as a basis for hand-eye coordination. In *Proc. IEEE Conf. Comp. Vision and Patt. Recog.*, pages 533–539. IEEE Computer Society Press, 1994.

[8] G. D. Hager. A modular system for robust hand-eye coordination. *IEEE Trans. Robot. Automat*, 13(4):582–595, 1997.

[9] G. D. Hager and Z. Dodds. A projective framework for constructing accurate hand-eye systems. In *Proc. IROS Workshop on New Trends in Image-based Robot Servoing*, pages 71–82, 1997.

[10] J. Hespanha, Z. Dodds, G. D. Hager, and A. S. Morse. What visual tasks are performable without calibration? DCS RR-1138, Yale University, New Haven, CT, October 1997.

[11] S. Hutchinson, G. D. Hager, and P. Corke. A tutorial introduction to visual servo control. *IEEE Trans. Robot. Automat*, 12(5), 1996.

[12] M. Jagersand, O. Fuentes, and R. Nelson. Experimental evaluation of uncalibrated visual servoing for precision manipulation. In *Proc., ICRA*, pages 2874–2880, 1997.

[13] L. Robert, C. Zeller, and O. Faugeras. Applications of non-metric vision to some visually guided robotics tasks. Technical Report 2584, INRIA, Sophia-Antipolis, June 1995.

[14] C. Samson, M. Le Borgne, and B. Espiau. *Robot Control: The Task Function Approach.* Clarendon Press, Oxford, England, 1992.