



Visual Control Using Projective Kinematics

Andreas Ruf, Frederick Martin, Bart Lamiroy, Radu Horaud

► To cite this version:

Andreas Ruf, Frederick Martin, Bart Lamiroy, Radu Horaud. Visual Control Using Projective Kinematics. 9th International Symposium on Robotics Research (ISRR '99), Oct 1999, Snowbird, United States. pp.97-104. inria-00590124

HAL Id: inria-00590124

<https://inria.hal.science/inria-00590124>

Submitted on 3 May 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Visual Servoing Using Projective Kinematics

Andreas Ruf¹, Frédérick Martin², Bart Lamiroy³ and Radu Horaud⁴

INRIA Rhône-Alpes and GRAVIR-CNRS, 655, avenue de l'Europe, 38330 Montbonnot St. Martin, FRANCE,

¹Andreas.Ruf@inrialpes.fr, ²Frederick.Martin@inrialpes.fr, ³Bart.Lamiroy@inrialpes.fr, ⁴Radu.Horaud@inrialpes.fr

Abstract

This paper presents a new approach to controlling a robot from video-feedback. In contrast to classical approaches, the entire system is modeled by means of projective geometry. For the robot's geometry in particular, we introduce a new formalism "*projective kinematics*". As a result, motions in joint- and image-space can be related without metric calibration, and a corresponding visual control law can be derived. We present experiments for projective robot calibration and visual servoing, both using uncalibrated cameras.

1. Introduction

Visual servoing of robot manipulators is a key technique where the appearance of an object in the image plane is used to control the velocity of the end-effector, aiming to guide it to a desired position. The majority of visual servoing methods proposed so far use calibrated robots and calibrated cameras. Although it has been shown that the behavior of visual control loops does not degrade too much in the presence of coarse calibration, the latter remains a complex and time-consuming procedure, requiring special-purpose devices such as theodolites and calibration jigs.

In this paper we introduce a new representation of robot manipulators and of articulated mechanisms in general. Unlike previous approaches, a non-metric description of robot motion based on projective transformations is investigated. To control a robot using projective representations rather than Euclidean ones, both the robot's direct kinematic map and the Jacobian matrix must be defined non-metrically.

The elementary joint motions that can be performed by a robot manipulator are pure rotations and translations. They give rise to corresponding projective mappings which can be parameterized as special kinds of homographies, called *projective rotations* and *projective translations*. The Lie-group properties of these projective representations of elementary motions allow us to characterize the direct kinematic map and the Jacobian matrix of a manipulator, leading to projective equivalents to the classical metric ones. They replace

in a single expression both the classical robot Jacobian (joint to end-effector velocities), and the classical Jacobian of image-based visual servoing (3D to 2D point-velocities).

Finally, we provide a practical method in order to estimate the projective kinematic model, to calculate the projective Jacobian, and we report on an experimental system for visual control using projective kinematics and a pair of uncalibrated cameras.

2. Metric and Projective Camera Spaces

In this section, we recapitulate the properties of a stereo camera. We will view it as a triangulation device which can be operated either in metric or in projective mode, depending on the degree of calibration, and perform a respective reconstruction of the three-dimensional workspace. A stereo rig being calibrated allows a metric camera space to be defined. It being uncalibrated allows only a projective camera space to be defined, in which Euclidean notions such as orthogonality, angle, and length, as well as affine notions such as parallelism, and length-ratio lack definition. But in both sorts of ambient spaces stereo projection can be expressed by a pair of projection matrices, mapping 3D world-points M or $N = (x, y, z, 1)^T$ onto 2D image-points $m = (u, v, 1)^T$, $m' = (u', v', 1)^T$. In homogeneous¹ coordinates they are:

$$\text{metric: } m \simeq [K|o] N, \quad m' \simeq [K'R'|K't'] N, \quad (1)$$

$$\text{projective: } m \simeq PM, \quad m' \simeq P'M, \quad (2)$$

where K, K' and R', t' hold an intrinsic and an extrinsic calibration, respectively, and P, P' hold an epipolar calibration, only.

Both cases, (1) and (2), describe a pair of projection rays, and solving for M or N amounts to triangulating

¹ \simeq denotes the homogeneous equality "up-to-scale".

in the respective ambient space:

$$\text{projective:} \quad \begin{bmatrix} 1 & 0 & -u \\ 0 & 1 & -v \\ 1 & 0 & -u' \\ 0 & 1 & -v' \end{bmatrix} \cdot \begin{bmatrix} \mathbf{P} \\ \mathbf{P}' \end{bmatrix} \lambda \mathbf{M} = \mathbf{o} \quad (3)$$

The so reconstructed points are related by a constant “upgrade” mapping from projective to Euclidean camera space, which encapsulates calibration of a rigid stereo rig [1]: a pair of rigidly linked cameras with fix intrinsic parameters, defining the special plane \mathbf{a}^T described later on,

$$\mathbf{N} \simeq \mathbf{H}_{PE} \mathbf{M}, \quad \mathbf{H}_{PE} \simeq \begin{bmatrix} \mathbf{K}^{-1} & \mathbf{o} \\ \mathbf{a}^T & 1 \end{bmatrix}. \quad (4)$$

Switching between ambient spaces also means switching between transformation groups, the consequences of which will be illustrated now. For robotics, coordinates and transformations being homogeneous allows displacements to be linearly represented by matrixes $\mathbf{T} \in SE(3)$ at a price no higher than adding a fourth coefficient 1 and a fourth row $(0, 0, 0, 1)$. For computer vision, coordinates being homogeneous and projective allows perspective projection² – being actually non-linear –, projective reconstruction (3), or a general projective transformation (“homographies”) to be linearly represented by matrix equations. The price is no higher than allowing a free scalar in all vectors and matrixes, e.g. all $\lambda \mathbf{M}$ solve the reconstruction (3).

Consequently, a 4×4 homography in projective space has 15 degrees-of-freedom (*dof*), thus requires **five points** to be completely defined, in contrast to three points required to define a displacement. Here, we only sketch a linear method to determine a homography given five 3D-points \mathbf{A}_p and their coordinates \mathbf{B}_p after a transformation, e.g. a motion or a change of perspective. For greater detail consult [2].

$$\begin{aligned} \lambda_p \mathbf{A}_p &= \mathbf{H} \mathbf{B}_p, & p &= 1, \dots, 5 \\ \mathbf{B} &= [\mathbf{B}_1 \ \mathbf{B}_2 \ \mathbf{B}_3 \ \mathbf{B}_4], & \mathbf{A} &= [\mathbf{A}_1 \ \mathbf{A}_2 \ \mathbf{A}_3 \ \mathbf{A}_4], \\ \mathbf{H}' &= \mathbf{A}^{-1} \mathbf{B}, & \mathbf{D} &= \text{diag}(\lambda_1, \lambda_2, \lambda_3, \lambda_4) \\ \mathbf{H} &= \gamma \mathbf{D} \mathbf{H}', & \mathbf{D} \mathbf{A}_5 &\stackrel{!}{=} \mathbf{H} \mathbf{B}_5. \end{aligned}$$

In particular, the fix homographic relation \mathbf{H}_{PE} between projective and metric camera space determines a *projective camera frame*, that can be imagined as five points in rigid linkage with the cameras. Therefore, rigid displacements \mathbf{T}_{RT} in metric camera space act on projective camera space as a *projective displacements* \mathbf{H}_{RT} , a group of homographies conjugate to $SE(3)$.

²The image plane of a pinhole camera is commonly represented as a 2D projective plane \mathbb{P}^2 , such that homogeneous coordinates $(\lambda x, \lambda y, \lambda z)^T \simeq (x/z, y/z, 1)^T$ represent a point and implicitly its perspective projection.

They inherit its Lie group structure, and corresponding Lie-algebra representations $\hat{\mathbf{H}}_{RT}$ [3]

$$\mathbf{H}_{RT} \simeq \mathbf{H}_{PE}^{-1} \mathbf{T}_{RT} \mathbf{H}_{PE}, \quad \mathbf{H}_{RT} = \exp(\hat{\mathbf{H}}_{RT}). \quad (5)$$

Naturally, points $(x, y, z, 0)^T$ in metric camera space lie on a plane $(0, 0, 0, 1)$ called “at infinity”. Finite displacements leave this topological property invariant, expressed by a fourth coefficient either 0 or 1. Switching to projective means that the infinity plane has the fix but unknown coordinates \mathbf{a}^T in \mathbf{H}_{PE} , but still a finite projective displacement always maps the plane at infinity onto itself.

3. Projective Kinematics

In this section, closed form expressions for the motion and kinematics of a robot manipulator are recapitulated, where the chosen formalism based on the exponential of 4×4 matrixes is chosen in order to introduce first metric analogs to the notions of projective kinematics introduced subsequently. In contrast to the popular Denavit-Hartenberg method, the joint-transforms are expressed relative to a single reference frame: here (metric) camera space. Therefore, switching the ambient space from metric to projective allows to elegantly and directly generalize to a projective formulation of robot kinematics. Its building-blocks are “projective rotations” which represent elementary revolutions of the joints.

For concreteness, consider a robot at the origin *zero* $\mathbf{q} = \mathbf{o}$ of its joint-angle space $\mathbf{q} \in \mathbb{R}^6$. Its joint axes, being located at points \mathbf{c}_i and having unit directions $\boldsymbol{\omega}_i$, determine twists $\hat{\mathbf{T}}_{Ri}$ [4] which act on camera space as a pure rotation $\mathbf{T}_{Ri}(\theta_i)$ ($[\]_{\times}$ is 3×3 anti-symmetric form).

$$\hat{\mathbf{T}}_R = \begin{bmatrix} [\boldsymbol{\omega}_i]_{\times} & [\boldsymbol{\omega}_i]_{\times} \mathbf{c}_i \\ \mathbf{o}^T & 0 \end{bmatrix}, \quad \mathbf{T}_{Ri} = \exp(\theta_i \hat{\mathbf{T}}_{Ri}). \quad (6)$$

Therefore, given a general joint-space motion $\mathbf{q} \in \mathbb{R}^6$, the revolutions of all joints can be concatenated to a product-of-exponentials (*poe*) [5], describing the end-effector’s displacement

$$\mathbf{T}(\mathbf{q}) = \prod_{i=1}^6 \exp(\theta_i \hat{\mathbf{T}}_{Ri}) \quad (7)$$

Obviously, for the robot haveing moved to a configuration \mathbf{Q} other than zero, the twists change say to $\hat{\mathbf{T}}_{Rj}(\mathbf{Q})$, reflecting the axis’ new locations. Thus the formulation of the robot Jacobian in (8) describing the end-effector’s velocity for the robot at \mathbf{Q} and driving joint-velocity $\dot{\mathbf{q}}$:

$$\hat{\mathbf{T}}(\dot{\mathbf{q}}) = \dot{\theta}_1 \hat{\mathbf{T}}_{R1}(\mathbf{Q}) + \dots + \dot{\theta}_6 \hat{\mathbf{T}}_{R6}(\mathbf{Q}). \quad (8)$$

Let's now consider what happens to the above formulae if we switch to projective camera space, where the interested reader is referred to [3] or [6] for greater detail. Formally, the projective robot motion writes simply as $\mathbf{H}(\mathbf{q}) = \mathbf{H}_{PE}^{-1} \mathbf{T}(\mathbf{q}) \mathbf{H}_{PE}$ (5), (7). Moreover, the fact that all revolutions are represented in camera space permits also to conjugate the inner factors of the *poe* (7). So, a new formulation of forward kinematics is obtained which is purely projective in the sense that each exponential represents a joint revolution relative to projective camera space as only ambient space:

$$\mathbf{H}(\mathbf{q}) = \mathbf{H}_{PE}^{-1} \prod_{i=1}^6 \exp(\theta_i \hat{\mathbf{T}}_{Ri}) \mathbf{H}_{PE} \quad (9)$$

$$= \prod_{i=1}^6 \mathbf{H}_{PE}^{-1} \exp(\theta_i \hat{\mathbf{T}}_{Ri}) \mathbf{H}_{PE} = \prod_{i=1}^6 \mathbf{H}_{Ri}(\theta_i) \quad (10)$$

Having a closer look at the *projective rotations* $\mathbf{H}_R(\theta)$ in (10), (see [3]) for greater detail, reveals that each of them gives rise to a one-parameter group of homographies, which is algebraically characterized by its *Jordan matrix* \mathbf{J}_R and respective *Jordan decomposition*

$$\mathbf{H}_R(\theta) = \mathbf{H}_{PE}^{-1} \mathbf{T}_J^{-1} \begin{bmatrix} \cos \theta & -\sin \theta & 0 & 0 \\ \sin \theta & \cos \theta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \mathbf{T}_J \mathbf{H}_{PE}, \quad (11)$$

$$\mathbf{H}_J^{-1} \quad \mathbf{J}_R(\theta) \quad \mathbf{H}_J, \quad (12)$$

where $\exp(\theta \hat{\mathbf{T}}_R) = \mathbf{T}_J^{-1} \mathbf{J}_R \mathbf{T}_J$ (10). We can interpret this geometrically like: first \mathbf{H}_{PE} changes from projective to metric camera space, \mathbf{T}_J changes rigidly to a frame whose z -axis is on the joint axis, and finally \mathbf{J}_R is the joint's action in this frame. However, an entirely projective representation (12) cannot separate the calibration \mathbf{H}_{PE} , since camera and kinematic parameters are “scrambled” in \mathbf{H}_J . All the same, the groups $\mathbf{H}_R(\theta)$ being conjugate forms of $\mathbf{J}_R(\theta)$, a matrix representation of $SO(2)$, form a Lie group, having a Lie algebra (14) and an exponential form (13) with closed-form solution in “Rodrigues-form”³

$$\mathbf{H}_R(\theta) = e^{\theta \hat{\mathbf{H}}_R} = \mathbf{I} + \sin \theta \hat{\mathbf{H}}_R + (1 - \cos \theta) \hat{\mathbf{H}}_R^2 \quad (13)$$

$$\hat{\mathbf{H}}_R = \mathbf{H}_J^{-1} \hat{\mathbf{J}}_R \mathbf{H}_J, \quad \hat{\mathbf{J}}_R = \begin{bmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad (14)$$

Therefore, we propose these Lie algebras as a faithful and purely projective representations of revolute joints. Fortunately, their matrix representation can be recovered from a group element other than by an unstable Jordan decompositions:⁴

$$\hat{\mathbf{H}}_R = \frac{1}{2 \sin \theta} (\mathbf{H}_R - \mathbf{H}_R^{-1}), \cos \theta = \frac{1}{2} (\text{tr } \mathbf{H}_R - 2). \quad (15)$$

³C.f. $\mathbf{I} + \sin \theta [\mathbf{w}]_{\times} + (1 - \cos \theta) [\mathbf{w}]_{\times}^2$ to go from $so(3)$ to $SO(3)$.

⁴C.f. $\frac{1}{2 \sin \theta} (\mathbf{R} - \mathbf{R}^T)$ to go from $SO(3)$ to $so(3)$.

In practice (13) is used to define and calculate the forward kinematic map and (15) is used to calculate the projective model of a joint from a single trial motion.

As in the metric case (8), an operator $\hat{\mathbf{H}}_{Ri}$ represents a joint in the spatial position it has at zero, but also encodes how this rotation acts on projective camera space. For the robot having moved to configuration \mathbf{Q} , the projective model gives also the necessary update to the axis' new locations (C.f. $\hat{\mathbf{T}}_{Ri}(\mathbf{Q})$):

$$\hat{\mathbf{H}}_{Ri}(\mathbf{Q}) = \mathbb{H}_i(\mathbf{Q}) \hat{\mathbf{H}}_{Ri} \mathbb{H}_i^{-1}(\mathbf{Q}), \quad \text{where} \quad (16)$$

$$\mathbb{H}_i(\mathbf{Q}) = \prod_{j=1}^i e^{\mathbf{Q}_j \mathbf{H}_{Rj}}, \quad \text{and } \hat{\mathbf{H}}_{Ri}(\mathbf{o}) = \hat{\mathbf{H}}_{Ri}. \quad (17)$$

In practice, this formalism is used in projective robot-calibration (section 5.1.) and in the computation of the Jacobian matrix (section 4.1.).

4. Visual Control

In this section, we derive and explain the relationship between velocities of joint-space motion, projective motion, and image motion. We will formulate a control law in projective camera space that allows for directly servoing robot joint velocities using video-feedback from a rigid but uncalibrated stereo rig.

4.1. Projective motor-image Jacobian

The Jacobian matrix will be derived as a closed form expression, depending on the robot's current configuration $\mathbf{q}(t)$ and the current coordinates $\mathbf{M}(t)$ of the considered 3D point. Its model parameters are the projective operators $\hat{\mathbf{H}}_{Ri}$ representing the joints. The projective kinematic model is update in (18) for the robot now at a general configuration \mathbf{Q} using updated operators $\hat{\mathbf{H}}_{Ri}(\mathbf{Q})$ described in (16). Since currently $\mathbf{Q} = \mathbf{q}(t)$, all exponentials obviously evaluate to identity, such that the partial derivatives of $\mathbf{H}(\mathbf{q} - \mathbf{Q})$ (18) with respect to θ_i actually equal the updated operators:

$$\mathbf{H}(\mathbf{q} - \mathbf{Q}) = \prod_{i=1}^6 \exp((\mathbf{q}_i - \mathbf{Q}_i) \hat{\mathbf{H}}_{Ri}(\mathbf{Q})), \quad (18)$$

$$\frac{d\mathbf{H}(\mathbf{q}(t) - \mathbf{Q})}{dt} = \sum_{i=1}^6 \frac{\partial \mathbf{H}(\mathbf{q}_i(t) - \mathbf{Q}_i)}{\partial \theta_i} \cdot \frac{\partial \theta_i}{\partial t},$$

$$\hat{\mathbf{H}}(\dot{\mathbf{q}}) = \sum_{i=1}^6 \dot{\theta}_i \cdot \hat{\mathbf{H}}_{Ri}(\mathbf{Q}), \quad \text{at } \mathbf{q}(t) = \mathbf{Q}. \quad (19)$$

Equation (19) is to be read: for a general articulated motion at an instant t , a “velocity” of projective end-

effector motion is obtained as instantaneous linear superposition of velocity components caused by each of the joints.

Consider now a 3D-point M undergoing the trajectory $\mathbf{H}(\mathbf{q}(t)) M$. A matrix form of M , its “velocity” in projective camera space, follows immediately, again an instantaneous linear superposition (21)

$$\dot{M} = \frac{\mathbf{H}(\mathbf{q}(t) - \mathbf{Q})M}{dt} = \frac{\partial \mathbf{H}(\mathbf{q} - \mathbf{Q})M}{\partial \mathbf{q}} \cdot \dot{\mathbf{q}}, \quad (20)$$

$$= \underbrace{[\hat{\mathbf{H}}_{R1}(\mathbf{Q})M, \dots, \hat{\mathbf{H}}_{R6}(\mathbf{Q})M]}_{\mathbf{J}_H^{4 \times 6}(\mathbf{M}, \mathbf{Q})} \cdot \dot{\mathbf{q}}. \quad (21)$$

The free scale λ in the projective reconstruction of each point (3) allows the velocity of a point to be defined likewise up to this scalar, only [3]. This unpleasant ambiguity however vanishes as soon as its image-projections $\mathbf{s} = G(\mathbf{P}\lambda M)$ (22) and yhe respective image velocity $\dot{\mathbf{s}} = dG(\mathbf{P}M(t))/dt$ are considered:

$$\mathbf{s} = G(\mathbf{m}) = \begin{pmatrix} m_1/m_3 \\ m_2/m_3 \end{pmatrix}, \quad \frac{\partial G(\mathbf{m})}{\partial \mathbf{m}} = \underbrace{\begin{bmatrix} \frac{1}{m_3} & 0 & -\frac{m_1}{m_3^2} \\ 0 & \frac{1}{m_3} & -\frac{m_2}{m_3^2} \end{bmatrix}}_{\mathbf{J}_G(\mathbf{m})} \quad (22)$$

$$\dot{\mathbf{s}} = \underbrace{\frac{\partial G(\mathbf{P}M)}{\partial \mathbf{m}} \mathbf{P}\dot{M}}_{\mathbf{J}_P^{2 \times 4}(M)}, \quad \dot{\mathbf{s}}' = \underbrace{\frac{\partial G(\mathbf{P}'M)}{\partial \mathbf{m}'} \mathbf{P}'\dot{M}}_{\mathbf{J}_{P'}^{2 \times 4}(M)}$$

Globally, the Jacobians of perspective projection \mathbf{J}_P , $\mathbf{J}_{P'}$, and that of articulated motion \mathbf{J}_H combine to the motor-image Jacobians, $\mathbf{J}(\mathbf{M}, \mathbf{Q})$ and $\mathbf{J}'(\mathbf{M}, \mathbf{Q})$, which once again show an instantaneous linear superposition of the image-velocities caused by each of the joints:

$$\dot{\mathbf{s}} = \mathbf{J}_G(\mathbf{M})\mathbf{J}_H(\mathbf{Q}, \mathbf{M}) \cdot \dot{\mathbf{q}} = \mathbf{J}(\mathbf{M}, \mathbf{Q}) \cdot \dot{\mathbf{q}}, \quad (23)$$

$$= \sum_{i=1}^6 \theta_i \cdot [\mathbf{J}_G(\mathbf{M})\hat{\mathbf{H}}_{Ri}(\mathbf{Q})M]_{2 \times 6}. \quad (24)$$

It is most important to understand this *Jacobian as a sound analytic expression* in \mathbf{Q}, \mathbf{M} . It is sound for an arbitrary configuration \mathbf{Q} of the robot and for an arbitrary point \mathbf{M} , e.g. on the tool. Therefore, it allows for global visual seroving since it is valid over the entire configuration space and allows for arbitrary shapes of the tool to be recovered on-the-fly. In contrast to existing formulations of visual seroving, the Jacobian is neither an on-line estimated linear model [7], nor an a-priori given approximation around the target [8], [9]. Above that, metric-representations, which would require a-priori calibration, were successfully eliminated in favor of projective camera space and projective kinematics.

4.2. Control law

Finally, we explicit a control law based on the projective kinematic formalism. It controls the joint-velocities of the robot arm in order to move the tool to a desired workspace position. The latter is implicitly specified by the image targets $\mathbf{s}^*, \mathbf{s}'^*$, i.e. the image projections of tool markers pre-recorded at the goal. The feedback consists of the current left- and right images \mathbf{s}, \mathbf{s}' of the markers, as they are extracted from the video stream. Along the lines of [8], a law with exponential convergence rate (Fig. 6) is obtained by applying joint velocities $\dot{\mathbf{q}}$ that impose the image-velocities $\dot{\mathbf{s}}, \dot{\mathbf{s}}'$ to be proportional to the current error vectors $\mathbf{s}^* - \mathbf{s}, \mathbf{s}'^* - \mathbf{s}'$, and so leveling it down to zero

$$\dot{\mathbf{q}} = - \begin{pmatrix} \mathbf{s}_1 - \mathbf{s}_1^* \\ \mathbf{s}'_1 - \mathbf{s}'_1^* \\ \mathbf{s}_2 - \mathbf{s}_2^* \\ \mathbf{s}'_2 - \mathbf{s}'_2^* \\ \mathbf{s}_3 - \mathbf{s}_3^* \\ \mathbf{s}'_3 - \mathbf{s}'_3^* \end{pmatrix} \begin{bmatrix} \mathbf{J}(\mathbf{M}_1, \mathbf{Q}) \\ \mathbf{J}'(\mathbf{M}_1, \mathbf{Q}) \\ \mathbf{J}(\mathbf{M}_2, \mathbf{Q}) \\ \mathbf{J}'(\mathbf{M}_2, \mathbf{Q}) \\ \mathbf{J}(\mathbf{M}_3, \mathbf{Q}) \\ \mathbf{J}'(\mathbf{M}_3, \mathbf{Q}) \end{bmatrix}^+. \quad (25)$$

The Jacobian matrix is of size $4n \times 6$, where n is the number of control points, but for the Jacobian matrix to be non-singular generally at least three points are required. Intuitively, a pair of image-velocities for a single point results through $\mathbf{J}_P, \mathbf{J}_{P'}$ in a 3D point-velocity, thus further constrains the solution through \mathbf{J}_H to a 3-*dof* linear form of joint-velocities. A second pair however, due to an additional rigid constraint, results just in a 2-*dof* pencil of 3D point-velocities, thus constrains the solution to a 1-*dof* linear form. The third pair, taking into account further two rigidity constraints, fixes the last *dof* of the solution. The used pseudo-inverse assures a best-fit to the desired “linear” image-velocities in the least-squares sense. A constant gain yields exponential convergence.

5. Experiments

A number of experiments including self-calibration of a robot (the robot moves its articulations, one by one, in front of an uncalibrated stereo rig) and a visual seroving experiment (the target position and robot trajectories are specified in the image planes) will illustrate the interest of the new paradigm outlined in this paper.

Throughout all the experiments, the input data consists of the image-points of four white markers on a black plate which is rigidly attached onto the end-effector. They are extracted from the raw images by first thresholding the intensity image and then localizing the centers of gravity with sub-pixel accuracy. Additionally, the robot’s joint-configuration $\mathbf{q}(t)$ is read

$$\text{for all } j \min_{\hat{\mathbf{H}}_{Rj}} \sum_k d\left(\mathbf{s}^{jk}, \mathbf{P} \exp((\theta^{jk} - \theta^{j1}) \hat{\mathbf{H}}_{Rj}) \mathbf{M}^{j1}\right) + d\left(\mathbf{s}'^{jk}, \mathbf{P}' \exp((\theta^{jk} - \theta^{j1}) \hat{\mathbf{H}}_{Rj}) \mathbf{M}^{j1}\right), \quad (26)$$

$$\min_{\text{all } \hat{\mathbf{H}}_{Ri}} \sum_j \sum_k d\left(\mathbf{s}^{jk}, \mathbf{P} \mathbf{H}_{RT}(\mathbf{q}^{jk}) \mathbf{M}_Q\right) + d\left(\mathbf{s}'^{jk}, \mathbf{P}' \mathbf{H}_{RT}(\mathbf{q}^{jk}) \mathbf{M}_Q\right), \quad (27)$$

$$\min_{\text{all four } \mathbf{M}_Q} \sum_j \sum_k d\left(\mathbf{s}^{jk}, \mathbf{P} \mathbf{H}_{RT}(\mathbf{q}^{jk}) \mathbf{M}_Q\right) + d\left(\mathbf{s}'^{jk}, \mathbf{P}' \mathbf{H}_{RT}(\mathbf{q}^{jk}) \mathbf{M}_Q\right) \quad (28)$$

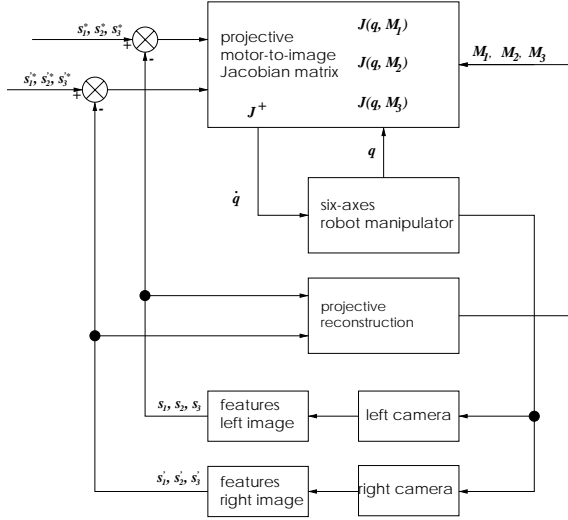


Figure 1: Schema of the projective kinematics based visual control law.

at each time t a stereo images pair is taken. For the sake of simplicity, we use single letters to generically denote all four points: \mathbf{M} for the reconstruction, \mathbf{m} , \mathbf{m}' for its theoretical, and \mathbf{s} , \mathbf{s}' for its actual image-measurements in pixel coordinates.

5.1. Recovering Projective Kinematics

The basic principle is to define an objective function which takes model parameters and physical measurements, while at the same time giving a physically meaningful error measure. Minimizing this objective then determines a best fit of model parameters. Projective quantities, such as reconstructions \mathbf{M} , allow a-priori only an “algebraic” error to be defined, which lacks a physical or statistical meaning. In our case, the measurements are the robot joint angles, and the measured pixel-coordinates of the markers’ images. The angles are considered accurate, whereas the image measurements are assumed to be disturbed by random Gaussian noise. Therefore, the main objective to be minimized is the Euclidean pixel distance $d(\mathbf{s}, \mathbf{m})$ between the actually measured image points $\mathbf{s} = (s_u, s_v)$

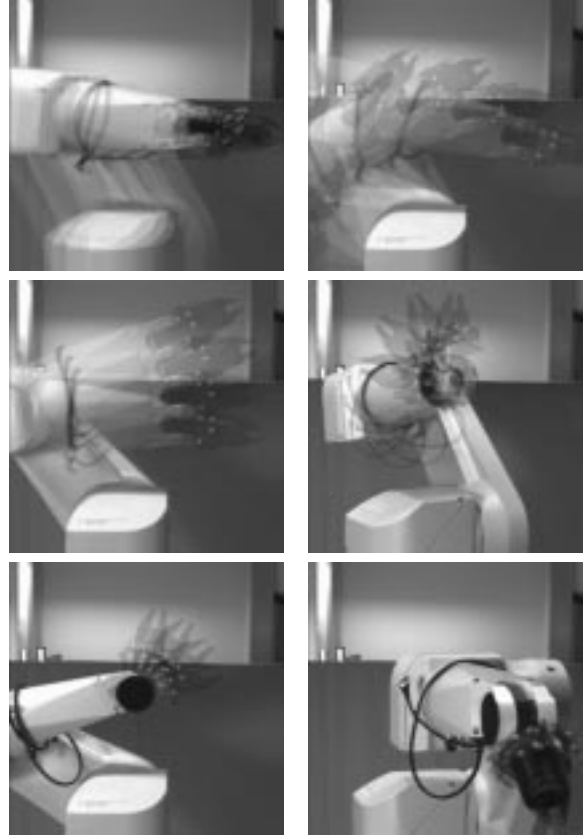


Figure 2: The image data of the six trial-motions.

and the projective image points \mathbf{m} , which are a function of the model parameters. For the sake of simplicity, we use d to denote the total error over the four image measurements.

$$d(\mathbf{s}, \mathbf{m}) = \left(\left(s_u - \frac{m_1}{m_3} \right)^2 + \left(s_v - \frac{m_2}{m_3} \right)^2 \right)^{1/2}. \quad (29)$$

For concreteness, consider the $j = 1 \dots 6$ joints and a trial motion (Fig. 2) of each joint j from configuration \mathbf{q}^{j1} to configurations \mathbf{q}^{jk} , where all but the j^{th} joint remain locked, which itself moves from angle θ^{j1} to angles θ^{jk} . The positions \mathbf{q}^{j1} are chosen in order to obtain nice circular trajectories in the image. Analogously, the reconstructed markers in the

Table 1: Results of projective robot calibration. Note, that angles and link length result from upgrading to a metric model, subject to stability matters.

residual image-error	fix 3D-points image-err [px]	free 3D-points image-err [px]
axis 1	0.55	0.13
axis 2	0.14	0.10
axis 3	0.15	0.09
axis 4	0.16	0.12
axis 5	0.16	0.10
axis 6	0.24	0.16
axis angles	Factory [$^{\circ}$]	Lab [$^{\circ}$]
axis 1, 2	90	89.72
axis 2, 3	0	0.28
axis 3, 4	90	89.71
axis 4, 5	90	90.002
axis 5, 6	90	90.10
link lengths	Factory [mm]	Lab [mm]
link 1	0	0.5522
link 2 (scale)	400	400
link 3	800	800.5457
link 4	0	0.0633
link 5	0	0.0347
link axis 4, 6	0.3	0.1895

first position are indexed as M^{j1} , their subsequent 3D-positions as M^{jk} , and their projective images as m^{jk} and m'^{jk} (2). Hence, given M^{j1} , all the subsequent positions are fully determined by equations (18) and (2), while the operators $\hat{H}_{Rj} = \hat{H}_{Ri}(q^{j1})$ are the unknown model parameters. Having also indexed the image measurements as s^{jk} and s'^{jk} , the above characterized objective function is obtained in (26). An initial guess is found by applying (15) to a single rotation. Still, the implementation of this minimization is non-trivial, since the 16 coefficients of \hat{H}_{Rj} have only 10 degrees of freedom, due to algebraic constraints and free “gauge parameters”. In [10] we devise a minimal parameterization, together with explicit solutions of the arising constraints. Note that in the experiments, the numerical implementations of equations (26)-(28) are over sets of 10 parameters which expand to a 4×4 operator.

We devise a three-step estimation process, in the first of which the joints \hat{H}_{Rj} were treated separately, and the markers M_{j1} were a-priori reconstructed, separately for each joint. In the second step, the so far independent operators are gathered to an explicit and complete projective kinematic model, expressed in terms of operators at zero $\hat{H}_{Ri} = \hat{H}_{Ri}(o)$. An initial

guess can be obtained from the results of the first step using the formalism in (16). In addition, the multiply reconstructed markers are replaced by a single a-priori reconstruction, e.g. $M_Q = M_{j1}$. Thus, step two reflects that the joints are part of one and the same articulated chain and that the four markers are rigidly linked with its end. As a results, we get a first estimate for the total of 60 parameters in the projective kinematic model after minimizing equation (27).

In an additional third step, also the 3D-coordinates of points M_Q are allowed to vary and are refined a-posteriori together with the kinematic parameters (28). This removes the bias resulting from a-priori privileging points M_Q as done in (27) and (26), and effectively improved precision (Table 1). In order to evaluate the metric dimensions “hidden” by the projective model, we have also quoted in Table 1 the results of a recent experiment on self-calibration of a Stäubli RX90 robot having a PUMA-alike geometry [11]. There, the camera parameters and kinematic parameters are separated such that the robot’s Denavit-Hartenberg parameters can be recovered up-to-scale. Conclusively, a projective calibration of the robot has been fit that agrees with the measurements down to their absolute precision, i.e. subpixel-accuracy of the markers. A metric self-calibration, known for its inherent unstability, shows good agreement with factory-given Denavit-Hartenberg parameters.



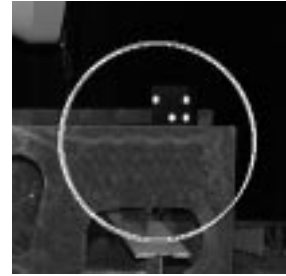
stereo rig and six-axes robot.



robot in target position.



robot in start-position.



reprojected target image.

Figure 3: Overview of system for stereo visual servoing

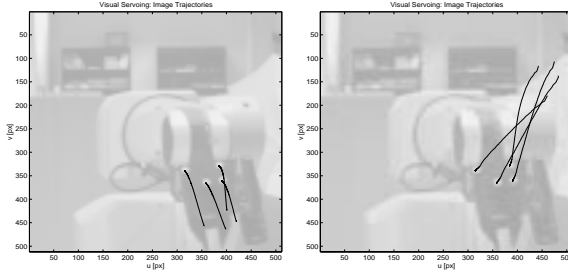


Figure 4: Image trajectories of marker points.

5.2. Visual Control using Projective Kinematics

We present results obtained with the visual servoing system under development in the VIGOR-project (ES-PRIT LTR 26247). The general setup consists of an independent stereo camera controlling a six-axes Stäubli robot RX90. The task under consideration is a 6-*dof* alignment of tool and workpiece, in our case of a welding torch or a gripper and a metal ship-part (Fig. 3). The alignment is once recorded or just calculated from CAD-data, and reprojected onto the images of the work-piece taken at task-time [12]. Also at task-time, the projective kinematics relating the robot to the stereo system is established (section 5.1.). Then running the visual servoing loop of section 4.2. performs the alignment. Thanks to the precise model estimate, equations (2) and (18) allow to predict the marker positions from joint-angle measurements so that the burden of feature extraction and tracking is greatly reduced, resulting in increased robustness and tracking range. Figure 3 gives an idea of the complexity of displacements that can be reached. Nevertheless, the resulting joint-space trajectories are still very smooth (Fig. 5), and image-trajectories are close to the expected linear behaviour, as far as rigidity permits (Fig. 4). Additionally, the gain can be largely varied (Fig. 6) without stability of control or robustness of tracking being really affected, allowing even for an adaptive gain with steep close-to-linear convergence.

6. Discussion

We have shown how to projectively represent a kinematic model and how to recover it from image measurements only. Neither a-priori knowledge, nor metric information is required. This representation of robot kinematics is predestinated for visual control of a robot in order to perform basic reaching, grasping, or manipulation tasks. In comparison to previous work, our new approach has a number of advantages. Visual servoing can be along trajectories covering the whole

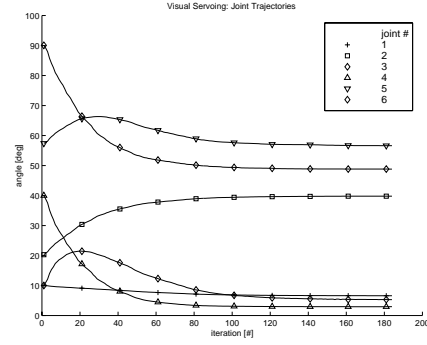


Figure 5: Joint-space trajectories.

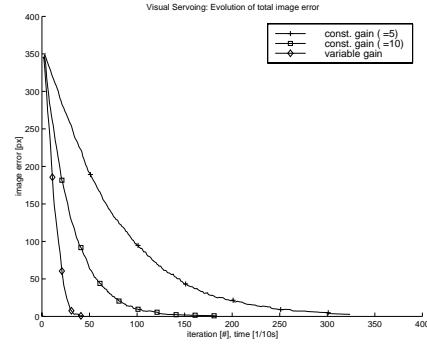


Figure 6: Exponential convergence at different gains.

visible configuration space. The sound relationship between joint and image-space prepares the ground for more intelligent control laws that incorporate joint-space reasoning. So, avoidance of joint-limits, singularities, or “workspace limits” (collisions, occlusions projected onto joint-space) become feasible, in addition to optimal control techniques, such as shortest-path, shortest-time, etc. Last but not least, this formalism opens perspectives for the design of “visual mechanisms”, meaning that a camera is assuring by means of a visual control loop that a visual kinematic constraint is respected. For instance the control of a robot arm reaching through a hole, manipulating without occluding the tele-operators’ view, or co-operating with a mobile observer robot.

Acknowledgements

The authors are very grateful towards the European Commission for financial support through the Marie-Curie fellowship FMBICT972281, and the Esprit LTR project VIGOR 26247.

References

- [1] R. Horaud and G. Csurka. Self-calibration and Euclidean reconstruction using motions of a stereo rig. In *Proc. of the 6th International Conference on Computer Vision (ICCV'98)*, pages 96–103, Bombay, India, January 1998. IEEE.
- [2] G. Csurka, D. Demirdjian, and R. Horaud. Finding the collineation between two projective reconstructions. *Computer Vision and Image Understanding*, 75(3):260–268, September 1999.
- [3] Andreas Ruf and Radu Horaud. Rigid and articulated motion seen with an uncalibrated stereo rig. In *Proc. of the 7th International Conference on Computer Vision (ICCV'99)*, pages 789 – 796, Korfu, Greece, September 1999.
- [4] R.M. Murray, Z. Li, and S.S. Sastry. *A Mathematical Introduction to Robotic Manipulation*. CRC Press, 1994.
- [5] R. W. Brockett. Robotic manipulators and the product of exponentials formula. In *Proc. of International Symposium on Mathematical Theory of Networks and Systems*, number 58 in Lecture Notes in control and information sciences, pages 120–129. Springer, June 1983.
- [6] 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.
- [7] M. Jagersand, O. Fuentes, and R. Nelson. Acquiring visual-motor models for precision manipulation with robot hands. In *Proc. of 4th European Conference on Computer Vision (ECCV'96)*, volume II, pages 603–612, Cambridge, England, April 1996. Springer Verlag.
- [8] B. Espiau, F. Chaumette, and P. Rives. A new approach to visual servoing in robotics. *IEEE Trans. on Robotics and Automation*, 8(3):313–326, June 1992.
- [9] G.D. Hager. A modular system for robust positioning using feedback from stereo vision. *IEEE Trans. on Robotics and Automation*, 13(4):582 – 595, August 1997.
- [10] Andreas Ruf and Radu Horaud. Projective rotations applied to a pan-tilt stereo head. In *Proc. of Intl. Conf. on Computer Vision and Pattern Recognition (CVPR' 1999)*, volume II, pages 144–150, Fort Collins, CO, June 1999.
- [11] Frédéric Martin. Identification numérique du modèle métrique et non-métrique d'un robot par vision stéréoscopique. rapport de "dea", ENSIMAG, UJF, June 1999.
- [12] R. Horaud, F. Dornaika, and B. Espiau. Visually guided object grasping. *IEEE Transactions on Robotics and Automation*, 14(4):525–532, August 1998.