

Visual Trajectories from Uncalibrated Stereo

Andreas Ruf, Radu Horaud

► **To cite this version:**

Andreas Ruf, Radu Horaud. Visual Trajectories from Uncalibrated Stereo. Workshop on "New Trends in Image-based Robot Servoing" (IROS '97), IEEE (Institute of Electrical and Electronics Engineers) / RSJ (Robotics Society of Japan), Sep 1997, Grenoble, France. pp.83–91. inria-00590079

HAL Id: inria-00590079

<https://hal.inria.fr/inria-00590079>

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 Trajectories from Uncalibrated Stereo

Andreas Ruf Radu Horaud

ruf@inrialpes.fr horaud@inrialpes.fr

GRAVIR-IMAG, INRIA Rh^{one}-Alpes, 38330 Montbonnot St. Martin, FRANCE March 31,
1999

Abstract

This paper proposes a methodological framework for trajectory generation in projective space. The framework enables trajectories that respect rigidity constraints of structure and motion, that are furthermore uniform and optimal in terms of Euclidean configuration, even though no camera calibration is assumed.

In a projective reconstruction of the scene, obtained e.g. from an uncalibrated stereo-rig, the homography between corresponding feature points on a rigid object in different positions is employed to represent its spatial configuration. From its algebraic structure -it is similar to a rigid motion -decoupling of rotation and translation, and uniform interpolation of trajectories are derived.

Speaking in robotic terms, the merely point-to-point definitions of goals, as present in current approaches to visual servoing, are extended to a guarded motion along via points on complete-path trajectories which result from interpolation of a demanded type of motion. Translationonly, rotation-first, rotation-last task-functions are further explicated.

Introduction and motivation

Computer vision in robotics started as off-line system for part-or landmark-recognition, trajectory planning or mobile robot navigation, and computer-aided quality control. From the continuing increase in the capabilities of these systems and especially from the decrease of run-time towards real-time, the integration of computer vision into the on-line loop has become feasible. In consequence, the domain of vision systems has extended to object detection, object tracking and pose estimation for the purposes of obstacle avoidance, target tracking, execution monitoring, and finally for position-based control. The respective control laws run an open loop which consists of successive but well separated stages, one for determining the position of the manipulator and one for correcting this position by actuator motions towards the control objective. These approaches are classified as "look-and-move" [6].

1.1 Closed-loop image-based approach

Further reduction of cycle time nowadays allows a full integration of dynamic visual sensor input into closed-loop feedback control [3]. In contrast to open-loop control, a closed-loop is less sensitive to inaccuracies in calibration, is robust to disturbances during execution, requires however careful design of its dynamics in order to ensure convergence and stability [12]. Indeed, such dynamic visual servoing systems have recently been developed. They are called "image-based", since the target is given merely in terms of images, e.g. by a set of image points, since the error-function is a distance between the current and the desired configuration of these image points, and since the control-law reduces this error by servoing a robot in closedloop mode without explicitly determining pose or spatial structure [5]. For these control-laws, local convergence and stability in presence of calibration errors has been demonstrated [2]. The principal deficiencies are that these approaches enable only point-to-point goals, that trajectories are not further confined, particularly that they are not optimal, that vanishing features due to occlusion or loss of focus often result in divergence, and that even in case of complete visibility convergence is guaranteed only locally. All these are consequences of the fact that rigidity of structure and motion is no longer ensured when

calculations are solely image-based. More figuratively, the straight-forward descend on the gradient of image error in P-controllers does not “look-ahead”. So, either local minima might trap the controller before reaching the final goal or sudden changes in error due to occlusion or loss of focus cause divergence. Furthermore, trajectories induced by image-based gradient descend are neither physically valid nor optimal, i.e. they tend to deform structure by violating rigidity of motion and additionally fail to constrain motion to the shortest path towards the goal and direction of approach.

1.2 Generic task-functions

In this paper, we propose a method for complete-path goal definition and subsequent generation of optimal trajectories for image-based visual servoing. Although just weakly calibrated stereo-rigs are assumed, trajectories respect rigidity and guarantee shortest path-to-goal. At first, trajectories are calculated in projective space from a spatial reconstruction of the scene. They are later reprojected to the images to serve as set-points for image-based control.

The investigated classes of complete-path trajectories

include translation-only, parallel-translation, rotation-first, rotation-last trajectories. A Cartesian-move mode in projective space should also be feasible. These occur quite often in practice. For instance, the list of orders given by an operator in video-based telemanipulation constitutes generic robotic tasks that correspond to certain trajectory classes:

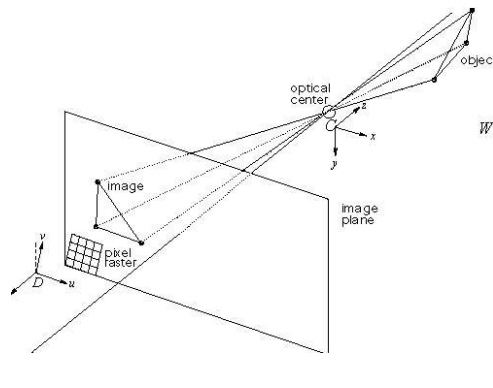


Figure 1: Pin-hole camera model.

2.2 Intrinsic geometry

A canonical affine frame D is attached to the video-camera (CCD), such that the u - v -origin is the pixel origin, the u -axis is parallel to pixel columns, the v -axis is again parallel to image scan-lines, and the w -axis is parallel to the optical axis with unit length f in negative z -direction. Units on the u - and v -axis are in pixel of width $1/k_u$ and of height $1/k_v$. The transition from camera-frame C to video-frame D is by the affine transform A (Fig. 1).

$$\begin{bmatrix} y \\ U \\ V \\ W \end{bmatrix} = \begin{bmatrix} k_u & k_{uv} & -u_0/f & 0 \\ 0 & k_v & -v_0/f & 0 \\ 0 & 0 & -1/f & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ Y \\ Z \\ 1 \end{bmatrix}$$

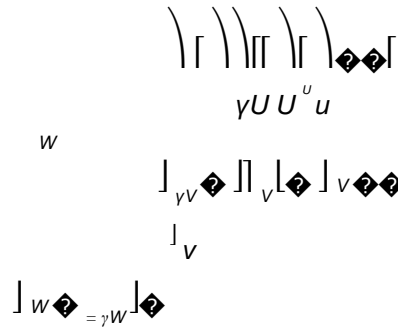
$$\gamma N = ATRT X,$$

(1)

The deviation of the u -axis from orthogonality to v introduces an affine skew in the focal plane by k_{uv} .

2.3 Image projection

In D , normalization with γ yields affine space coordinates $[U, V, W]^T$ of N . Normalizing with γW results in image coordinates $[u, v]^T$ in pixel.



$$| \diamond = \gamma | \diamond = \gamma W$$

- . • grasp object, e.g. fix, load part
- . • lift object, e.g. car disassembly
- . • put-down object on table, e.g. surface mount
- . • insert peg into hole, e.g. plug in socket
- . • turn object, e.g. turn valve
- . • drag object, e.g. move heavy load
- . • parallel alignment of tool to target, e.g. driver to screw
- . • perpendicular alignment of tool and planar target, e.g. drill normal to surface
- . • angular alignment of tool and target, e.g. spray-can normal to surface

1.3 Outline

The paper is divided into six sections. After the introduction in section 1, some algebraic and geometric fundamentals are laid down in section 2. In section 3 the representation of pure translations in a projective frame is investigated. An algebraic algorithm for translation-based affine auto-calibration follows in section 4, including some consequences on the generation of point-to-point translations and R-T decomposition, both in a projective frame with known affine properties. The relevance of the preceding results for visual servoing is demonstrated in section 5, which develops the continuous, uniform, completepath task-function for a straight-line translation, observed by weakly calibrated cameras. Section 6 concludes with a discussion of the presented framework.

$$\begin{aligned} & \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} W \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \\ & \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} W \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \\ & \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} W \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \end{aligned}$$

Fundamentals

We will revisit this purely scalar relation between image coordinates and affine space coordinates in D in section

5. The usual formulation of perspective projection evokes now as a special case of the representation in the affine frame D , mainly because the u - v -plane is parallel to the

(2)

$$X = PE^T X,$$

(3)

2.1 Extrinsic geometry

The geometry of the scene is given in a world-frame of reference W and points X in W are in homogeneous coordinates ${}^W X = [X, Y, Z, 1]^T$. A Euclidean camera-frame C is canonically attached to each camera, such that the origin of C coincides with the optical center, the Z -axis coincides with the optical axis, the X - Y plane becomes the focal plane, parallel to the image plane, and the Y -axis is parallel to image scan-lines (Fig. 1). The displacement between world-and camera-frame is written as homogeneous transform ${}^{TR} T$, which maps W -coordinates onto C -coordinates: ${}^C X = {}^{TR} T {}^W X$.

with $\diamond_{uv} = -fk_u$, $\diamond_{v} = -fk_v$: (U_0, V_0) is the pixel where plane equation $\diamond_{uv} N = 0$, so $\diamond_{uv} = [0, 0, 0, 1]$ in the optical axis hits the image plane, and \diamond_{uv} describes the affine coordinates. Projective coordinates of \diamond_{uv} follow affine skew between columns and rows of the CCD: the five from applying the infinity constraint to rectified coordinintrinsic parameters of the camera. rates $N = SE^T M$

2.4 Projective reconstruction

$$\begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}^T \begin{matrix} SE^T \\ SE^T \\ SE^T \\ SE^T \end{matrix} \begin{bmatrix} M \\ M \\ M \\ M \end{bmatrix} = 0, \quad (7)$$

It is well known, that from at least two uncalibrated images and eight common points, a reconstruction of the scene in which reveals the forth

focal plane (Fig. 1).

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} u \\ v \\ 1 \end{bmatrix}$$

row $S_{E(4)}$ of SE to hold the co-a-freely chosen frame R of projective three-space $P(3)$ can ordnates of \diamond in R . \diamond rests fix under projective displacement calculated [4], [11]. They are however a projectivity SE placements, i.e. it is an eigenvector of any H^{-T}

away from a Euclidean reconstruction in $WYN = ASESE^{-1}X$,

choice of the projective frame in which the scene was reconstructed. Hence, its coordinates depend mainly on the

$$YN = HA M, \quad \text{2.7 Canonic representation} \quad (4)$$

where M is the representation of X in R and S_E is a The so far arbitrarily fixed Euclidean world-frame can be "Euclideanizing" homography $X = S_E M$, constrained

2.5 Projective displacements

projective displacement, the similarity class writes as Consider now a set of k points X on a rigid body and in a canonical way, as it is done by the "Jordan canonical form" (JCF) of similar matrices. In case of a

let this body undergo a rigid displacement TRT , such that

$$X_i = TRT X_i$$

$$\begin{matrix} \cos \diamond - \sin \diamond & 00 \\ \sin \diamond & \cos \diamond & 00 \\ 0 & 0 & 1 \end{matrix} \quad \begin{matrix} | \\ | \\ | \end{matrix} \quad \begin{matrix} | \\ | \\ | \end{matrix} \quad (8)$$

$$= YS_{RT} \begin{bmatrix} \diamond S_{RT} \\ 0 & 0 & 1 \end{bmatrix}$$

$$SE^{-1} X_i = SE^{-1} TRT SE SE^{-1} X_i \quad (5)$$

$$M_i = Y HRT M_i$$

$$HRT = YSRT^{-1} JRT SRT$$

shows that the Euclidean representation T_{RT} and the pro-where the canonical frame is such that rotation is by an angle \diamond around the Z-axis and translation is of unit length in Z-ment are related by a scaled algebraic similarity with SE direction. The X and Y-axis are chosen mutually orthogo

as similarity transform (see also [13], [1]) nal in a plane perpendicular to the Z-axis. This corresponds to representing HRT in the frame of normalized screw- H motion, i.e. the origin is shifted to lie on the screw-axis,

$$= YS^{-1} T^{-1} S \cdot (6)$$

which then coincides with the Z-axis, and scales are such

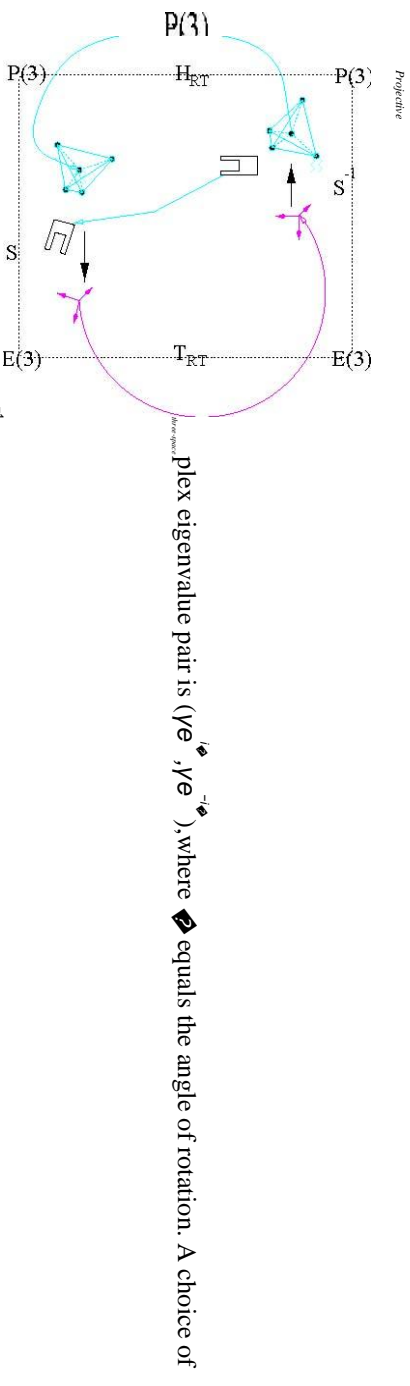
Call "projective displacements" the subgroup of projective

that the screw-pitch is of unit-length (Fig. 3). The algebraic

transforms, which are similar to a rigid displacement (Fig.

2).

characterization of H_{RT} is obvious from its JCF . In the general case, the only real eigenvalue is γ , with algebraic multiplicity two, but geometric multiplicity one. The com-



eigenvectors are the columns of S_{RT}^{-1} . The first two span an orthonormal basis of the X - Y -plane, the third is parallel to the Z -axis. Together they form an orthogonal triad spanning Euclidean vector space.

S

3 Pure translations

By the JCF some special similarity classes of H_{RT} are identified. A *projective translation* HT is a projective

transform, which has a JCF JT defined like

$$JT = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (9)$$

1000 Figure 2: Euclideanizing homography SE and

the resulting sim-
 $= \gamma S_T^{-1} \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

ilarity between a rigid RT and a projective displacement H_{RT} .

2.6 Plane at infinity

$$H_T = \begin{bmatrix} \gamma & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \gamma S_T^{-1} J_T S_T,$$

A point is said to be "at infinity" iff its affine coordinates

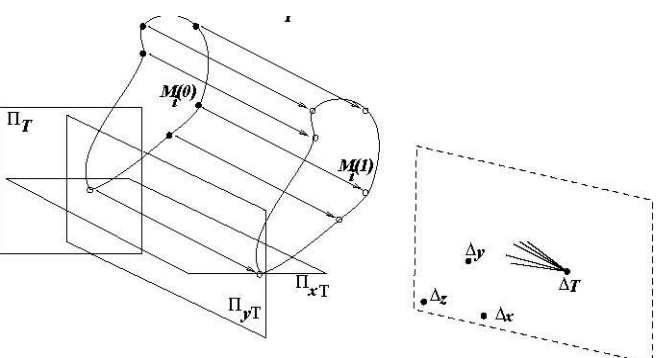
T

The Jordan form is canonical up to permutations of the Jordan blocks.

fulfill $N = (U, V, W, 0)$. The set of all points at infinity

Without loss of generality, we concentrate on the permutation as given by form a "plane at infinity" $\diamond \diamond$, which is defined by the above.

displacement H_T



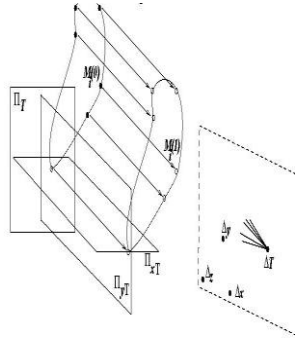


Figure 4: Projective translation H_T , mapping points $M_i(C)$ onto $M_i(1)$. The eigenspace of the point-translation H_T spans the plane at infinity Π_∞ , e.g. with the vanishing points of coordinate axis $\Delta_x, \Delta_y, \Delta_z$ as basis. The eigenspace of the

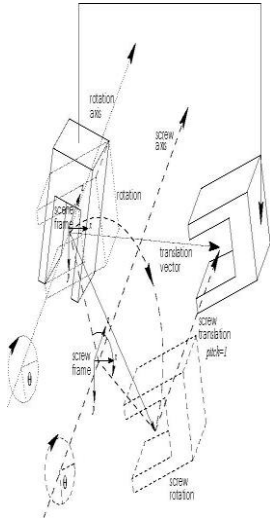


Figure 3: Screw-geometry of a rigid displacement of a gripper (solid lines). The displacement in the original frame is drawn in dotted lines, the respective screw displacement in dashed lines.

8

where γ is an arbitrary scale factor. ST rectifies projective coordinates, such that translation is of unit length in z -direction.

Algebraically JT is characterized by $\det(JT) = 1$, $\text{trace}(JT) = 4$, the quadruple eigenvalue $\lambda = 1$, and the $\text{dim} = 3$ eigenspace $E^1 = [x, y, z, 0]^T$, which spans pointwise the plane at infinity $\lambda = [0, 0, 0, 1]$, which is at the same time the eigenspace's orthogonal complement

$$\lambda^\perp = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}^T$$

$E^1 = 0$. From duality, the plane transformation corresponding to HT follows as

$$\begin{aligned} & \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}^T \\ & \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}^T \\ & = 1 \gamma ST^{-r} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}^T, (10) \\ & \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}^T \end{aligned}$$

Now, the dual eigenspace $E^{-T} = [X, Y, 0, 1]$ is the sub

1

space of planes parallel to the translation, including the plane at infinity, and the orthogonal complement $N = [0, 0, 1, 0]^T$ is the direction of translation.

The translation direction is in the point eigenspace E^T as well as the plane at infinity is in the plane eigenspace E^{-T} .

1

Although they cannot be distinguished from any other vector of the eigenspace, their property of being orthogonal complement of the respective dual eigenspace allows their calculation from a given HT . We will propose an efficient algorithm for doing so in section 4.

The similarity of HT to JT by ST implies directly $\det(HT) = Y$, $\text{trace}(HT) = 4Y$, a quadruple eigen-value $\lambda = Y$, and a transformed eigenspace $E^T = S^{-1}$

$[X, Y, Z, 0]^T$. Analogous arguments apply to the dual transform. In the following, we introduce a reduced parameterization for a projective translation. Based on this, the geo-

Figure 4: Projective translation HT , mapping points $M(0)$ onto $M(1)$. The eigenspace of the point-translation HT spans the plane at infinity $\lambda = 1$, e.g. with the vanishing points of coordinate axis $\lambda = X$, $\lambda = Y$, $\lambda = Z$ as basis. The eigenspace of the plane-translation H^T is spanned by the plane at infinity and

Two planes parallel to the translation, e.g. $\lambda = 1$, $\lambda = XT$, $\lambda = YT$ and the vanishing point of the translation $\lambda = T$ is the intersection of these three planes.

metric interpretations of the parameter vectors follow easily and a factorization-based algorithm for affine autocalibration of a stereo-rig is derived. Then, a way to calculate the projective translation between a given point pair and the decomposition of a general projective displacement into rotation and translation is shown.

After that, the task-function of translational motion is explicated. The often found point-to-point definition of a positioning task is extended to a rigid translation of the complete object along a continuous-path, which yields a globally admissible task-function in the image domain. The uniform, shape-preserving interpolation of image trajectories and the incorporation of a-priori velocity profiles are finally developed.

3.1 Reduced parameterization

In the new form, the number of parameters of HT reduces to eight, four of them are fix for all translations, the remaining four characterize the present translation, plus an arbitrary scale factor.

After writing the columns of ST^{-1} as c_i , the rows of ST as r_i , and separating the identity matrix I , equation (10) becomes:

$$HT = Y \begin{pmatrix} c_1 & c_2 & c_3 & c_4 \\ r_1 & r_2 & r_3 & r_4 \end{pmatrix} \begin{pmatrix} I & & & \\ & I & & \\ & & I & \\ & & & I \end{pmatrix} \begin{pmatrix} ST^{-1} \\ ST^{-1} \\ ST^{-1} \\ ST^{-1} \end{pmatrix} \quad (11)$$

0001 0000

$\begin{pmatrix} r_1 \\ r_2 \\ r_3 \\ r_4 \end{pmatrix}$

$$= Y \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + \begin{bmatrix} c_1 & c_2 & c_3 & c_4 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} E_{34} \\ \vdots \\ \vdots \\ \vdots \end{bmatrix} \quad (12)$$

$$= Y (I + c_3 r^4) \quad (13)$$

The vectors c_3 and r^4 however are not fully independent: $\text{trace}(H^T) = 4\gamma$ follows from similarity, and consequently $\text{trace}(H_0) = 0$. This can likewise be seen from

$$r^4 c_3 = 0, \quad (14)$$

since $S^T S^{-1} = 0$. Hence, H_0 is a “zero”-matrix in the sense that it has rank 1 and trace 0.

3.2 Plane at infinity

As shown in section 2.6, r^4 is the plane at infinity \diamond in the projective frame R . Multiplying (12) from the left r^4 $H^T = Y r^4$ shows that r^4 is a left eigenvector of H^T , i.e. r^4 is eigenvector of H^T to eigenvalue $1/\gamma$. In geometric terms, r^4 is a fixed plane of the plane-transform H^T .

What is more, r^4 is a plane of fixed points, as is easily seen by multiplying (12) from the right by a point M in this plane, i.e. by a point which satisfies $r^4 M = 0$.

3.3 Vanishing point

An endlessly continued translation of an affine point $N(t)$ along direction vector $\diamond = [u, v, w]^T$ results at the limit in the “vanishing point” N_\diamond , which corresponds to the direction of translation

$$N_\diamond := \lim_{t \rightarrow \infty} N(t) = \lim_{t \rightarrow \infty} \begin{bmatrix} u + tv \\ v \\ w \end{bmatrix} = \begin{bmatrix} \infty \\ v \\ w \end{bmatrix} \quad (15)$$

$${}^{\infty}P \diamond \diamond \lceil W + tW$$

10

It obviously lies in the plane at infinity $Pi \diamond \diamond M \diamond = 0$. Its image $PN \diamond$ is called “focus of expansion” of the given translation.

In the affine frame of (10), translation is along the z-axis and the vanishing point becomes $M \diamond = [0, 0, 1, 0]$. Going back to the projective frame R by S^{-1} shows that the column $G3$ appearing in (12) holds the projective coordinates $M \diamond$ of the vanishing point, which characterizes projective translations along this direction, independent of their amplitude. More formally, $G3$ is an eigenvector of HT , which is evident by multiplying (12) from the right.

To sum up, a projective translation HT depends merely on a scalar factor r and two vectors r^4 and $G3$, which appear as a row of the rectifying homography and a column of its inverse, respectively. They obey a single constraint $r^4 \cdot G3 = 0$. In geometric terms, $G3$ is the direction of translation, r^4 is the plane at infinity, and the constraint places $G3$ at infinity, i.e. it is the limit of all points translating in this direction. The vanishing point $G3$ is eigenvector of the point-translation HT , the plane at infinity r^4 is eigenvector of the plane-translation H^{-1} . The eigenspace

T of HT is the plane at infinity and HT leaves all vanishing points unchanged. Dually, the eigenspace of H^{-1} is plane-wise fix and consists of all planes parallel to the translation, including the plane at infinity.

5

4 Affine calibration

4.1 The algorithm

The relationship in equation (11) gives raise to a subspace method for affine auto-calibration of perspective cameras from one or more translation observed with a weakly calibrated stereo system. In theory, affine calibration corresponds to determining the plane at infinity, which appears in $H0$ calculated like

$$H0 = HT N - I, \quad (16)$$

where $Y = trace(HT)/4$. In each row the plane at infinity r^4 appears scaled by the components of $G3$, and likewise the direction $G3$ appears in each column scaled by the components of r^4 .

In practice, an estimation $H0 \diamond$ will be disturbed from fuzzy data and round-off errors. Straight-forward accumulation by componentwise sums (17) and componentwise products (18) is potentially unstable.

$$\begin{aligned} & \prod_{i=1}^4 H0(i, 1) \quad \diamond \\ & \parallel H0(i, 2) \\ & \parallel H0(i, 3) \\ & r^4 = \prod_{i=1}^4 (17) \end{aligned}$$

$$\begin{aligned} & \lceil \\ & \lceil_{i=1}^4 H0(i, 3) \\ & H0(i, 4) \end{aligned}$$

$$H_0(i, 1) \quad [1/4]$$

$$P_4 = \begin{bmatrix} \diamond \\ (18) \end{bmatrix}$$

$$\begin{bmatrix} H_0(i, 3) \\ \vdots \\ H_0(i, 4) \end{bmatrix} \quad [1]$$

$$\diamond \begin{bmatrix} H_0(i, 2) \end{bmatrix}$$

As \mathcal{C}_3 might have components close to zero, numerical effacements in (17) or instabilities due to sign changes in (18) are likely to occur. This effect has likewise been observed when calculating $\det(H^T)$ instead of $\text{trace}(H^T)$ to determine γ . Even in presence of minor disturbances, the determinant became unstable.

Unlike the ill-conditioned methods above, we propose to exploit $\text{rank}(H_0) = 1$ in the singular value decomposition (SVD) of H_0

$$H_0 \begin{matrix} \diamond \\ \end{matrix} = U^T \text{diag}(\alpha_1, \alpha_2, \alpha_3, \alpha_4) V, \quad (19)$$

with $U = (u_1, u_2, u_3, u_4)$, $V = (v_1, v_2, v_3, v_4)$, and the singular values in descending order. The rank constraint permits to neglect α_2, α_3 , and α_4 , and H_0 becomes

$$H_0 = \alpha_1 u_1 v_1^T, \quad (20)$$

while at the same time minimal error $\|H^* - H_0\|_0$ in the

sense of the matrix norms $\|\cdot\|_2$ and $\|\cdot\|_F$ is guaranteed. Comparison with (11) reveals

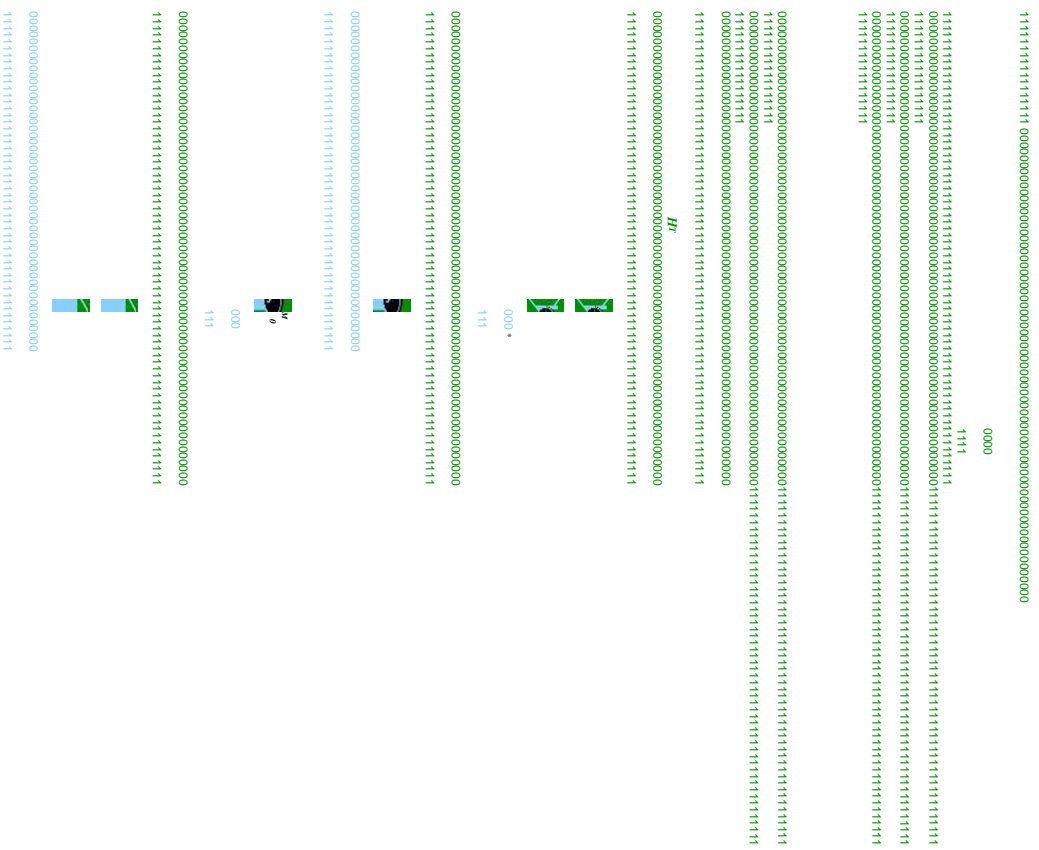
$$c_3 = u_1, \quad (21)$$

$$P_4 = v_1. \quad (22)$$

Singular value decomposition is equally applicable for “stacked” $H_0^{(4) \times n}$, which combines different H^T estimates from various translational motions expressed in the same projective base. In practice, the calibration procedure is as follows:

1. Take n image pairs during arbitrary translations of the camera system and the scene.
2. Calculate a projective reconstruction of the translating points in one and the same projective base for consecutive image pairs.

$$M = YHT M_0 \quad (23)$$



Translation imposes parallelism on all point traces. This is

will come in. Solving for c_3 with (12) yields

$$c_3 = (24) \quad Y M^T - M_0$$

$$r_4^T M_0$$

The choice of γ has to guarantee (14), so

$$Y = \begin{matrix} r_4^T & M_0 \\ r_4^T & M_1 \end{matrix}$$

$$(25) \quad M^T$$

and Hr as a function of M_0, M_1 and r_4 becomes

$$r_4^T M_0 \quad M_1^T \quad M_0^T$$

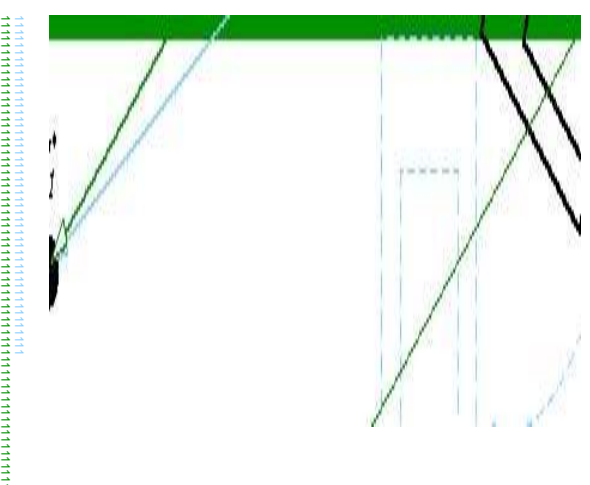


Figure 6: RT-Decomposition depending on the chosen center of rotation.

$$HT = I + \dots (26)$$

Now, considering the calculated projective translation as a $M_i = HT M_0$. The corresponding projective translarobotic task, the whole object is moved rigidly in the scene tion HT is determined following section 4.2. The resid-

-1

and contact, collision, or occlusion could be checked (Fig. ual displacement $H = HT HT$ is clearly the pro5). jective representation of a pure rotation H_R . To obtain

-1

a rotation-first task, decomposition is $H = HT HT$ **4.3 R-T Decomposition** and $H = HT HT^{-1}$ for a rotation-last task. The condi-We will show how to decompose a projective displacement $M_i = H M_i$ obviously holds, so the axis of rotation ment with respect to a central point, which moreover con-passes through M_i . strains the rotation axis to pass through it. One could think of generating rotation-first, rotation-last trajectories and a

In summary, affine calibration by determining the plane

Cartesian-move mode. at infinity permits calculating a projective transform which rigidly moves a given point to a destination by a translata as a sequence of one projective rotation H_R and one projectional displacement. It furthermore enables to decompose a general displacement into a translation HT

An arbitrary projective displacement HT is expressed

and rotation H_R , whose axis is constrained to pass through the distinct- $H_{Rt} = H_R H_T$ (27) gushed point. This is of particular importance in robotics,

be an ideal point robots' kinematics, e.g. the robot-wrist's center point, al-or an observed point – the destination M_i is calculated by lows to decouple the control-law.

5 The task of translation

We are now considering a robotic task that consists of a uniform translational displacement of an object. Given the reconstruction of a point of reference $Q(0)$, and in its target position $Q(1)$, the corresponding projective translation HT is determined by (26) and fully defines the task. The direction becomes $M_i = C3$, according to (24). Distance and orientation are implied by the target point $Q(1)$. If translation is to be defined by its direction only, taking an arbitrary via-point as $Q(1)$ suffices and extrapolation yields the motion beyond this point. Noteworthy, \diamond is an unoriented direction only, while orientation is defined by the order of $Q(0)$ and $Q(1)$. In addition, the reference point is not required to lie on the object itself, i.e. any pair of points defines a translational task.

5.1 Singleton set-point definition

Consider M_i on the object and their initial positions $M_k(0)$ in the reconstruction. Compose the configuration matrix $S(\cdot)$

$$S(0) = M_i(0) M_k(0) \dots M_k(0) \quad (28)$$

which represents pose and shape of the object. Apply HT to obtain the set-point configuration $S(1) = HT S(0)$. Since HT preserves rigidity, $S(1)$ is, up to occlusion, an admissible configuration, i.e. corresponds to the image of the rigid object in the respective pose. The image configuration sand the image

set-points \mathbf{s}^* follow by simple backprojection

$$\mathbf{s}(0) = PS(0), \mathbf{s}^*(1) = PS^*(1), \quad (29)$$

where PS is the projection matrix (3), that is available for each observing camera from the reconstruction step.

So far, this corresponds to the usual task definition by “point-point-incidence”, present in known approaches to visual servoing [3]. Neither trajectory, nor direction of approaching are defined. Unlike existing approaches, the definition of the task by the projective translation HT allows for the generation of multi-point set-points for an arbitrary rigid object. At the same time, the translation task HT can be calculated from any observed translation vector.

5.2 Complete-path task-function

The definition of the task given by the projective translation HT now is extended to a continuous-path, uniform, straight-line translation. This completely constrains the trajectory and additionally the direction of approach. Having

(10) in mind, the definition [10] of our continuous taskfunction $HT(\theta)$ becomes

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$HT(\theta) = Y S T^{-1} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} S^T, t \in [0, 1]. \quad (30)$$

$$\begin{bmatrix} 1 & 0 & 0 & t \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

⁷ Please note that uniformity is with respect to spacial Euclidean distance, i.e. $HT(\theta)$ corresponds to a t -fractile of the task's total distance. The formulation in (30) equally applies for moving backwards $t < 0$, towards the set-point $0 < t < 1$, and beyond the set-point $t > 1$. In contrast to simple interpolation of image set-points, rigidity is guaranteed at any t .

5.3 Uniform set-point interpolation

Thanks to the parameterization in (13), the calculation of $HT(\theta)$ becomes a weighted linear interpolation between initial and final configuration $HT(0)$ and $HT(1)$ with the scale ratio Y

$$HT(\theta) = Y(1 - \theta)HT(0) + \theta HT(1), \quad (31)$$

Unless Y is respected, $HT(\theta)$ is neither uniform, nor “rigid”. The same holds for the interpolation of \mathbf{S}^* and \mathbf{S} .

$$\mathbf{S}^*(\theta) = Y(1 - \theta)\mathbf{S}^*(0) + \theta \mathbf{S}^*(1), \quad (32)$$

$$\mathbf{S}(\theta) = Y(1 - \theta)\mathbf{S}(0) + \theta \mathbf{S}(1), \quad (33)$$

The singleton set-points \mathbf{S}^* and \mathbf{s}^* become a set-point function $\mathbf{S}^*(t)$ and $\mathbf{s}^*(t)$, the latter in homogeneous image coordinates. The trace of $\mathbf{s}^*(t)$ coincides with the image lines connecting initial and target image points. The trajectory of $\mathbf{s}^*(t)$, a dynamically varying set of kimage points, is however the projection of a uniform, rigid translation. This implies that at any t the current set-point $\mathbf{S}^*(t)$ corresponds to an admissible pose of the object, which lies on the shortest path to the target. Moreover, the parameter t corresponds to relative Euclidean distance along the path. Purely image-based interpolation, in contrast, deforms the object and fails to reflect spatial distance.

5.4 Velocity profiles

The dynamic formulation now allows to impose any velocity profile on the trajectory. Direct application of $\mathbf{S}^*(t) \rightarrow [0,1]$ yields constant velocity translation parallel to the Z -axis. Composing $\mathbf{S}^*(t)$ with the integral function of a desired “velocity profile” $V(U)$ generates a set-point function of variable gain. Figure 7, for instance, depicts an initial acceleration until saturation, followed by a phase of constant velocity, and a final deceleration until complete set-point overlap $\mathbf{S}^*(1)$. More formally, let $V(U) : U \rightarrow [0,1]$ be the desired velocity profile. The resulting time-dilating function X is defined as

$$X(t) = \int_0^t V(U) dU, \text{ with } X(1) = 1. \quad (34)$$

In fact, this procedure corresponds to enrolling the velocity profile by integration, resulting in a one-dimensional trajectory over time, which is finally projected onto the direction of translation.

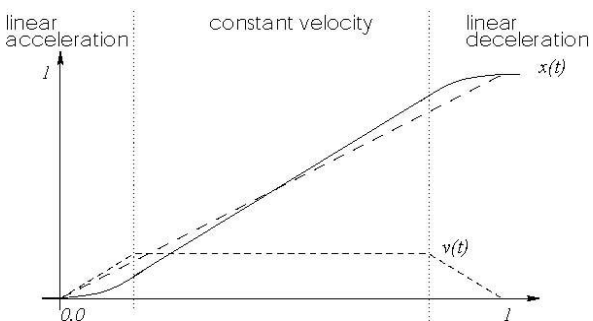


Figure 7: Three-phase velocity profile $V(U)$ and the time-dilation $X(t)$.

6 Discussion and Conclusions

In this framework we propose generating task-functions that correspond to complete-path trajectories solely and immediately from the visual information of the scene, without knowledge of the calibration parameters. In classic approaches to task planning, completely known Euclidean structure of the work-space is assumed. Therefore, planning and generation of trajectories are moreless equivalent. In particular, no additional difficulty arises from trajectory generation and from their transfer to the sensor-space.

Unlike these approaches, the presented framework does without Euclidean scene structure, and does without a calibrated sensor. Hence, care has to be taken that the constraints resulting from observing a Euclidean world by a perspective visual sensor are respected by the generated task-functions, or more figuratively, that a link between sensor-space and work-space is kept, that reflects as far as possible the implicit relations between both, as their explicit relationship is not accessible in an uncalibrated environment.

6.1 A generated task

We distinguish generating a task-function from planning a task in the sense that no combinatoric search takes place, that no decisions are made. Generating a visual taskfunction corresponds rather to transforming image information, such that the obtained task-function is a visual instantiation of a generic task, with the current visual configuration of the scene as input parameters (see Fig. 8). The specification of the above mentioned generic tasks is however thought to be in terms of rigid motions of the object, articulated motions of the actuator, or Euclidean geometry of the scene. Thus, at any time the relationship between the visual configuration of features and the geometric and kinematic configuration of tool and actuator have to be identified to the largest extend possible in an uncalibrated setup.

Apart from explicit task planning, the decomposition of observed displacements and composition from observed displacements is included as “construction phase” in the framework of visual trajectory generation. They sometimes require further knowledge about the scene structure. In this 8 report, we have given some consequences and examples of determining the affine properties of the scene. However, the explicit step of affine reconstruction is avoided.

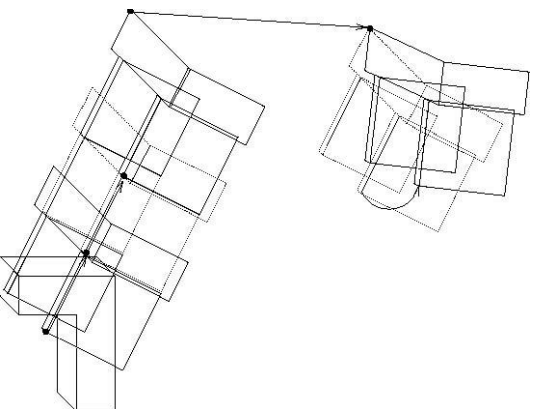


Figure 8: The instantiation of a generic grasping task. Starting from the object to be grasped, the gripper position can be transferred as shown in [5]. Subsequently, a pure translation of twice the claw's edge is instantiated, which avoids an object collision during approaching. Finally, the present position of the gripper is taken and a rotation-first trajectory to reach the initial position of approach is generated.

6.2 A plane at infinity

The affine properties of the scene have been proven to be of particular use in the construction and decomposition of projective displacements, the “uncalibrated” representation of rigid spatial motion. The identification of the plane at infinity introduces the concepts of parallelism and oriented directions. They allow to generate the projective translation between an ordered pair of points, and consequently to decompose a general projective displacement into a translation and a rotation around a given center-point. An efficient algebraic method of motion-based affine auto-calibration was given, which nicely fits in the considered system-setup,

i.e. controlled robot motions observed by a weakly calibrated stereo-rig.

A further affine property are length ratios. They allow not just for the rigid interpolation of projective motion. Above that, uniformity of the corresponding trajectories in the sense of Euclidean configuration is achieved.

6.3 A complete path

It was demonstrated how to inter- and extrapolate projective motion to generate the complete-path of a rigid motion. Here, the path consists primarily of the traces of rigidly moving points. The interpolation of the path is of particular interest since potential local minima in usually point-to-point-wise defined functions of the image error are bridged. An additional contribution to globally convergent control

²The ISO defines “path” as “...*spatial locus drawn by the movement of any point on the robot or the workpiece along which orientation of the end-effector may or may not be variable.*” is that the path corresponds to motions with monotonically decreasing error in configuration. In conclusion, we consider a complete-path as a step towards globally convergent control.

6.4 A uniform trajectory

The difference between a trajectory and a path is that the trajectory describes the progression of motion over time and the object's configuration at a time instant, whereas the path describes the swept trace of motion, only. Thus, the proposed interpolation of projective motion ensures that the reprojected, now dynamically varying set-points correspond at any time to an admissible configuration of the object, i.e. to a rigidly moving object in varying poses. Thus rigidity is preserved, trajectories are physically valid.

Over and above that, the interpolation parameter t establishes an implicit relation between visual and Euclidean configuration, and the corresponding error functions. In other words, the visual set-points at t correspond to the fractile of distance in Euclidean configuration. One can also view this relationship as an implicit calibration of and only of the currently instantiated visual task with respect to (relativ) error in configuration.

We consider the established relationship between visual and Euclidean configuration as a step towards weaker requirements on hand-eye and kinematic actuator calibration [7], [8]. Furthermore, there is now an apparent connection between the gain in visual configuration and the gain in object configuration, which allows for optimizing with respect to dynamic properties, e.g. maximal acceleration, speed limits of the actuator [9].

References

- [1] F. Devernay and O. Faugeras. From projective to euclidean reconstruction. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, San Francisco, California, USA, June 1996.*

- [2] B. Espiau. Sur les erreurs en asservissement visuel. Rapport de recherche, INRIA, July 1995.
- [3] 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, 1992.
- [4] O. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig? In G. Sandini, editor, *Proceedings of the 2nd European Conference on Computer Vision, Santa Margherita Ligure, Italy*, pages 563–578. Springer-Verlag, May 1992.
- [5] R. Horaud, F. Dornaika, and B. Espiau. Visually guided object grasping. *IEEE Transactions on Robotics and Automation*, 1997. accepted.
- [6] S. Hutchinson, G. D. Hager, and P. I. Corke. A tutorial on visual servo control. *IEEE Transactions on Robotics and Automation*, 12(5):651–669, October 1996.
- [7] M. Jaegersand. On-line estimation of visual-motor models using active vision. In *Proceedings of the ARPA Image Understanding Workshop*, 1996.
- [8] M. Jaegersand. Visual servoing using trust region methods and estimation of the full coupled visualmotor jacobian. In *Proceedings of IASTED Applications of Control and Robotics*, pages 105–108, 1996.
- [9] E. Marchand, A. Rizzo, and F. Chaumette. Avoiding robot joint limits and kinematic singularities in visual servoing. In *Proceedings of 13th IARP/IEEE International Conference in Pattern Recognition*, volume A, pages 297 – 301, 1996.
- [10] C. Samson, B. Espiau, and M. Le Borgne. *Robot Control: The Task Function Approach*. Clarendon Press, Oxford, 1991.
- [11] B. Triggs. Factorization methods for projective structure and motion. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, San Francisco, California, USA*, pages 845–851, 1996.
- [12] L.E. Weiss, A.C. Anderson, and C.P. Neuman. Dynamic sensor-based control of robots with visual feedback. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 3(5):404–417, October 1987.
- [13] A. Zisserman, P.A. Beardslley, and I.D. Reid. Metric calibration of a stereo rig. In *Workshop on Representation of Visual Scenes, Cambridge, Massachusetts, USA*, pages 93–100, June 1995.

