Rigid and Articulated Motion Seen with an Uncalibrated Stereo Rig

Andreas Ruf, Radu Horaud

Rigid and Articulated Motion Seen with an Uncalibrated Stereo Rig

Andreas Ruf * Radu Horaud*

Andreas.Ruf@inrialpes.fr Radu.Horaud@inrialpes.fr
GRAVIR-IMAG, INRIA Rhone-Alpes
655, avenue de l'Europe
38330 Montbonnot St.Martin, France

Abstract

This paper establishes a link between uncalibrated stereo vision and the motion of rigid and articulated bodies. The variation in the projective reconstruction of a dynamic scene over time allows an uncalibrated stereo rig to be used as a faithful motion capturing device. We introduce an original theoretical framework – projective kinematics – which allows rigid and articulated motion to be represented within the transformation group of projective space. Corresponding projective velocities are defined in the tangent space. Most importantly, these projective motions inherit the Lie-group structure of the displacement group.

These theoretical results lead immediately to non-metric formulations of visual servoing, tracking, motion capturing and motion synthesis systems, that no longer require the metric geometry of a stereo camera or of the articulated body to be known. We report on such a non-metric formulation of a visual servoing system and present simulated experimental results.

1 Introduction

In this paper we address the problem of representing and controlling the motion of robot manipulators or, more generally, of articulated mechanical devices using image measurements and continuous feedback from an uncalibrated stereo camera pair - a stereo rig. It is well-known that such a camera pair can recover the 3D projective structure of an observed object from point-to-point matches between the two images - this result has been simultaneously shown by Faugeras [5] and Hartley [8]. The relationship between the projective structure thus recovered has been further investigated by a number of authors: Zisserman revealed how to upgrade projective structure to metric if the stereo rig undergoes a general displacement [18] and to affine if the rig undergoes planar motion [1], Devernay & Faugeras revealed important algebraic properties associated with the similarity between rigid and projective motion [3]. Horaud & Csurka devised a closed-form solution for computing the internal parameters from general motion [11], and from a single planar motion [2].

In parallel other authors investigate the relationship between a stereo rig and the visual control of a robot manipulator. Hollinghorst and Cipolla considered an affine approximation of the perspective camera model [10] and Hager et al. investigated both, the representation of alignments using projective invariants and the sensibility of stereo-based visual control in the presence of coarse calibration [7].

Therefore, based on the current state-of-the-art, a possible solution for stereo-guided visual servoing using uncalibrated cameras would be to calibrate the stereo rig first and then to use the classical Euclidean robot to establish a visual control law [4].

The work described in this paper takes a different approach, where neither internal camera parameters, nor Euclidean robot representations are required anymore. Since the motions of a robot are combinations of elementary rotations (revolute joints) or translations (prismatic joints), we introduce two special projective transformations, namely, *projective rotations* and *projective translations*. These transformations are special parameterizations of 4×4 homographies that arise when a weakly calibrated stereo rig observes either rotations or translations. We reveal the Lie-group structure of these transformations and we analyze the action of these groups onto 3D projective space. We explicitly devise the projective tangent operators associated with them. Next, we establish the projective forward kinematics and inverse models of a robot manipulator. Note that this model applies to articulated motion in general. Moreover, we devise the projective velocity associated with such an articulated motion and we explicitly

*Submitted to ICCV 99. 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.

derive the relationship between the joint velocities and the $2D$ velocities observed in the two images. Finally, we introduce the concept of non-metric (projective) visual servoing, where no Euclidean representations (neither for the cameras nor for the robot) are required.

1.1 Notation

Bold type $\mathbf{H}, \mathbf{T}$ is used for matrices, bold italic $\boldsymbol{M}, \boldsymbol{k}$ for vectors, calligraphic $\mathcal{F}, \mathcal{P}$ for frames, and Roman a, b, θ for scalars, angles etc. Vectors $\boldsymbol{k}$ are column vectors, and row vectors are written by the transpose $\boldsymbol{k}^T$. " $\simeq$" denotes equality up to scale.

We give without proof a well-known matrix identity

$$\exp(\mathbf{A}^{-1}\mathbf{X}\mathbf{A}) = \mathbf{A}^{-1}\exp(\mathbf{X})\mathbf{A} \tag{1}$$

2 Projective reconstruction

A calibrated stereo rig is modeled as two pinhole cameras which have intrinsic parameters $\mathbf{K}$, $\mathbf{K}'$, and which are rigidly linked by $(\mathbf{R}', \boldsymbol{t}')$. A point $\boldsymbol{N}$ in Euclidean space projects onto the points $\boldsymbol{m}$ and $\boldsymbol{m}'$ in the left and right projective image plane $\mathbb{P}^2$ [6]. Solving the projection constraints (2) for $\boldsymbol{N}$ yields a *Euclidean reconstruction* in the *Euclidean camera frame* $\mathcal{E}$

$$\boldsymbol{m} \simeq \left[\mathbf{K}|\boldsymbol{0}\right]\boldsymbol{N}, \qquad \boldsymbol{m}' \simeq \left[\mathbf{K}'\mathbf{R}'|\mathbf{K}'\boldsymbol{t}'\right]\boldsymbol{N}. \tag{2}$$

A weakly calibrated stereo rig is modeled as a pair of cameras whose epipolar geometry $\mathbf{F}$ is supposed to be known [9]. This allows two projection matrices $\mathbf{P} = [\mathbf{I}|\boldsymbol{0}]$ and $\mathbf{P}'$ to be calculated, such that the corresponding *projective reconstruction*, solving

$$\boldsymbol{m} \simeq \mathbf{P}\boldsymbol{M}, \qquad \boldsymbol{m}' \simeq \mathbf{P}'\boldsymbol{M} \tag{3}$$

for $\boldsymbol{M}$, is relative to a *projective camera frame* $\mathcal{P}$. The frame $\mathcal{P}$ is defined by $\mathbf{F}$ and can be thought of as five rigid points attached to the stereo rig [8], [13].

If the stereo rig remains fixed, i.e. constant $\mathbf{K}$, $\mathbf{K}'$, and $(\mathbf{R}', \boldsymbol{t}')$, the so-called *projective-Euclidean link (PE-link)* $\mathbf{H}_{PE}$ is the well-defined homography

$$\boldsymbol{N} \simeq \mathbf{H}_{PE}\boldsymbol{M}, \quad \mathbf{H}_{PE} \simeq \begin{bmatrix} \mathbf{K}^{-1} & \boldsymbol{0} \\ \boldsymbol{a}^T & 1 \end{bmatrix}, \tag{4}$$

that upgrades the projective reconstruction (3) in $\mathcal{P}$ to a Euclidean one (2) in $\mathcal{E}$. The PE-link encapsulates the complete geometry of the stereo camera, and recovering $\mathbf{H}_{PE}$ amounts to metric calibration [18], [3].

Throughout this paper, we consider a stereo camera observing and reconstructing a dynamic scene containing rigid and articulated motion. We assume a fixed but unknown stereo geometry $\mathbf{H}_{PE}$ (4).

3 Metric rigid motion

At a time instant t, generic points N on a rigid body have coordinates $\boldsymbol{N}(t) = [X(t), Y(t), Z(t), 1]^T$ with respect to the Euclidean camera frame $\mathcal{E}$. The trajectory of a *rigid motion* is described by

$$\boldsymbol{N}(t) = \mathbf{T}_{RT}(t)\boldsymbol{N}(0), \qquad \mathbf{T}_{RT}(t) = \begin{bmatrix} \mathbf{R}(t) & \boldsymbol{t}(t) \\ \boldsymbol{0}^T & 1 \end{bmatrix}, \tag{5}$$

where $\mathbf{T}_{RT}(t)$ is a differentiable trajectory in the displacement group $SE(3)$. Therefore, *spatial point-velocities* are defined by the tangent

$$\dot{\boldsymbol{N}}(t) = [\dot{X}, \dot{Y}, \dot{Z}, 0]^T = \dot{\mathbf{T}}_{RT}(t)\boldsymbol{N}(0), \tag{6}$$

where $\dot{\boldsymbol{N}}$ as well as $\boldsymbol{N}$ are relative to the frame $\mathcal{E}$. Moreover, the Lie-group structure of $SE(3)$ allows a *spatial body-velocity* to be defined by the tangent operator

$$\hat{\mathbf{T}}_{RT}(t) = \dot{\mathbf{T}}_{RT}\mathbf{T}_{RT}^{-1}, \tag{7}$$

which indeed yields the motion tangent $\dot{\mathbf{T}}_{RT}(t)$ and the point-velocity of $\boldsymbol{N}(t)$ by simple left-multiplication

$$\dot{\boldsymbol{N}}(t) = \hat{\mathbf{T}}_{RT}(t)\boldsymbol{N}(t) = \dot{\mathbf{T}}_{RT}(t)\boldsymbol{N}(0). \tag{8}$$

The instantaneous body-velocity as written in (7) is a matrix representation of the Lie-Algebra $se(3)$ of $SE(3)$. It has generally the form

$$\hat{\mathbf{T}}_{RT}(t) = \begin{bmatrix} 0 & -\omega_z & \omega_y & v_x \\ \omega_z & 0 & -\omega_x & v_y \\ -\omega_y & \omega_x & 0 & v_z \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{9}$$

which is geometrically interpreted as follows. The vectos $\boldsymbol{w} = [\omega_x, \omega_y, \omega_z]^T$ and $\boldsymbol{v} = [v_x, v_y, v_z]^T$ are the instantaneous angular- and linear velocity of the frame $\mathcal{E}$ as if it were rigidly moving with the body at instant t [14]. This has to be distinguished from the common usage of kinematic screws, where they represent the velocity of a rigid body in its own, body-fixed frame and not in a spatial reference frame; here the frame $\mathcal{E}$.

4 Rigid projective motion

In this section, we no longer assume a calibrated stereo camera and the Euclidean frame $\mathcal{E}$, but consider an uncalibrated rig. The reconstruction is then projective, relative to the projective camera frame $\mathcal{P}$. Nevertheless, the PE-link (4) between the points $\boldsymbol{N}(t)$ and their projective coordinates $\boldsymbol{M}(t)$ allows us to define the notions of *projective kinematics*.

4.1 Projective displacements

The projective coordinates $\boldsymbol{M}(t)$ and $\boldsymbol{M}(0)$ of rigidly moving points are related by a 4×4 homography matrix $\mathbf{H}_{RT} \in PGL(3)$ – the three-dimensional projective group – which can be calculated from at least 5 such point and their correspondences

$$\boldsymbol{M}(t) = \mathbf{H}_{RT}(t)\boldsymbol{M}(0). \tag{10}$$

Such homographies are algebraically similar to a displacement by the PE-link up to the scalar γ

$$\mathbf{H}_{RT}(t) = \gamma\, \mathbf{H}_{PE}^{-1}\, \mathbf{T}_{RT}(t)\, \mathbf{H}_{PE}. \tag{11}$$

By similarity, the trace and determinant determine γ and allow normalizing $\mathbf{H}_{RT}$ to unit scale

$$\gamma = sign(trace(\mathbf{H}_{RT}))(|\mathbf{H}_{RT}|)^{1/4}. \tag{12}$$

We refer to normalized homographies $\mathbf{H}_{RT}$ that have the form (11) and $\gamma = 1$ as *projective displacements (p-displacement)*. It is straight forward to see that they form a subgroup of $PGL(3)$, modulo a fix matrix $\mathbf{H}_{PE}$ embodying the stereo geometry. What is more, thanks to normalization, equation (11) is a well-defined differentiable homeomorphism from $SE(3)$ into $PGL(3)$, such that the p-displacements inherit this structure and become a subgroup, moreover a Lie-subgroup in $PGL(3)$. They are simply a matrix representation of $SE(3)$ in $PGL(3)$, which we will denote as $PSE(3)$.

4.2 Projective body-velocity

In analogy with equation (7), a notion of *projective body-velocities* can be defined as

$$\hat{\mathbf{H}}_{RT}(t) = \dot{\mathbf{H}}_{RT}\mathbf{H}_{RT}^{-1}. \tag{13}$$

Geometrically, these are spatial velocities with respect to the projective frame $\mathcal{P}$. Algebraically, they are a matrix representation of $PSE(3)$'s Lie-Algebra, denoted $pse(3)$, and generally have the form

$$\hat{\mathbf{H}}_{RT}(t) = \mathbf{H}_{PE}^{-1} \begin{bmatrix} 0 & -\omega_z & \omega_y & v_x \\ \omega_z & 0 & -\omega_x & v_y \\ -\omega_y & \omega_x & 0 & v_z \\ 0 & 0 & 0 & 0 \end{bmatrix} \mathbf{H}_{PE}. \tag{14}$$

4.3 Orbit of projective points

From the PE-link(4), a scale factor ρ can be identified in the homogeneous coordinates $\boldsymbol{M}$ of a projective point, that is related to the unknown vector $[\boldsymbol{a}^T 1]$ of the plane at infinity.

$$\rho\boldsymbol{N} = \mathbf{H}_{PE}\boldsymbol{M}_\rho, \qquad \rho = [\boldsymbol{a}^T 1]\boldsymbol{M}_\rho. \tag{15}$$

All vectors $\boldsymbol{M}_\rho$ have ρ as an implicit property, which we call *height*. Neither $\boldsymbol{a}^T$ for the height are known, but the *height is invariant* under p-displacements.

$$\mathbf{H}_{RT}\boldsymbol{M}_\rho = \mathbf{H}_{PE}^{-1}\mathbf{T}_{RT}\mathbf{H}_{PE}\boldsymbol{M}_\rho = \mathbf{H}_{PE}^{-1}\mathbf{T}_{RT}\rho\boldsymbol{N} = \rho\mathbf{H}_{PE}^{-1}\boldsymbol{N}' = \boldsymbol{M}'_\rho.$$

So $\boldsymbol{M}_\rho$'s orbit under the action of $PSE(3)$ lies entirely within the hyperplane $[\boldsymbol{a}^T 1]^T\boldsymbol{M}_\rho = \rho$ of $\mathbb{R}^4$. Hence ρ is its *orbital height*.

4.4 Projective point-velocity

For a point on such a rigid orbit $\boldsymbol{M}_\rho(t)$, a *projective point-velocity* is well-defined by $\dot{\boldsymbol{M}}_\rho(t)$. It is related to its metric velocity (6) up to its fix but unknown height

$$\rho\dot{\boldsymbol{N}} = \mathbf{H}_{PE}\dot{\boldsymbol{M}}_\rho. \tag{16}$$

Finally, applying a body velocity $\hat{\mathbf{H}}_{RT}(t)$ to a point-vector $\boldsymbol{M}_\rho(t)$, yields again its point-velocity

$$\begin{aligned} \hat{\mathbf{H}}_{RT}(t)\boldsymbol{M}_\rho(t) &= \dot{\mathbf{H}}_{RT}(t)\boldsymbol{M}_\rho(0) \\ &= \mathbf{H}_{PE}^{-1}\, \dot{\mathbf{T}}_{RT}(t)\, \mathbf{H}_{PE}\, \rho\, \mathbf{H}_{PE}^{-1}\boldsymbol{N}(0) \\ &= \rho\mathbf{H}_{PE}^{-1}\boldsymbol{N}(0) = \dot{\boldsymbol{M}}_\rho(t). \end{aligned}$$

5 Articulated projective motion

In this section, the projective motions arising from revolute and prismatic joints are derived, then the projective motions of articulated chains are composed from these projective joint motions.

5.1 Projective revolute joints

Projective revolute joints are represented by means of a projective formulation of pure rotational twists [14]. Geometrically, a *general pure rotation* is a revolution around an axis at a general position in space. It can no longer be represented by $\mathbf{R}(\theta) \in SO(3)$, since this would constrain the axis to go through the origin. Instead, it has a linear representation in $SE(3)$

$$\mathbf{T}_R(\theta) = \mathbf{T}^{-1} \begin{bmatrix} \mathbf{R}(\theta) & \boldsymbol{0} \\ \boldsymbol{0}^T & 1 \end{bmatrix} \mathbf{T}, \tag{17}$$

where the position of the joint is specified by means of its displacement $\mathbf{T}$ away from the origin.

We define a *projective rotation (p-rotation)* to be

$$\mathbf{H}_R(\theta) = \mathbf{H}_{PE}^{-1}\, \mathbf{T}_R(\theta)\, \mathbf{H}_{PE}, \tag{18}$$

the p-displacement (11) corresponding to the general pure rotation $\mathbf{T}_R(\theta)$.[1] A similarity relation like (18) has the *Jordan decomposition of a p-rotation*

$$\mathbf{H}_R(\theta) = \mathbf{H}_{PE}^{-1}\mathbf{T}^{-1} \quad \begin{bmatrix} \cos\theta & -\sin\theta & 0 & 0 \\ \sin\theta & \cos\theta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \mathbf{T}\mathbf{H}_{PE} \quad (19)$$

$$\mathbf{H}_J^{-1} \qquad \mathbf{J}_R(\theta) \qquad \mathbf{H}_J, \quad (20)$$

Geometrically, this is achieved by writing $\mathbf{R}(\theta)$ in (17) w.l.o.g.[2]as a rotation around the z-axis. Algebraically, any p-rotation is similar to the Jordan matrix $\mathbf{J}_R$, which implies its *rotation angle* to be determined by

$$\cos(\theta) = 1/2(trace\ \mathbf{H}_R - 2). \quad (21)$$

However, there is a whole family $\mathbf{H}_J = \mathbf{C}_R\mathbf{T}\mathbf{H}_{PE}$ of decompositions like (20), which is spanned by the commutator group $\mathbf{C}_R$ of the Jordan matrix $\mathbf{J}_R$.

$$\mathbf{C}_R = \begin{bmatrix} a & -b & 0 & 0 \\ b & a & 0 & 0 \\ 0 & 0 & c & d \\ 0 & 0 & e & f \end{bmatrix}, \qquad rank(\mathbf{C}_R) = 4. \quad (22)$$

This ambiguity is undesirable, since $\mathbf{H}_J$ encapsulates the projective geometry of the joint. Nevertheless, let $\boldsymbol{h}_j^T$ and $\boldsymbol{k}_i$ denote the j^{th} row of some $\mathbf{H}_J$ and the i^{th} column of some $\mathbf{H}_J^{-1}$.

The Jordan matrix $\mathbf{J}_R(\theta)$ is the $SE(3)$ representation of z-rotations, a 1D Lie-group of SE. It is generated over the trivial Lie-Algebra with elements $\theta\hat{\mathbf{J}}_R$:

$$\mathbf{J}_R = \exp(\theta\hat{\mathbf{J}}_R), \qquad \hat{\mathbf{J}}_R = \begin{bmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \in se(3). \quad (23)$$

As soon as (23) is substituted in (20), we have a differentiable homomorphism from $SE(3)$ to $PSE(3)$. Therefore, the p-rotations form a family of 1D Lie-subgroups of $PSE(3)$, modulo an axis position and a fixed stereo geometry, both embodied by $\mathbf{H}_J$.

Formally applying the matrix identity (1) to $\mathbf{H}_J^{-1}\exp(\theta\hat{\mathbf{J}}_R)\mathbf{H}_J$, yields the *exponential form* of a p-rotation and its *generator* $\hat{\mathbf{H}}_R$:

$$\mathbf{H}_R = \exp(\theta\hat{\mathbf{H}}_R), \qquad \hat{\mathbf{H}}_R = \mathbf{H}_J^{-1}\hat{\mathbf{J}}_R\mathbf{H}_J. \quad (24)$$

So, the *generator* $\hat{\mathbf{H}}_R$ of a p-rotation can be writes as

$$\hat{\mathbf{H}}_R = \boldsymbol{k}_2\boldsymbol{h}_1^T - \boldsymbol{k}_1\boldsymbol{h}_2^T, \quad (25)$$

Despite of the ambiguity of $\mathbf{H}_J$ (22), $\hat{\mathbf{H}}_R$ itself is unique, since $\mathbf{C}_R$ commutates also with $\hat{\mathbf{J}}_R$ (24), (23).

[1] a joint can be positioned w.l.o.g. relative to frame $\mathcal{E}$

[2] since $\mathbf{T}$ allows for free reorientation of the axis

We now prove the *Rodriguez form* of p-rotations

$$\mathbf{H}_R(\theta) = \mathbf{I} + sin\theta\hat{\mathbf{H}}_R + (1 - cos\theta)\hat{\mathbf{H}}_R^2, \quad (26)$$

Notice that $-\hat{\mathbf{H}}_R^2 = \boldsymbol{k}_1\boldsymbol{h}_1^T + \boldsymbol{k}_2\boldsymbol{h}_2^T$ since $\boldsymbol{k}_i\boldsymbol{h}_j^T = \delta_{ij}$. Substitute $(\mathbf{J}_R - \mathbf{I}) + \mathbf{I}$ into (20), expand in the sum, and collect the rows and columns corresponding to $sin\theta$ and $(cos\theta - 1)$ to obtain (26) – q.e.d.

On the one hand, given the joint angle θ, the Rodriguez form (26) allows us to analytically calculate the projective motion of a revolute joint. Its projective geometry is uniquely represented by the generator $\hat{\mathbf{H}}_R$ of the corresponding Lie-subgroup of $PSE(3)$.

On the other hand, given a single p-rotation $\mathbf{H}_R$, calculated from a trial motion of the revolute joint (Fig. 3), the *logarithm of a p-rotation*

$$\hat{\mathbf{H}}_R = \frac{1}{2\sin\theta}\,(\mathbf{H}_R - \mathbf{H}_R^{-1}), \quad (27)$$

allows us to recover its generator in $pse(3)$. C.f. $\frac{1}{2\sin\theta}\,(\mathbf{R} - \mathbf{R}^T)$ to go from $SO(3)$ to $so(3)$.

5.2 Projective prismatic joints

The projective motion of prismatic joints is derived along the lines of section 5.1. It is only summarized here and can be found in greater detail in [15]. Starting from a pure translation of $\tau = 1$ along the z-axis, the *Jordan decompositions of p-translations* have the form

$$\hat{\mathbf{H}}_T(q) = \mathbf{H}_{PE}^{-1}\mathbf{T} \quad \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \tau \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \mathbf{T}\mathbf{H}_{PE} \quad (28)$$

$$\mathbf{H}_J^{-1} \qquad \mathbf{J}_T(q) \qquad \mathbf{H}_J, \quad (29)$$

They form a 1D Lie-subgroup of $PSE(3)$, modulo the translation direction and the affine stereo geometry, both embodied by $\mathbf{H}_J$.

The *generator* $\hat{\mathbf{H}}_T$ of p-translations is the product of the coordinate vectors $\boldsymbol{k}_3$, their vanishing point, and $\boldsymbol{h}_4^T$, the plane at infinity,

$$\hat{\mathbf{H}}_T = \mathbf{H}_J^{-1}\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}\mathbf{H}_J, \qquad \hat{\mathbf{H}}_T = \boldsymbol{k}_3^T\boldsymbol{h}_4. \quad (30)$$

Their *exponential form* and *logarithm* are trivial, where unit τ has the length of the trial motion $\mathbf{H}_T$

$$\mathbf{H}_T(q) = \exp(q\hat{\mathbf{H}}_T) = \mathbf{I} + \tau\hat{\mathbf{H}}_T, \quad \hat{\mathbf{H}}_T = \mathbf{H}_T - \mathbf{I}.$$

5.3 Projective motion of articulated chains

Projective articulated motion is expressed by means of a projective generalization of the twist model of articulated chains and the product-of-exponentials expansion of their zero-reference kinematics [14],[17], In particular, we concisely prove this projective model of the kinematics of articulated chains using exclusively our projective kinematics.

Consider an articulated chain whose elements are serially linked by either revolute or prismatic joints. Consider further a stereo rig rigidly fixed with respect to one end of the chain. This end is called the *base* of the chain, whereas the opposite end is called its *tip*. The n joints are indexed with $i = 1 \dots n$ in base-to-tip order, and the element linking joint i with joint $i+1$ is indexed with i.

This convention allows us to uniformly cover both, the independent-eye case, where the robot and the stereo rig are independently but rigidly installed in the workspace, and the eye-in-hand case, where the rig is rigidly mounted on the robot hand, which is now taken as the base of the chain. To help intuition, focus on a six-axes robot manipulator moving in front of a stereo rig.

A vector $\boldsymbol{q} = [q_1 \dots q_n]^T$ of joint variables q_i, which stand for θ_i or τ_i depending whether the i^{th} joint is revolute or prismatic, describes the *configuration* of the chain relative to its *zero-reference*: an arbitrary configuration chosen as the origin $\boldsymbol{q} = 0$ of joint-space. Now, assign to generic points on the i^{th} element the coordinates $\boldsymbol{M}_i(\boldsymbol{0})$ in zero-reference, and $\boldsymbol{M}_i(\boldsymbol{q})$ in configuration $\boldsymbol{q}$. The p-displacement $\mathbb{H}_i(\boldsymbol{q})$ of the i^{th} element, and such for all the elements, represent the *articulated projective motion* for $\boldsymbol{q}$ (Fig. 1)

$$\boldsymbol{M}_i(\boldsymbol{q}) = \mathbb{H}_i(\boldsymbol{q})\boldsymbol{M}_i(\boldsymbol{0}). \tag{31}$$

The projective motion of a single articulation $\mathbf{H}_i(q_i)$ is directly expressed using the respective projective joint $\hat{\mathbf{H}}_i(q_i)$, since in zero-reference each joint has a specific position in space. More precisely, moving only the i^{th} joint to q_i results in either a p-rotation $\mathbf{H}_{Ri}(q_i)$ or a p-translation $\mathbf{H}_{Ti}(q_i)$, that is generated either by $\hat{\mathbf{H}}_{Ri}$ or by $\hat{\mathbf{H}}_{Ti}$, depending on the type of joint. The generators encapsulate the projective geometry of the joint in the zero-reference.

$$\boldsymbol{M}_i(q_i) = \mathbf{H}_i(q_i)\, \boldsymbol{M}_i(\boldsymbol{0}). \tag{32}$$

$$\mathbf{H}_i(q_i) = \exp(q_i\hat{\mathbf{H}}_i), \;\; \hat{\mathbf{H}}_i = \hat{\mathbf{H}}_{Ri} \text{ or } \hat{\mathbf{H}}_i = \hat{\mathbf{H}}_{Ti}, \tag{33}$$

Given an articulated chain and its projective joints $\hat{\mathbf{H}}_i$ in zero-reference, the p-displacementof the i^{th} element can be written as a product-of-exponentials (Fig. 1)

$$\mathbb{H}_i(\boldsymbol{q}) = \exp(q_1\hat{\mathbf{H}}_1)\cdots\exp(q_i\hat{\mathbf{H}}_i). \tag{34}$$

The argument is a simple induction. The points on the 1^{st} element are affected only by the motion of the 1^{st} joint: $\boldsymbol{M}_1(\boldsymbol{q}) = \mathbf{H}_1(q_1)\boldsymbol{M}_1(\boldsymbol{0})$. Thus the hypothesis is that the elements preceeding the i^{th} one move like (34). Now, points on the i^{th} element move like $\boldsymbol{M}_i(q_i) = \mathbf{H}_i(q_i)\boldsymbol{M}_i(\boldsymbol{0})$ if only the i^{th} joint is actuated. And after the i^{th} joint is locked to q_i, the points $\boldsymbol{M}_i(q_i)$ on the i^{th} element move rigidly with element $i-1$. Conclusively, we can apply the hypothesis for $i-1$, to see that elementary joint motions have to be left-multiplied in tip-to-base order (34). – q.e.d.

Since each $\mathbf{H}_i(q_i)$ in (34) is expressed w.r.t. to frame $\mathcal{P}$, their conjugate forms (19), (28) can be substituted into (34). Since the inner pairs $\mathbf{H}_{PE}\mathbf{H}_{PE}^{-1}$ cancel out, $\mathbb{H}_i(\boldsymbol{q})$ is in fact the p-displacementof the i^{th} element's articulated motion.

$$\mathbb{H}_i(\boldsymbol{q}) = \mathbf{H}_{PE}^{-1}\exp(q_1\hat{\mathbf{T}}_1)\cdots\exp(q_i\hat{\mathbf{T}}_i)\mathbf{H}_{PE}. \tag{35}$$

Here, the $\hat{\mathbf{T}}_i$ in the Euclidean POE are the twists (14) representing the zero-reference of each joint w.r.t. $\mathcal{E}$. It is essentially this particular modelling of the chain's geometry, that allows us to directly exploit the correspondence of the Euclidean and projective motion.

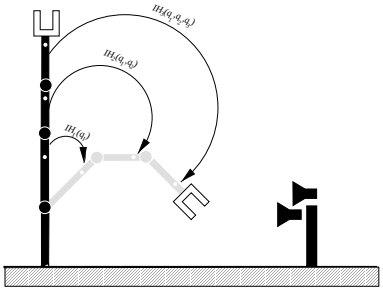

Figure 1: Zero-reference model of articulated chain.

5.4 Projective velocity of articulated chain

The instantaneous motion of an articulated body along a differentiable trajectory $\boldsymbol{q}(t)$ in joint space is characterized by the projective body velocities $\hat{\mathbb{H}}_i$ of its elements. Consider a time instant t, at which the chain instantaneously has the configuration $\boldsymbol{q} = \boldsymbol{q}(t)$ and the joint-space velocity $\dot{\boldsymbol{q}} = \dot{\boldsymbol{q}}(t)$.

We obtain the *motion tangent* $\dot{\mathbb{H}}_i(\boldsymbol{q}, \dot{\boldsymbol{q}})$ from the temporal derivative of the POE-formula (34)

$$\dot{\mathbb{H}}_i(\boldsymbol{q}, \dot{\boldsymbol{q}}) = \frac{d\mathbb{H}_i(\boldsymbol{q})}{d\boldsymbol{q}}\;\dot{\boldsymbol{q}} = \sum_{k=1}^{i}\dot{q}_k\,\frac{\partial\mathbb{H}_i(\boldsymbol{q})}{\partial q_k} \tag{36}$$

$$\frac{\partial\mathbb{H}_i}{\partial q_k} = \mathbf{H}_1(q_1)\cdots\hat{\mathbf{H}}_k\exp(q_k\hat{\mathbf{H}}_k)\cdots\mathbf{H}_i(q_i) \tag{37}$$

The *projective body-velocity* $\hat{\mathbb{H}}_i(\boldsymbol{q})$ of the i^{th} element

$$\hat{\mathbb{H}}_i(\boldsymbol{q},\dot{\boldsymbol{q}}) = \dot{\mathbb{H}}_i(\boldsymbol{q},\dot{\boldsymbol{q}})\mathbb{H}_i^{-1}(\boldsymbol{q}) \tag{38}$$

$$= \sum_{k=1}^{i} \dot{q}_k \underbrace{\left(\mathbb{H}_{k-1}(\boldsymbol{q})\, \hat{\mathbf{H}}_k\, \mathbb{H}_{k-1}^{-1}(\boldsymbol{q})\right)}_{\hat{\mathbf{H}}_{k,\boldsymbol{q}}}. \tag{39}$$

follows from definition (13). The expression $\hat{\mathbf{H}}_{k,\boldsymbol{q}}$ in (39) follows by inserting (34) (37) into (38). It represents the projective motion of the k^{th} joint for its new position in configuration $\boldsymbol{q}$. Consequently, the body velocity as written in (39) can be seen as the joint-wise linear superposition of projective motions $\hat{\mathbf{H}}_{k,\boldsymbol{q}}$ weighted by the joint-velocities q_k.

Finally, points $\boldsymbol{M}_i(\boldsymbol{q})$ rigidly moving with the i^{th} element have the *projective point-velocity*

$$\dot{\boldsymbol{M}}_i(\boldsymbol{q},\dot{\boldsymbol{q}}) = \hat{\mathbb{H}}_i(\boldsymbol{q},\dot{\boldsymbol{q}})\boldsymbol{M}_i(\boldsymbol{q}) = \dot{\mathbb{H}}_i(\boldsymbol{q},\dot{\boldsymbol{q}})\boldsymbol{M}_i(\boldsymbol{0}), \tag{40}$$

$$= \sum_{k=1}^{i} \dot{q}_k\, \hat{\mathbf{H}}_{k,\boldsymbol{q}}\boldsymbol{M}_i(\boldsymbol{q}) = \sum_{k=1}^{i} \dot{q}_k \frac{\partial \mathbb{H}_i(\boldsymbol{q})}{\partial q_k}\boldsymbol{M}_i(\boldsymbol{0}),$$

$$= \sum_{k=1}^{i} \dot{q}_k\, \frac{\partial \boldsymbol{M}_i(\boldsymbol{q})}{\partial q_k}. \tag{41}$$

Equation (40) agrees with direct differentiation of (34). Again, (41) rewrites is as a joint-wise linear superposition of respective velocity components.

6 Non-metric Visual Servoing

In this section, we introduce a *non-metric formulation* of the visual servoing paradigm, based on the projective representation of articulated motions. Classically, this paradigm consists in servoing the end-effector to a target position by means of aligning its velocity screw with the difference between the current and the target image of its features. To fix ideas, consider a gripper, mounted on a six-axis robot arm, moving under visual control of a stereo rig.

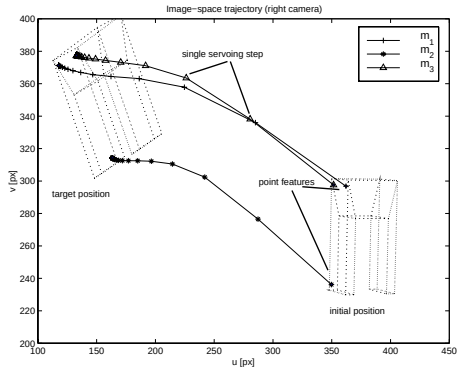

Figure 2: Initial and target image of the gripper and the feature trajectories over the iterations of the control loop.

We call our formulation non-metric for three reasons: First, generally speaking, the geometry of the entire system is modeled w.r.t. the projective camera frame $\mathcal{P}$. Metric frames do not appear anymore. Second, no a-priori knowledge about the geometry of the system is required. The camera geometry, $\mathbf{P}$,$\mathbf{P}'$, the geometry of gripper features $\boldsymbol{M}_6$, and the robot geometry in terms of $\hat{\mathbf{H}}_i$ are acquired on-line. Third, the actual control law no longer servos the robot's Cartesian velocity, but servos the manipulator's joint-velocities. Therefore, we seek to derive a Jacobian that relates image-velocities $\dot{\boldsymbol{s}}$ to joint ones: $\dot{\boldsymbol{s}} = \mathbf{J}(\boldsymbol{q},\boldsymbol{M}_6)\dot{\boldsymbol{q}}$.

Most important is to understand the *Jacobian as an analytic expression* in $\boldsymbol{q},\boldsymbol{M}_6$, which ensures its soundness over the robot's entire configuration space. In contrast to existing systems, it is neither an on-line estimated linear model [12], nor an a-priori given approximation around the target [4], [7].. First of all, for each joint k, its projective motion is developed around the current configuration $\boldsymbol{q}$, i.e. for its current position in space. The calculation is that of $\hat{\mathbf{H}}_{k,\boldsymbol{q}}$ in (39). Second, for each point $\boldsymbol{M}_6$ on the gripper and for each joint k, a point-velocity component $\hat{\mathbf{H}}_{k,\boldsymbol{q}} \cdot \boldsymbol{M}_6$ is developed around $\boldsymbol{q}$ and the current position $\boldsymbol{M}_6$ of the point. Now, their superposition (41) can be expressed by a Jacobian matrix $\mathbf{J}_H$

$$\dot{\boldsymbol{M}}_6(\boldsymbol{q},\dot{\boldsymbol{q}}) = \mathbf{J}_H(\boldsymbol{q},\boldsymbol{M}_6)\dot{\boldsymbol{q}} \tag{42}$$

$$= \left[\frac{\partial \boldsymbol{M}_6(\boldsymbol{q})}{\partial q_1} \;\ldots\; \frac{\partial \boldsymbol{M}_6(\boldsymbol{q})}{\partial q_6}\right]\dot{\boldsymbol{q}}. \tag{43}$$

$$\frac{\partial \boldsymbol{M}_6(\boldsymbol{q})}{\partial q_k} = \hat{\mathbf{H}}_{k,\boldsymbol{q}} \cdot \boldsymbol{M}_6 \tag{44}$$

Finally, the camera Jacobian $\mathbf{J}_C$ between spatial velocity $\dot{\boldsymbol{M}}_6$ and image velocity $\dot{\boldsymbol{s}}$ is developed around the point's current image $\boldsymbol{m} = \mathbf{P}\boldsymbol{M}_6$ in the projective plane $\mathbb{P}2$. This first step has the trivial Jacobian $\mathbf{P}$. The map $C : \boldsymbol{m} \rightarrow \boldsymbol{s} = [m_1/m_3, m_2/m_3]^T$ from $\mathbb{P}^2$ onto the pixel plane de-

scribes the actual perspective projection, which has Jacobian $\mathbf{J}_C(\boldsymbol{m})$

$$\dot{\boldsymbol{s}} = \mathbf{J}_C\Big(\boldsymbol{m}\Big)\,\dot{\boldsymbol{m}},\;\; \mathbf{J}_C\Big(\boldsymbol{m}\Big) = \begin{bmatrix} \frac{1}{m_3} & 0 & -\frac{m_1}{m_3^2} \\ 0 & \frac{1}{m_3} & -\frac{m_2}{m_3^2} \end{bmatrix}. \qquad (45)$$

Finally, the Jacobian $\mathbf{J}$ is a function of $\boldsymbol{q}$ and $\boldsymbol{M}_6$

$$\begin{bmatrix} \dot{\boldsymbol{s}} \\ \dot{\boldsymbol{s}}' \end{bmatrix} = \begin{bmatrix} \mathbf{J}_C(\mathbf{P}\,\boldsymbol{M}_6) & \mathbf{P} & J_H(\boldsymbol{q}, \boldsymbol{M}_6) \\ \mathbf{J}_C(\mathbf{P}'\boldsymbol{M}_6) & \mathbf{P}' & J_H(\boldsymbol{q}, \boldsymbol{M}_6) \end{bmatrix} \dot{\boldsymbol{q}}, \qquad (46)$$

where the projective geometry of the system is encapsulated in $\mathbf{P}, \mathbf{P}'$ and in the $\hat{\mathbf{H}}_i$.

We now report on *simulations* of such a non-metric visual servoing system. The metric geometry of the simulated setup is roughly a stereo system having $20cm$ baseline, 20^o vergence angle, a $3/4''$ CCD, and $12.5mm$ lenses, capturing over time the motion of a PUMA-like robot from $1m$ distance. The latter has an arm with links of $36cm$, $48cm$, and $40cm$ length, and carries a gripper of $10cm$ in size. In contrast, the only inputs used by the non-metric system are joint angle measurements and image projections of gripper features. To take into account real imaging conditions, independent Gaussian noise with $\sigma = 1px$ is added, whereas joint angle are supposed to be accurate.

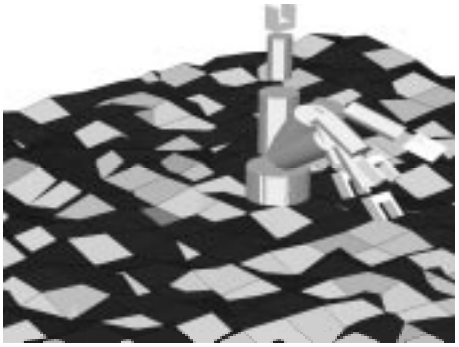

Figure 3: Robot in zero-reference and trial motions.

During the *acquisition phase* (Fig. 3) , six joint-wise trial-motions $[30, 20, 30, 40, 60, 60]^T$ are acquired through the p-rotation homographies $\mathbf{H}_i$ estimated for 7 point features $\boldsymbol{M}_6$ of the gripper. In a first step, the generators $\hat{\mathbf{H}}_i$ are calculated algebraically (27), but have to be refined using a non-linear numerical method to obtain stable and accurate results in presence of noise. The method employed has already proved its performance on real image data [15], [16].

During the *servoing phase*, three points on the visible face of the gripper are tracked, and the point-wise stack of the error vectors from their current images $\boldsymbol{s}, \boldsymbol{s}'$ to their goal images $\boldsymbol{s}_*, \boldsymbol{s}'_*$ constitutes the overall image-error. This stack is used to invert the Jacobian relation (46) for joint-velocities $\dot{\boldsymbol{q}}$. This proportional control law causes the image-error to decrease exponentially [4].

$$\begin{bmatrix} \dot{q}_1 \\ \vdots \\ \dot{q}_6 \end{bmatrix} = -\mathbf{J}^+ \begin{bmatrix} \boldsymbol{s} - \boldsymbol{s}_* \\ \boldsymbol{s}' - \boldsymbol{s}'_* \\ \vdots \end{bmatrix} \qquad (47)$$

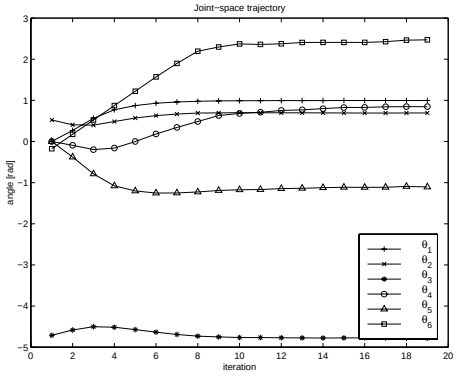

Figure 4: Joint-space trajectory

We show the result for a servoing task which translates the gripper by $82cm$ and rotates it by 88^o. The goal was attained after 20 iterations. In joint-space (Fig. 4), a smooth, almost monotonous trajectory covers a distance of $\boldsymbol{q} = [57, 10, 5, 51, -60, 158]^T$ and attains the target to within $\Delta\boldsymbol{q} = [0.04, 0.15, 0.23, 0.31, 0.10, 1.5]^T$ for $0.5px$ image noise, and to within $\Delta\boldsymbol{q} = [0.01, 0.04, 0.09, 0.19, 0.21, 0.35]^T$ for $0.1px$ image noise. In the image (Fig. 2), the approximately linear trajectory covers a distance of about $220px$. The image error (Fig. 5) shows exponential decay in all its components until convergence is attained with the residual error below the noise levels.

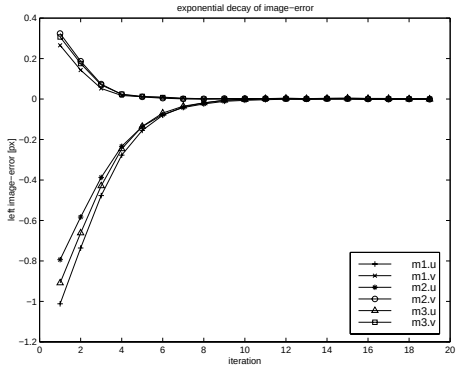

Figure 5: Exponential convergence in image-error.

7 Summary and Conclusions

We have shown how an uncalibrated stereo rig sees rigid and articulated motion.The introduced original formalism of *projective kinematics* has proven to be almost as powerful as the classical kinematics of a metric space. In detail, projective formulations for displacements, for body- and point velocities, as well as for revolute and prismatic joints have been introduced. Most importantly, a projective model for the geometry of an articulated chain has been presented that lead immediately to an original approach to "non-metric visual servoing", which is formulated without any knowledge about the metric geometry of the system, at all.

We hope this theoretical work will give foundations and motivations for the integration of uncalibrated visual sensors into perception-action cycles. We judge the present simulations and formerly published practical experiments very promising. Future work will hence concentrate on further developing the practical and numerical means to better validate the contribution of non-metric systems in practice.

References

[1] P. A. Beardsley, I. D. Reid, A. Zisserman, and D. W. Murray. Active visual navigation using non-metric structure. In *Proc. 5th ICCV*, pages 58–64. IEEE Computer Society Press, June 1995.

[2] G. Csurka, D. Demirdjian, A. Ruf, and R. Horaud. Closed-form solutions for the euclidean calibration of a stereo rig. In *Proc. 5. ECCV*, volume I, pages 426–442. Springer, 1998.

[3] F. Devernay and O. Faugeras. From projective to Euclidean reconstruction. In *Proc. IEEE Conf. CVPR*, pages 264–269, San Francisco, CA., June 1996.

[4] B. Espiau, F. Chaumette, and P. Rives. A new approach to visual servoing in robotics. *IEEE Transactions on Robotics and Automation*, 8(3):313–326, June 1992.

[5] O. D. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig. In G. Sandini, editor, *Computer Vision – Proc. 2. ECCV*, pages 563–578. Springer Verlag, May 1992.

[6] O. D. Faugeras. *Three Dimensional Computer Vision: A Geometric Viewpoint*. MIT Press, Boston, 1993.

[7] G. D. Hager. A modular system for robust positioning using feedback from stereo vision. *IEEE Transactions on Robotics and Automation*, 13(4):582–595, August 1997.

[8] R. I. Hartley. Euclidean reconstruction from uncalibrated views. In M. Z. Forsyth, editor, *Applications of Invariance in Computer Vision*, pages 237–256. Springer Verlag, Berlin Heidelberg, 1994.

[9] R. I. Hartley. In defence of the 8-point algorithm. In *Proc. 5. ICCV*, pages 1064–1070, Cambridge, Mass., June 1995. IEEE Computer Society Press.

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

[11] R. Horaud and G. Csurka. Self-calibration and euclidean reconstruction using motions of a stereo rig. In *Proc. 6. ICCV*, pages 96–103. IEEE Computer Society Press, January 1998.

[12] M. Jagersand and R. Nelson. Aquiring visual-motor models for precision manipulation with robot hands. In Buxton-Cipolla, editor, *Proc. 4. ECCV*, volume II, pages 603–612. Springer, April 1996.

[13] Q.-T. Luong and T. Viéville. Canonic representations for the geometries of multiple projective views. *Computer Vision and Image Understanding*, 64(2):193–229, September 1996.

[14] R. Murray, Z. Li, and S. Sastry. *A Mathematical Introduction to Robotic Manipulation.* CRC Press, 1994.

[15] A. Ruf, G. Csurka, and R. Horaud. Projective translations and affine stereo calibration. In *Proc. IEEE Conf. CVPR*, pages 475–481. IEEE Computer Society Press, June 1998.

[16] A. Ruf and R. Horaud. Projective rotations applied to a non-metric pan-tilt head. In *Proc. IEEE Conf. CVPR, accepted*, June 1999.

[17] J. Selig. *Geometrical Methods in Robotics*. Springer, 1996.

[18] A. Zisserman, P. A. Beardsley, and I. D. Reid. Metric calibration of a stereo rig. In *Proc. IEEE Workshop on Representation of Visual Scenes*, pages 93–100, June 1995.