Proceedings of the 37th IEEE
Conference on Decision & Control
Tampa, Florida USA • December 1998

Decidability of Robot Positioning Tasks Using Stereo Vision Systems*

J. P. Hespanha joao.hespanha@yale.edu
Z. Dodds zachary.dodds@yale.edu
G. D. Hager gregory.hager@yale.edu
A. S. Morse morse@sysc.eng.yale.edu

Center for Computational Vision and Control
Yale University, New Haven, CT 06520

Abstract

Conditions are given which enable one to decide on the basis of images of point features observed by an imprecisely modeled two-camera stereo vision system, whether or not a prescribed robot positioning task has been accomplished with precision. It is shown that for a stereo vision system whose two-camera model is known only up to a projective transformation, one can decide with the available data whether or not such a positioning task has been accomplished, just in case the function which specifies the task is a projective invariant. It is then shown that, under suitable technical conditions, every task with this property can be constructed from primitive tasks by using only a small number of distinct operations.

1 Introduction

Feedback control systems employing video cameras as sensors have been studied in the robotics community for many years {c.f. recent tutorial on visual servoing [1]}. An especially interesting feature of such systems is that *both* the process state {e.g., the position and orientation of the robot in its workspace} and the reference set-point {e.g., the landmark determined target} are typically observed through the *same* sensors {i.e., cameras}. Because of this unusual architectural feature, it is sometimes possible to achieve *precise* positioning, despite sensor/actuator and process model imprecision [2, 3, 4, 5], just as it is in the case of a noise-free conventional set-point control system with a loop-integrator and fixed exogenous reference.

The aim of this paper is to give conditions which enable one to decide on the basis of images of point features observed by an imprecisely modeled two-camera stereo vision system, whether or not a prescribed positioning task has been accomplished. To make precise what the issue is, we formalize the concept of a "positioning task" and then introduce the notion of a "decidable positioning task." By a *positioning task* is meant the objective of bringing the pose of a robot to a specified "target" in the robot's workspace. The pose of the robot and the target are both determined by a list of point features simultaneously observed by an imprecisely modeled stereo vision system. Formally, a positioning task is then represented by an equation of the form $T(f) = 0$, where T is a *task function* and f a list of observed robot and target point features [6, 7, 8, 9, 1, 10, 11, 12]. A positioning task is then said to be *decidable* if it is possible to determine whether or not $T(f) = 0$ on the basis of an observed image of f. We show that for a projective stereo vision system with known epipolar geometry {i.e. weakly calibrated [13]} a positioning task is decidable just in case the task function T is a suitably defined projective invariant. This result is inspired by [14] whose findings suggest that for a weakly calibrated stereo vision system, there ought to be a close relationship between the decidability of a given task $T(f) = 0$ and the *invariant properties of* T under projective transformations. This result also underscores the observation made in [10, 15, 16, 17] that accurate metric information is not needed for the accomplishment of many types of positioning tasks with a stereo vision system.

This paper continues a line of research started in [18] and subsequently pursued in [19, 20, 21].

Notation: Throughout this paper, prime denotes matrix transpose, $\mathbb{R}^m$ is the real linear space of m-dimensional vectors, and $\mathbb{P}^m$ is the real projective space of one-dimensional subspaces of $\mathbb{R}^{m+1}$. Recall that the elements of $\mathbb{P}^m$ are called *points, lines* in $\mathbb{P}^m$ are two-dimensional subspaces of $\mathbb{R}^{m+1}$, and for $m > 2$ planes are three-dimensional subspaces of $\mathbb{R}^{m+1}$. A point $p \in \mathbb{P}^m$ is said to be *on* a line ℓ (respectively plane ψ) in $\mathbb{P}^m$ if p is a linear subspace of ℓ (respectively ψ) in $\mathbb{R}^{m+1}$. For each nonzero vector $x \in \mathbb{R}^m$, $\mathbb{R}x$ denotes both the one-dimensional linear span of x, and also the point in $\mathbb{P}^m$ which x *represents.* The spe-

*This research was supported by the National Science Foundation, the Army Research Office, and the Air Force Office of Scientific Research

cial Euclidean group of rigid body transformations is denoted by SE(3).

2 Formulation

This paper is concerned with the problem of controlling the pose x of a robot which moves in a prescribed *workspace* $\mathcal{W} \subset \mathrm{SE}(3)$ using two cameras functioning as a position measuring system. The data available for this purpose consists of projections onto the cameras' image planes of robot point features as well as point features in the environment[1]. All such features lie within the two cameras' *joint field of view* $\mathcal{V}$. Typically $\mathcal{V}$ is taken to be either a nonempty subset of $\mathbb{R}^3$ or of $\mathbb{P}^3$.

Point features are mapped into the two cameras' *joint image space* $\mathcal{Y}$ through a fixed but imprecisely known *two-camera model* $C_{\text{actual}} : \mathcal{V} \to \mathcal{Y}$ where, depending on the problem, $\mathcal{Y}$ may be either $\mathbb{R}^2 \oplus \mathbb{R}^2$ or $\mathbb{P}^2 \times \mathbb{P}^2$. Typically many point features are observed all at once. If f_i is the ith such point feature in $\mathcal{V}$, then f_i's observed position in $\mathcal{Y}$ is given by the measured output vector $y_i = C_{\text{actual}}(f_i)$. The two-camera model C_{actual} is a fixed but unknown element of a prescribed set of injective functions $\mathcal{C}$ which map $\mathcal{V}$ into $\mathcal{Y}$. In the sequel $\mathcal{C}$ is called the *set of admissible two-camera models.* For the present, no constraints are placed on the elements of $\mathcal{C}$ other than they be injective functions mapping $\mathcal{V}$ into $\mathcal{Y}$.

2.1 Tasks

By a positioning task or simply a "task" is meant, roughly speaking, the objective of bringing the pose of a robot to a "target" in $\mathcal{W}$. Both the pose of the robot under consideration and the target to which it is to be brought, are determined by a list of simultaneously observed point features in $\mathcal{V}$. As in [1, 10, 6], tasks are represented mathematically as equations to be satisfied. By a "task function" is meant, loosely speaking, a function which maps ordered sets {i.e., lists} of n simultaneously appearing point features $\{f_1, f_2, \ldots, f_n\}$ in $\mathcal{V}$ into the integer set $\{0, 1\}$. We use an un-subscripted symbol such as f to denote each such list and we henceforth refer to f as a *feature.* The set $\mathcal{F} \triangleq \mathcal{V}^n = \underbrace{\mathcal{V} \times \mathcal{V} \times \cdots \times \mathcal{V}}_{n \text{ times}}$ of all such lists of interest is called the *admissible feature space.* A *task function* is then a given function T from $\mathcal{F}$ to $\{0, 1\}$. The *task* specified by T is the equation

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

In case (1) holds we say that the task is *accomplished at* f. Examples of tasks defined in this manner can be be found in [6, 7, 8, 9, 1, 10, 11, 12].

In order to complete the problem formulation, it is helpful to introduce the following. For each C in $\mathcal{C}$, let $\bar{C}$ denote the function from $\mathcal{F}$ to the set $\mathcal{Y}^n = \underbrace{\mathcal{Y} \times \mathcal{Y} \times \cdots \times \mathcal{Y}}_{n \text{ times}}$ which is defined by the rule

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

We sometimes call $\bar{C}$ the *extension* of C to $\mathcal{F}$. The aim of this paper is then to give conditions which enable one to decide on the basis of the a priori information, namely $\mathcal{C}$, T and the measured data

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

whether or not task (1) has been accomplished.

2.2 Decidability

Since C_{actual} is not presumed to be known with precision, the exact locations of point features cannot be reconstructed from the observed data. Therefore it is not clear if, on the basis of this data, it is possible to decide whether or not a given task has been accomplished. To make precise what the issue is, we need to formalize the notion of an "encoded task". Toward this end, let us call a function $E_T : \mathcal{Y}^n \to \mathbb{R}$ an *encoded task function* if it can be constructed solely from knowledge of the available a priori information, namely the set of admissible two-camera models $\mathcal{C}$ and the task function T. In particular, it must be possible to construct E_T without knowledge of the actual camera model C_{actual}. With E_T so constructed, the equation

$$E_T(y) = 0 \tag{3}$$

is said to be an *encoding* of task (1) or simply the *encoded task.* In case (3) holds we say that the encoded task is *accomplished at* y. A task $T(f) = 0$ is said to be *verifiable on* $\mathcal{C}$ with an encoding $E_T(y) = 0$ if,

$$T(f) = 0 \Leftrightarrow E_T(y)|_{y=\bar{C}(f)} = 0, \quad \forall f \in \mathcal{F},\ C \in \mathcal{C} \tag{4}$$

In other words, $T(f) = 0$ is verifiable on $\mathcal{C}$ with a given encoding $E_T(y) = 0$, if for each feature $f \in \mathcal{F}$ and each admissible camera model C in $\mathcal{C}$, the task $T(f) = 0$ is accomplished at f just in case the encoded task $E_T(y) = 0$ is accomplished at $y = \bar{C}(f)$. The reader is referred to [20, 21] for a discussion on methods of building encodings.

A given task is then said to be *decidable on* $\mathcal{C}$, if it is verifiable on $\mathcal{C}$ with some encoding. In other words, $T(f) = 0$ is decidable on $\mathcal{C}$ if there exists an encoding $E_T(y) = 0$ for which (4) holds. The notion of

[1] Depending on the problem being considered, a "point feature" may be represented by either a point in $\mathbb{R}^3$ or a point in $\mathbb{P}^3$ {i.e. a one-dimensional subspace of $\mathbb{R}^4$}. In the examples which follow, points in $\mathbb{R}^m$ are related to points in $\mathbb{P}^m$ by the injective function $x \longmapsto \mathbb{R}\bar{x}$ where $\bar{x}$ is the vector $[x' \quad 1]'$ in $\mathbb{R}^{m+1}$. With this correspondence, geometrically significant points in $\mathbb{R}^3$ such as a camera's optical center can be unambiguously represented as points in $\mathbb{P}^3$.

decidability thus singles out those task which are verifiable, without regard to which particular encodings they might be verified with. In [21], necessary and sufficient conditions to check for task verifiability and decidability are given.

3 Projective Camera Models

In this section we specialize the decidability question to the geometrically meaningful case when the two-camera models of interest are pairs of projective camera models which map subsets of $\mathbb{P}^3$ containing $\mathcal{V}$, into $\mathbb{P}^2 \times \mathbb{P}^2$. Projective models of this type are widely used in the computer vision field, mainly because they include as special cases many popular camera models such as those appropriate to perspective, affine and orthographic cameras. By restricting our attention to projective models, we are able to provide a complete and concise characterization of decidable tasks in terms of projective invariance [22].

3.1 Camera Models

In the sequel we are concerned with camera models whose fields of view are all the same subset $\mathcal{V} \subset \mathbb{P}^3$. We take $\mathcal{V}$ to be of the form

$$\mathcal{V} \triangleq \left\{ \mathbb{R} \begin{bmatrix} w \\ 1 \end{bmatrix} : w \in \mathcal{B} \right\},$$

where $\mathcal{B}$ is a suitably defined subset of $\mathbb{R}^3$. To avoid degenerate situations we assume that $\mathcal{B}$ has a nonempty interior.

For each real 3×4 full-rank matrix M, let $\mathbb{P}_M$ denote the set of all points in $\mathbb{P}^3$ except for the kernel of M, and write $\mathbf{M}$ for the function from $\mathbb{P}^3_M$ to $\mathbb{P}^2$ defined by the rule $\mathbb{R}x \longmapsto \mathbb{R}Mx$. We call $\mathbf{M}$ the *global camera model induced by* M. In the event that $\operatorname{Ker} M \notin \mathcal{V}$, the restricted function $\mathcal{V} \to \mathbb{P}^2$, $v \longmapsto \mathbf{M}(v)$ is well-defined. We denote this function by $\mathbf{M}|\mathcal{V}$ and refer to it as the *camera model determined by* $\mathbf{M}$ *on* $\mathcal{V}$. We call $\operatorname{Ker} M$ the *optical center* of $\mathbf{M}$. In the event that M is of the form

$$M = \begin{bmatrix} R & -Rc \end{bmatrix}$$

where R is a 3×3 rotation matrix and c a vector in $\mathbb{R}^3$, $\mathbf{M}$ models a perspective camera with unit focal length, optical center at c, and orientation defined by R. In this case the kernel of M is $\mathbb{R}\left[\begin{smallmatrix} c \\ 1 \end{smallmatrix}\right]$ which justifies calling it the optical center of $\mathbf{M}$.

3.2 Two-Camera Models

The aim of this section is to define what we mean by an uncalibrated two-camera projective model. We take each such model's joint image space and field of view to be $\mathcal{Y} \triangleq \mathbb{P}^2 \times \mathbb{P}^2$ and $\mathcal{V}$ respectively, where $\mathcal{V}$ is a subset of the form described in the last section. For each pair of real 3×4 full-rank matrices $\{M, N\}$ with distinct kernels, let $\mathbb{P}^3_{\{M,N\}}$ denote the set of all points in $\mathbb{P}^3$ except for $\operatorname{Ker} M$ and $\operatorname{Ker} N$. Let $\{\mathbf{M}, \mathbf{N}\}$ denote the function from $\mathbb{P}^3_{\{M,N\}}$ to $\mathbb{P}^2$ defined by the rule $\mathbb{R}x \longmapsto \{\mathbb{R}Mx, \mathbb{R}Nx\}$. We call $\{\mathbf{M}, \mathbf{N}\}$ the *global two-camera model induced by* $\{M, N\}$. For each model $\{\mathbf{M}, \mathbf{N}\}$ for which neither $\operatorname{Ker} M$ nor $\operatorname{Ker} N$ are in $\mathcal{V}$, it is possible to define the restricted function $\mathcal{V} \to \mathbb{P}^2$, $v \longmapsto \{\mathbf{M}, \mathbf{N}\}(v)$. We denote this function by $\{\mathbf{M}, \mathbf{N}\}|\mathcal{V}$ and refer to it as the *two-camera model determined by* $\{\mathbf{M}, \mathbf{N}\}$ *on* $\mathcal{V}$.

The line in $\mathbb{P}^3$ on which the optical centers of $\mathbf{M}$ and $\mathbf{N}$ lie is called the *baseline of* $\{\mathbf{M}, \mathbf{N}\}$. One can show that the mapping by $\{\mathbf{M}, \mathbf{N}\}$ into $\mathcal{Y}$, of points in $\mathbb{P}^3_{\{M,N\}}$ which do not lie on this baseline, is one to one [21]. In the sequel, for each line ℓ in $\mathbb{P}^3$, we write $\mathbb{P}^3[\ell]$ for the set of points in $\mathbb{P}^3$ which are not on ℓ and we say that the baseline of $\{\mathbf{M}, \mathbf{N}\}$ lies outside of $\mathcal{V}$ if there is no point on the model's baseline which is also in $\mathcal{V}$.

Let us note that $\mathcal{V}$ is automatically contained in the domain of any global two-camera model whose baseline lies outside of $\mathcal{V}$. Each such global model G thus determines a two-camera model $C = G|\mathcal{V}$. By the set of all *uncalibrated two-camera models on* $\mathcal{V}$, written $\mathcal{C}_{\text{uncal}}[\mathcal{V}]$, is meant the set of all two-camera models which are determined by global two-camera models whose baselines lie outside of $\mathcal{V}$. A stereo vision system whose two cameras admit a model that is known to be in this class but is otherwise unknown, is said to be *uncalibrated.*

3.3 Weakly Calibrated Stereo Vision Systems

It is well-known that, with a "weakly calibrated" stereo vision system, it is possible to reconstruct the position of point features in the two cameras' field of view from image measurements. However, this reconstruction is only unique up to a projective transformation [14].

These findings suggests that there is a connection between decidability of a task $T(f) = 0$ on a weakly calibrated stereo vision system and the properties of $T(f) = 0$ which are invariant under projective transformations. In this section, we demonstrate that this is the case. As a result we are able to characterize the set of all those tasks which are decidable using a weakly calibrated stereo vision systems.

Our immediate goal is to make precise what we mean by a weakly calibrated stereo vision system. Toward this end, let us write $\mathrm{GL}(4)$ for the general linear group of real, nonsingular, 4×4 matrices. For each such matrix A, $\mathbf{A}$ denotes the corresponding projective transformation $\mathbb{P}^3 \to \mathbb{P}^3$, $\mathbb{R}x \longmapsto \mathbb{R}Ax$.

For each global two-camera model $G_0 \triangleq \{\mathbf{M_0}, \mathbf{N_0}\}$ whose baseline ℓ lies outside of $\mathcal{V}$, let $\mathcal{C}[G_0]$ denote the set of two-camera models

$$\mathcal{C}[G_0] \triangleq \{(G_0 A)|\mathcal{V} : A \in \mathrm{GL}(4), \text{ and } \mathbf{A}(\mathcal{V}) \subset \mathbb{P}^3[\ell]\}$$

where $G_0 A$ is the global two-camera model induced by $\{M_0 A, N_0 A\}$. A stereo vision system whose two cameras admit a model which is known be in $\mathcal{C}[G_0]$ but is otherwise unknown, is said to be *weakly calibrated* [14, 13].

3.4 Main Result

We now define what is meant by a "projectively invariant task". For each $A \in \mathrm{GL}(4)$, let $\bar{\mathbf{A}}$ denote the extended function from $(\mathbb{P}^3)^n$ to $(\mathbb{P}^3)^n$ defined by the rule

$$\{p_1, p_2, \ldots, p_n\} \longmapsto \{\mathbf{A}(p_1), \mathbf{A}(p_2), \ldots, \mathbf{A}(p_n)\}$$

where $(\mathbb{P}^3)^n \triangleq \underbrace{\mathbb{P}^3 \times \mathbb{P}^3 \times \cdots \times \mathbb{P}^3}_{n \text{ times}}$. Call two points f and g in $\mathcal{F}$ *projectively equivalent* if there exists an A in GL(4) such that $f = \bar{\mathbf{A}}(g)$. Projective equivalence is an equivalence relation on $\mathcal{F}$. A task $T(f) = 0$ is said to be *projectively invariant on* $\mathcal{F}$ if for each pair of projectively equivalent points $f, g \in \mathcal{F}$,

$$T(f) = T(g) \tag{5}$$

In other words, $T(f) = 0$ is projectively invariant on $\mathcal{F}$, just in case T is constant on each equivalence class of projectively equivalent features within $\mathcal{F}$. Projectively invariant tasks are further studied in Section 4 below. On this topic see also [21].

The main result of this section is as follows.

Theorem 1 (Weak Calibration [21]) *Let G_0 be a fixed global two-camera model whose baseline lies outside of $\mathcal{V}$. A task $T(f) = 0$ is decidable on $\mathcal{C}[G_0]$ if and only if it is projectively invariant.*

In short, with a weakly calibrated stereo vision system, any projectively invariant task is verifiable with at least one encoding. Moreover, any task which is not projectively invariant is **not** verifiable with any type of encoding.

3.5 Uncalibrated Stereo Vision Systems

As it stands, Theorem 1 applies only to stereo vision systems which are weakly calibrated. But $\mathcal{C}[G_0]$ is a subset of $\mathcal{C}_{\text{uncal}}[\mathcal{V}]$. Thus any task that is decidable on $\mathcal{C}_{\text{uncal}}[\mathcal{V}]$ must also be decidable on $\mathcal{C}[G_0]$. Therefore since being projectively invariant is a necessary condition for the task $T(f) = 0$ to be decidable on $\mathcal{C}[G_0]$ it must also be a necessary condition for the task $T(f) = 0$ to be decidable on $\mathcal{C}_{\text{uncal}}[\mathcal{V}]$. We can thus state the following.

Proposition 1 *If $T(f) = 0$ is decidable on $\mathcal{C}_{\text{uncal}}[\mathcal{V}]$, then $T(f) = 0$ is projectively invariant.*

The reverse implication, namely that task invariance implies decidability on $\mathcal{C}_{\text{uncal}}$, is false. For example, suppose that the positioning objective is to make 3 point features collinear. This objective can be described mathematically by the task $T_{\text{col}}(f) = 0$, where T_{col} is the task function from $\mathcal{F}_{\text{col}} \triangleq \mathcal{V}^3$ to $\{0, 1\}$ defined by the rule

$$f \longmapsto \begin{cases} 0 & f_1, f_2, f_3 \text{ on the same line in } \mathbb{P}^3 \\ 1 & \text{otherwise} \end{cases}$$

The task $T_{\text{col}}(f) = 0$ is projectively invariant because projective transformations preserve collinearity. On the other hand this task is not decidable on $\mathcal{C}_{\text{uncal}}$. Indeed, there are camera models $C_1, C_2 \in \mathcal{C}_{\text{uncal}}$ and a pair of features $f, g \in \mathcal{F}$ at which $\bar{C}_1(f)$ and $\bar{C}_2(g)$ are equal and yet the task is accomplished at f but not at g. This can happen when all point features in the list f and the optical centers of the global camera models that determine C_1 are contained in a single 3-dimensional subspace of $\mathbb{R}^4$ and also when all point features in the list g and the optical centers of the global camera models that determine C_2 are contained in a single 3-dimensional subspace of $\mathbb{R}^4$.

4 Projectively Invariant Tasks

It was seen in the previous section that projectively invariant tasks are of special importance when dealing with sets of projective camera models. In fact, projective invariance is a necessary condition for decidability on the set of uncalibrated two-camera models, and a necessary and sufficient condition for decidability on any set of weakly calibrated two-camera models. The objective of this section is to show that every projectively invariant task can be constructed from primitive tasks by using only a small number of distinct operations. We proceed by defining these operations.

Given a task $T(f) = 0$, we call the task $\neg T(f) = 0$ specified by

$$\neg T : \mathcal{F} \to \{0, 1\}, \qquad f \longmapsto 1 - T(f),$$

the *complement* of $T(f) = 0$. Given a permutation $\pi \triangleq \{\pi(1), \pi(2), \ldots, \pi(n)\}$ of the set $\{1, 2, \ldots, n\}$, we call the task $\pi T(f) = 0$ specified by

$$\pi T : \mathcal{F} \to \{0, 1\}, \qquad f \longmapsto T(\pi f)$$

with $\pi f \triangleq \{f_{\pi(1)}, f_{\pi(2)}, \ldots, f_{\pi(n)}\}$, the *$\pi$-permutation* of $T(f) = 0$. Also, given two tasks $T_1(f) = 0$ and $T_2(f) = 0$, we call the task $(T_1 \wedge T_2)(f) = 0$ specified by

$$T_1 \wedge T_2 : \mathcal{F} \to \{0, 1\}, \qquad f \longmapsto T_1(f)T_2(f)$$

the *conjunction* of $T_1(f) = 0$ and $T_2(f) = 0$. The following proposition is straightforward to verify.

Proposition 2 *Given any set of admissible camera models $\mathcal{C}$, the following statments are true:*

1. *The complement of a task decidable on $\mathcal{C}$ is decidable on $\mathcal{C}$.*

2. *For any permutation π of $\{1, 2, \ldots, n\}$, the π-permutation of a task decidable on $\mathcal{C}$ is decidable on $\mathcal{C}$.*

3. *The conjunction of two tasks that are decidable on $\mathcal{C}$ is decidable on $\mathcal{C}$.*

Given a set of tasks $\mathcal{T}$ we say that a task $T(f) = 0$ can be *generated by* $\mathcal{T}$ if it can be obtained by applying any number of the three operations defined above to the tasks in $\mathcal{T}$. Because of Proposition 2, if all the tasks in $\mathcal{T}$ are decidable on a given set of admissible camera models $\mathcal{C}$, then every task generated by $\mathcal{T}$ is decidable on $\mathcal{C}$. It is also straightforward to show that if all tasks in $\mathcal{T}$ are projectively invariant then every task generated by $\mathcal{T}$ is also projectively invariant.

The following is the main result of this section.

Theorem 2 *Given the feature space $\mathcal{F} = \mathcal{V}^n$, the following statments are true:*

1. *For $n = 2$ any projectively invariant task can be generated by a set consisting of exactly one task, namely the one specified by the task function*

$$f \longmapsto \begin{cases} 0 & f_1 = f_2 \\ 1 & f_1 \neq f_2 \end{cases}$$

2. *For $n = 3$ any projectively invariant task can be generated by the set of tasks specified by the two task functions*

$$f \longmapsto \begin{cases} 0 & f_1 = f_2 \\ 1 & f_1 \neq f_2 \end{cases}$$

and

$$f \longmapsto \begin{cases} 0 & f_1, f_2, f_3 \text{ on the same line in } \mathbb{P}^3 \\ 1 & \text{otherwise} \end{cases}$$

3. *for $n = 4$ any projectively invariant task can be generated by the set of tasks specified by the three task functions*

$$f \longmapsto \begin{cases} 0 & f_1 = f_2 \\ 1 & f_1 \neq f_2 \end{cases}$$

$$f \longmapsto \begin{cases} 0 & f_1, f_2, f_3 \text{ on the same line in } \mathbb{P}^3 \\ 1 & \text{otherwise} \end{cases}$$

$$f \longmapsto \begin{cases} 0 & f_1, f_2, f_3, f_4 \text{ on the same plane in } \mathbb{P}^3 \\ 1 & \text{otherwise} \end{cases}$$

together with the family of tasks specified by

$$f \longmapsto \begin{cases} 0 & f_1, f_2, f_3, f_4 \text{ on the same line} \\ & \text{in } \mathbb{P}^3 \text{ with cross ratio [22] } \rho \\ 1 & \text{otherwise} \end{cases}$$

where ρ takes values in $\mathbb{R}$.

This theorem is a simple consequence of results in [21].

5 Concluding Remarks

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

- It is possible to verify that a robot positioning task has been accomplished with absolute accuracy using a weakly calibrated, noise-free, stereo vision system, if and only if the the task is invariant under projective transformations on $\mathbb{P}^3$.

- If it is possible to verify that such robot positioning task has been precisely accomplished using an *uncalibrated* stereo vision system, then the the task must be invariant under projective transformations on $\mathbb{P}^3$.

- Every projectively invariant task can be constructed from primitive tasks by using only a small number of distinct operations.

The issues addressed in this paper suggest a whole new line of inquiry within the area of visual servoing. The overriding question would seem to be something like this: Given a set of one or more imprecisely modeled cameras and a task which might be positioning, tracking or something else, under what conditions can it be decided that the task has been accomplished with precision using available images of features observed at **one or more** times? This question is more concerned with the *architecture* of a vision-based control system, than with the specific image processing and control *algorithms* which might be used to accomplish the task. Findings contributing to the answering of this question should serve to strengthen our understanding of basic issues within the emerging field of computational vision and control.

Acknowledgment: The authors thank Radu Horaud and David Kriegman for providing encouragement and useful insights contributing to this work.

References

[1] S. A. Hutchinson, G. D. Hager, and P. I. Corke. A tutorial on visual servo control. *IEEE Trans. Robot. Automat.*, 12(5):651–670, Oct. 1996.

[2] S. W. Wijesoma, D. F. H. Wolfe, and R. J. Richards. Eye-to-hand coordination for vision-guided robot control applications. *Int. J. of Robot. Res.*, 12(1):65–78, Feb. 1993.

[3] G. D. Hager, W.-C. Chang, and A. S. Morse. Robot hand-eye coordination based on stereo vision. *IEEE Contr. Syst. Mag.*, 15(1):30–39, Feb. 1995.

[4] G. D. Hager. Calibration-free visual control using projective invariance. In *Proc. Int. Conf. Comput. Vis.*, pp. 1009–1015, 1995.

[5] R. Kelly. Robust asymptotically stable visual servoing of planar robots. *IEEE Trans. Robot. Automat.*, 12(5):759–766, Oct. 1996.

[6] C. Samson, M. L. Borgne, and B. Espiau. *Robot Control: The Task Function Approach.* Number 22 in The Oxford engineering science series. Clarendon Press, 1991.

[7] B. Espiau, F. Chaumette, and P. Rives. A New Approach to Visual Servoing in Robotics. *IEEE J. of Robot. and Automat.*, 8:313–326, 1992.

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

[9] 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*, pp. 43–52, 1997.

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

[11] 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.

[12] J. Hespanha, Z. Dodds, G. D. Hager, and A. S. Morse. What can be done with an uncalibrated stereo system? In D. J. Kriegman, G. D. Hager, and A. S. Morse, editors, *The Confluence of Vision and Control*, number 237 in Lecture Notes in Control and Information Sciences, pp. 79–89. Springer-Verlag, 1998.

[13] L. Robert and O. Faugeras. Relative 3D Positioning and 3D Convex Hull Computation from a Weakly Calibrated Stereo Pair. In *Proceedings of the International Conference on Computer Vision*, pp. 540–543, Berlin, Germany, May 1993.

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

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

[16] N. Hollinghurst and R. Cipolla. Uncalibrated stereo hand-eye coordination. *Image and Vision Computing*, 12(3):187–192, Apr. 1994.

[17] S. B. Skaar, W. H. Brockman, and W. S. Jang. Three-Dimensional Camera Space Manipulation. *Int. J. of Robot. Res.*, 9(4):22–39, Aug. 1990.

[18] W.-C. Chang, J. P. Hespanha, A. S. Morse, and G. D. Hager. Task re-encoding in vision-based control systems. In *Proc. of the 36th Conf. on Decision and Contr.*, Dec. 1997.

[19] W.-C. Chang. *Vision-Based Control of Uncertain Systems.* PhD thesis, Yale University, Dec. 1997.

[20] J. Hespanha, Z. Dodds, G. D. Hager, and A. S. Morse. What can be done with an uncalibrated stereo system? In *Proc. of the Int. Conf. on Robotics and Automation*, pp. 1366–1372, May 1998.

[21] J. P. Hespanha. *Logic-Based Switching Algorithms in Control.* PhD thesis, Yale University, New Haven, CT, 1998.

[22] J. L. Mundy and A. Zisserman. *Geometric Invariance in Computer Vision.* MIT Press, 1992.