Stereo Calibration from Rigid Motions

Radu Horaud, Gabriella Csurka, and David Demirdjian

INRIA Rhône-Alpes & GRAVIR-CNRS

655, avenue de l'Europe

38330 Montbonnot Saint-Martin, FRANCE

Corresponding author: Radu Horaud

e-mail: Radu.Horaud@inrialpes.fr

phone: +33 476 61 52 26

fax: +33 476 61 54 54

Submitted to IEEE PAMI

Short paper version, 10 August 2000

Abstract

In this paper we describe method for calibrating a stereo pair of cameras using general or planar motions. The method consists of upgrading a 3-D projective representation to affine and to Euclidean without any knowledge neither about the motion parameters nor about the 3-D layout. We investigate the algebraic properties relating projective representation to the plane at infinity and to the intrinsic camera parameters when the camera pair is considered as a moving rigid body. We show that all the computations can be carried out using standard linear resolutions techniques. An error analysis reveals the relative importance of the various steps of the calibration process: projective-to-affine and affine-to-metric upgrades. Extensive experiments performed with calibrated and natural data confirm the error analysis as well as the sensitivity study performed with simulated data.

Keywords

Self-calibration, projective reconstruction, metric reconstruction, rigid motion, stereo vision, affine calibration, epipolar geometry.

I. Introduction, background, and contribution

This paper addresses the problem of metric calibration of a camera-pair. The overall approach consists of allowing the camera-pair to move while it observes a 3-D layout. Neither the motion parameters of the camera-pair nor the 3-D structure of the layout is known in advance. The fundamental hypothesis made and used throughout the paper is that the geometry of the camera-pair remains unchanged (the intrinsic parameters of both cameras as well as the position and orientation between the two cameras are fixed) and hence the camera-pair may well be viewed as a rigid body. Without loss of generality we will refer to such a fixed-geometry camera-pair as a *stereo rig*. Stereo-rig motions differ drastically from single-camera motions since in the latter case there is no rigidity associated with view pairs. There are several advantages of using a stereo rig rather than a single camera:

1. The epipolar geometry associated with the left and right images of the stereo rig remains unchanged over time and therefore it is possible to collect several image pairs and to compute a common fundamental matrix for all these image pairs rather than for individual pairs;

2. The tracking of image point-features over time is facilitated by the fact that the two cameras are rigidly attached to each other, and
3. The problem of upgrading 3-D projective structure to affine structure can be solved either in closed form or using linear minimization methods which is a great advantage over non linear methods generally associated with single camera calibration.

The latter of the above items is developed below and constitutes the main contribution of the paper.

Interesting enough, although the problem of camera calibration from rigid motions has been thoroughly investigated, there has been considerably less work dedicated to stereo rig calibration from rigid motions.

The Kruppa equations [7] relate the intrinsic camera parameters to the epipolar geometry associated with two views. However, solving these polynomial equations requires non-linear resolution methods. The absolute quadric can be used instead but it leads to non-linear resolution methods as well [12]. An alternative solution to these direct methods is to upgrade the projective structure to affine structure (affine calibration) and next to convert the latter to metric structure [8]. Affine calibration amounts to recover the infinite homography between two views. Once this homography has been estimated for at least two camera motions (or three views) it is possible to recover the internal camera parameters [2]. Unfortunately, affine calibration is a delicate process which requires either special camera motions such as rotations around the center of projection or specific knowledge about the observed scene features (parallel lines, midpoints of segments, etc.).

The major advantage of using a stereo rig rather than a single camera is that affine calibration becomes more tractable since linear algebraic solutions exist. Moreover, affine calibration is an *intrinsic* property of the rig – the infinite homography between the left and right views remains unchanged. Various techniques were suggested in the past to compute this left-view-to-right-view infinite homography. Moons et al. [9] describe a method based on vanishing point detection through pure translational motions of the stereo rig. One basic observation introduced by Beardsley et al. [1] and by Zisserman et al. [13] is that the projective and rigid motions of a stereo rig are conjugated. These authors investigated two types of motions: (1) planar motion and (2) general motion. In the first case the stereo rig is allowed to move in a plane perpendicular to a unique axis of rotation and the plane at infinity is defined by a line at infinity and a point at infinity. These line and point are the same, regardless of the number of motions. In the second case, the plane at infinity can be recovered as the unique eigenvector associated with the double eigenvalue (equal to 1) of a 3-D projective transformation. These authors have therefore a major contribution since they showed for the first time that affine calibration of a stereo rig amounts to a straightforward algebraic property.

The solution suggested in [13] computes both the epipolar geometry of the stereo rig **and** the epipolar geometry of the left camera motion. It will be shown that the latter computation is useless. Indeed, let us consider the left camera. The intrinsic parameters of the left camera are related to the infinite homography between views associated with the left camera motion **and not** to the left-view-to-right-view infinite homography. Below (proposition 2) we reveal the link (under the form of a simple analytic relation) between projective calibration of the rig and affine calibration associated with the motion of the left camera (or, equivalently, the motion of the right camera).

More insights into the affine and metric calibrations of a stereo rig are to be found in a paper by Devernay & Faugeras [3]. However, the intrinsic camera parameters do not appear explicitly in their formulation and hence it seems difficult to consider constraints associated with these parameters. Ruf & Horaud used a similar approach to parameterize projective translations, projective rotations, and projective articulated motion observed with a stereo camera pair [11].

More precisely, let $\mathcal{P}_1$ and $\mathcal{P}_2$ be two projective reconstructions of the same set of 3-D points obtained with an uncalibrated stereo rig before and after a rigid motion. Each one of these two reconstructions has a projective basis associated with it and hence they are related by a 4×4 homography $\mathbf{H}_{12}$ which is related to the 4×4 rigid motion $\mathbf{D}_{12}$ by ([13], [3]):

$$\mathbf{H}_{12} = \alpha \mathbf{H}_{PE}^{-1} \mathbf{D}_{12} \mathbf{H}_{PE} \tag{1}$$

where α is an arbitrary scale factor and $\mathbf{H}_{PE}$ is a 4×4 homography allowing the projective reconstruction to be upgraded to an Euclidean one. If a 3-D point M has projective coordinates $\boldsymbol{M}_1 \in \mathcal{P}_1$ and $\boldsymbol{M}_2 \in \mathcal{P}_2$ then $\boldsymbol{M}_2 \simeq \mathbf{H}_{12}\boldsymbol{M}_1$. The Euclidean coordinates of the same point are $\boldsymbol{N}_1 \simeq \mathbf{H}_{PE}\boldsymbol{M}_1$ and $\boldsymbol{N}_2 \simeq \mathbf{H}_{PE}\boldsymbol{M}_2$ with $\boldsymbol{N}_2 \simeq \mathbf{D}_{12}\boldsymbol{N}_1$. $\mathbf{H}_{12}$ *will be referred to as the projective motion of the rig from position 1 to position 2.*

This paper has the following contributions. First we show that with appropriate choices both for the projective basis and Cartesian frame associated with the stereo rig, matrix $\mathbf{H}_{PE}$ is parameterized by the plane at infinity and by the intrinsic parameters of the left camera. The result is known for single-camera in motion and it is of crucial importance for a moving stereo-rig. In particular we show that for a sequence of general motions or planar motions the plane at infinity of the stereo rig can be linearly estimated.

Second we show that $\mathbf{H}_{12}$ in eq. (1) leads to a simple analytic expression relating the projective motion of a stereo rig to the infinite homography associated with the motion of the left camera (or the right camera, depending on the origin of the Euclidean frame). This means that it is not necessary to determine the epipolar geometry associated with the left (or right) camera motion. Instead only the epipolar geometry associated with the stereo rig is required. As already mentioned, when a stereo rig moves, its "internal" epipolar geometry remains unchanged. Therefore one can incrementally improve the quality of the epipolar geometry as the stereo rig moves and compensate for flat scenes which are known to be an important source of numerical instabilities whenever the epipolar geometry is estimated.

Finally, the parameterization just mentioned allows for an error analysis which characterizes the relative importance of projective-to-affine upgrade with respect to affine-to-metric upgrade. In the light of this analysis we show that calibration errors generated by projective-to-affine calibration are likely to dominate over calibration errors generated by affine-to-metric calibration.

II. Notations, camera models, and reconstruction

Throughout the paper matrices are typeset in boldface ($\mathbf{H}$, $\mathbf{P}$, $\mathbf{I}$), vectors in slanted boldface ($\boldsymbol{m}$, $\boldsymbol{M}$), and scalars in italic. $\overline{\mathbf{P}}$ denotes a block of $\mathbf{P}$. 4-vectors are generally denoted by an upper case letter and 3-vectors are denoted either by a lower-case letter or by an overlined upper-case letter, $\overline{\boldsymbol{M}}$. $\mathbf{H}^\top$ is the transpose of $\mathbf{H}$ and $\mathbf{H}^{-\top}$ is the transpose of the inverse of $\mathbf{H}$.

A pinhole camera projects a point $\boldsymbol{M}$ from the 3-D projective space onto a point $\boldsymbol{m}$ of the 2-D projective plane. This projection can be written as a 3×4 homogeneous matrix $\mathbf{P}$ of rank equal to 3: $\boldsymbol{m} \simeq \mathbf{P}\boldsymbol{M}$ If we restrict the 3-D projective space to the Euclidean space, then it is well known that $\mathbf{P}$ can be written as (the origin and orientation of the Euclidean frame is arbitrarily chosen): $\mathbf{P}_E \simeq \mathbf{K}\begin{pmatrix} \mathbf{R} & \boldsymbol{t} \end{pmatrix}$. If we choose the standard camera frame then $\mathbf{R} = \mathbf{I}$ (rotation) and $\boldsymbol{t} = \mathbf{0}$ (translation). The projection matrix has the form given by eq. (5) below.

The most general form for the matrix of intrinsic parameters $\mathbf{K}$ is an upper triangular matrix defined by 5 parameters:

$$\mathbf{K} = \begin{pmatrix} \alpha & r\alpha & u_0 \\ 0 & k\alpha & v_0 \\ 0 & 0 & 1 \end{pmatrix} \tag{2}$$

where α is the horizontal scale factor, k is the ratio between the vertical and horizontal scale factors, $r\alpha$ is the image skew and u_0 and v_0 are the image coordinates of the center of projection. It will be useful to consider camera models with a reduced set of intrinsic parameters, as follows:

- *four-parameter camera* where either $r = 0$ or $k = 1$
- *three-parameter camera* with $r = 0$ and $k = 1$

A stereo rig is composed of two cameras fixed together. Let $\mathbf{P}$ and $\mathbf{P}'$ be the 3×4 projection matrices of the left and right cameras. One may distinguish between the projective model and the Euclidean or calibrated model:

- **In the uncalibrated case** and without loss of generality the two projection matrices can be written as:

$$\mathbf{P} \simeq \begin{pmatrix} \mathbf{I} & \mathbf{0} \end{pmatrix} \tag{3}$$
$$\mathbf{P}' \simeq \begin{pmatrix} \overline{\mathbf{P}}' & \boldsymbol{p}' \end{pmatrix} \tag{4}$$

- **In the calibrated (Euclidean) case** one can use the following projection matrices ($\mathbf{K}'$ is the matrix of right camera intrinsic parameters and $\mathbf{R}'$ and $\boldsymbol{t}'$ describe the orientation and position of the right camera frame with respect to the left camera frame):

$$\mathbf{P}_E \simeq \begin{pmatrix} \mathbf{K} & \mathbf{0} \end{pmatrix} \tag{5}$$

$$\mathbf{P}'_E \simeq \begin{pmatrix} \mathbf{K}'\mathbf{R}' & \mathbf{K}'\boldsymbol{t}' \end{pmatrix} \tag{6}$$

With these expressions for $\mathbf{P}$ and $\mathbf{P}'$ we obtain the following parameterizations for the infinity-plane homography and for the epipoles:

$$\mathbf{H}_\infty \simeq \mathbf{K}'\mathbf{R}'\mathbf{K}^{-1} \tag{7}$$

$$\boldsymbol{e} \simeq -\mathbf{K}\mathbf{R}'^{\top}\boldsymbol{t}' \text{ and } \boldsymbol{e}' \simeq \mathbf{K}'\boldsymbol{t}' \tag{8}$$

Given a stereo rig with two projection matrices as defined in eqs. (3) and (4) it is possible to compute the 3-D projective coordinates $\boldsymbol{M}$ of a point M from the equations $\mu\boldsymbol{m} = \mathbf{P}\boldsymbol{M}$ and $\mu'\boldsymbol{m}' = \mathbf{P}'\boldsymbol{M}$. Hartley and Sturm proposed an optimal solution both from numerical and statistical points of view [5].

Matrices $\mathbf{P}$ and $\mathbf{P}'$ can be estimated from point matches without any camera calibration: Indeed, given at least 8 left-right image point correspondences, one can estimate the fundamental matrix which encapsulates the epipolar geometry for a pair of uncalibrated views [4]. Several authors proved that the two projection matrices can be obtained from the epipolar geometry up to a 4-parameter projective mapping [8]:

$$\mathbf{P} \simeq \begin{pmatrix} \mathbf{I} & \mathbf{0} \end{pmatrix} \tag{9}$$

$$\mathbf{P}' \simeq \begin{pmatrix} \mathbf{H}_\infty + \boldsymbol{e}'\boldsymbol{a}^\top & a\,\boldsymbol{e}' \end{pmatrix} \tag{10}$$

where $\mathbf{H}_\infty$ and $\boldsymbol{e}'$ were defined above, $\boldsymbol{a}$ is an arbitrary 3-vector and a is an arbitrary scale factor. It will be shown below that the 4-vector $\boldsymbol{A}^\top = (\boldsymbol{a}^\top \; a)$ has a simple but important geometric interpretation.

III. Rigid versus projective motion

We consider the stereo rig, the projection matrices $\mathbf{P}$ and $\mathbf{P}'$ associated with it and two positions of the rig – position 1 and position 2 – before and after a motion. A fundamental observation is that the projection matrices can be choosen such that they are the same for all the rig positions. Without loss of generality, one may choose the matrices given by equations (9) and (10). The projective coordinates of a 3-D point are related to its Euclidean coordinates by $\boldsymbol{N}_i \simeq \mathbf{H}_{PE}\boldsymbol{M}_i$ where $\boldsymbol{M}_i$ and $\boldsymbol{N}_i$ are respectively, the projective and Euclidean homogeneous coordinates of the same point M when the stereo pair is in position i. Since the stereo rig is a rigid body, the homography $\mathbf{H}_{PE}$ *remains the same* for any position i and hence the projective displacement $\mathbf{H}_{12}$ is related to the rigid displacement $\mathbf{D}_{12}$ by the homogeneous similarity transformation defined in equation (1).

The fact that traces and determinants of matrices are invariant under similarity transformations will allow us to compute the scale factor α and define a *normalized projective displacement*. Indeed, we have: trace $(\mathbf{H}_{12})$ = trace $(\alpha\mathbf{D}_{12}) = 2\alpha(1+\cos\theta)$ and $\det(\mathbf{H}_{12}) = \det(\alpha\mathbf{D}_{12}) = \alpha^4$ where θ is the angle of rotation associated with the displacement $\mathbf{D}_{12}$. Since we have $1+\cos\theta \geq 0$ we obtain: $\mathrm{sign}\,(\alpha) = \mathrm{sign}\,(\mathrm{trace}\,(\mathbf{H}_{12}))$. Therefore, the scale factor α can uniquely be determined for each homography $\mathbf{H}_{12}$:

$$\alpha = \mathrm{sign}\,(\mathrm{trace}\,(\mathbf{H}_{12}))\sqrt[4]{\det(\mathbf{H}_{12})} \tag{11}$$

From now on we can replace the homogeneous homography by a normalized one:

$$\mathbf{H}_{12} \simeq \frac{1}{\alpha}\mathbf{H}_{12} = \mathbf{H}_{PE}^{-1}\mathbf{D}_{12}\mathbf{H}_{PE} \tag{12}$$

For simplifying notation, $\mathbf{H}_{12}$ will simply denote the normalized homography $\frac{1}{\alpha}\mathbf{H}_{12}$.

We defined a similarity relationship between a rigid displacement and a projective displacement. We already used the fact that traces and determinants are invariant under such a similarity transformation in order to determine the scale factor associated with the projective displacement. The eigenvalues are also invariant under matrix similarity and we are going to use some properties associated with them in order to characterize rigid motion based on the algebraic properties of the projective displacement $\mathbf{H}_{12}$.

A rigid displacement $\mathbf{D}$ is composed of a rotation matrix and a translation vector. The eigenvalues associated with a rigid displacement are $\lambda \in \{e^{i\theta}, e^{-i\theta}, 1, 1\}$ where θ is the angle of rotation. The algebraic multiplicity of $\lambda = 1$ is equal to 2 unless the motion is a pure translation, i.e., $\theta = 0$ in which case its algebraic multiplicity is equal to 4. The case of a pure translation is studied elsewhere [10]. In what follows we consider motions which have a non-null rotational component.

The geometric multiplicity of an eigenvalue is equal to the dimension of its associated eigenspace [6]. The geometric multiplicity of $\lambda = 1$ depends of the motion type:

- it is equal to 1 for a general motion and
- it is equal to 2 for a planar motion (the translation vector is perpendicular to the axis of rotation).

To summarize, since the geometric multiplicity of an eigenvalue remains invariant under similarity, the dimension of the eigenspace associated with the double eigenvalue $\lambda = 1$ of the projective transformation informs us about the type of motion undergone by the stereo pair.

In practice, let $\boldsymbol{Y}_i$ be the eigenvector of $\mathbf{D}_{12}$ associated with the eigenvalue λ_i: $\mathbf{D}_{12}\boldsymbol{Y}_i = \lambda_i \boldsymbol{Y}_i$. From eq. (12) we obtain $\mathbf{H}_{12}\boldsymbol{X}_i = \lambda_i \boldsymbol{X}_i$ with $\boldsymbol{X}_i = \mathbf{H}_{PE}^{-1}\boldsymbol{Y}_i$. If $\boldsymbol{Y}_i$ spans a vector space of dimension k then $\boldsymbol{X}_i$ will span a vector space of dimension k as well. Let $\boldsymbol{X}$ be the vector space associated with the eigenvalue $\lambda = 1$ of $\mathbf{H}_{12}$. $\boldsymbol{X}$ verifies $(\mathbf{H}_{12} - \mathbf{I})\,\boldsymbol{X} = \mathbf{0}$. The dimension of $\boldsymbol{X}$ is either 1 (for general motions) or 2 (for planar motions). The dimensionality of $\boldsymbol{X}$ can be computed from the singular values of $\mathbf{H}_{12} - \mathbf{I}$.

IV. From projective to metric structure

We are interested into the problem of converting 3-D projective structure into metric structure. This conversion is a projective mapping from the projective space onto its Euclidean sub-space and this mapping is the 4×4 homography $\mathbf{H}_{PE}$ which appeared already in eq. (1). The left and right camera projection equations can be written as:

$$\boldsymbol{m} \simeq \mathbf{P}\mathbf{H}_{PE}^{-1}\ \mathbf{H}_{PE}\boldsymbol{M} \tag{13}$$

$$\boldsymbol{m}' \simeq \mathbf{P}'\mathbf{H}_{PE}^{-1}\ \mathbf{H}_{PE}\boldsymbol{M} \tag{14}$$

Since $\boldsymbol{N} \simeq \mathbf{H}_{PE}\boldsymbol{M}$ is an Euclidean representation of $\boldsymbol{M}$, the projection matrices $\mathbf{P}\mathbf{H}_{PE}^{-1}$ and $\mathbf{P}'\mathbf{H}_{PE}^{-1}$ must have the structure given by eqs. (5) and (6). We can now state the following proposition:

Proposition 1: The 4×4 homography allowing the conversion of a projective reconstruction obtained with a stereo rig into an Euclidean reconstruction has the following structure:

$$\mathbf{H}_{PE} \simeq \begin{pmatrix} \mathbf{K}^{-1} & 0 \\ \boldsymbol{a}^\top & a \end{pmatrix} \tag{15}$$

where $\mathbf{K}$ is the matrix of intrinsic parameters of the left camera and $(\boldsymbol{a}^\top \ a)$ is the equation of the plane at infinity in the projective basis chosen to represent the projective reconstruction.

Proof: A projection matrix of the left camera can be written as the following product:

$$\mathbf{P} \simeq (\mathbf{I} \ \ \mathbf{0}) \simeq (\mathbf{K} \ \ \mathbf{0}) \begin{pmatrix} \mathbf{K}^{-1} & \mathbf{0} \\ \boldsymbol{a}^\top & a \end{pmatrix} \simeq \mathbf{P}_E \mathbf{H}_e$$

By substituting eq. (7) and eq. (8) into eq. (10) we obtain:

$$\mathbf{P}' \simeq (\mathbf{K}'\mathbf{R}\mathbf{K}^{-1} + \mathbf{K}'\boldsymbol{t}\boldsymbol{a}^\top \ \ a\mathbf{K}'\boldsymbol{t}) = \mathbf{P}'_E \mathbf{H}_e$$

Eqs. (13) and (14) become $\boldsymbol{m} \simeq \underbrace{\mathbf{P}_E \mathbf{H}_e}_{\simeq \mathbf{P}} \mathbf{H}_{PE}^{-1} \ \ \mathbf{H}_{PE} \boldsymbol{M}$ and $\boldsymbol{m}' \simeq \underbrace{\mathbf{P}'_E \mathbf{H}_e}_{\simeq \mathbf{P}'} \mathbf{H}_{PE}^{-1} \ \ \mathbf{H}_{PE} \boldsymbol{M}$ By simply taking $\mathbf{H}_{PE} \simeq \mathbf{H}_e$ proves the first part of the proposition.

In order to prove the second part of proposition 1 let us consider again the similarity transformation defined by eq. (12). By taking the transpose of the inverse we obtain:

$$\mathbf{H}_{12}^{-\top} = \mathbf{H}_{PE}^\top \mathbf{D}_{12}^{-\top} \mathbf{H}_{PE}^{-\top} \tag{16}$$

By duality, if $\mathbf{H}_{12}$ and $\mathbf{D}_{12}$ are point transformation matrices, $\mathbf{H}_{12}^{-\top}$ and $\mathbf{D}_{12}^{-\top}$ are plane transformation matrices. Matrix $\mathbf{D}_{12}$ represents a rigid motion and hence its eigenvalues are $\lambda \in \{e^{i\theta}, e^{-i\theta}, 1, 1\}$. Therefore the eigenvalues of $\mathbf{D}_{12}^{-\top}$ are $1/\lambda \in \{e^{-i\theta}, e^{i\theta}, 1, 1\}$. $\mathbf{D}_{12}^{-\top}$ has the following structure:

$$\mathbf{D}_{12}^{-\top} = \begin{pmatrix} \mathbf{R}_{12} & \mathbf{0} \\ -\boldsymbol{t}_{12}^\top \mathbf{R}_{12} & 1 \end{pmatrix}$$

The eigenspace associated with the double eigenvalue $\lambda = 1$ is given by the equation:

$$\mathbf{D}_{12}^{-\top} \boldsymbol{Y} = \boldsymbol{Y} \tag{17}$$

which yields:

$$\left\{ \begin{array}{rcl} \mathbf{R}_{12}\boldsymbol{y} & = & \boldsymbol{y} \\ \boldsymbol{t}^\top \mathbf{R}_{12}\boldsymbol{y} & = & 0 \end{array} \right.$$

with $\boldsymbol{Y}^\top = \begin{pmatrix} \boldsymbol{y}^\top & y \end{pmatrix}$. The first of the equations above ($\mathbf{R}_{12}\boldsymbol{y} = \boldsymbol{y}$) admits two solutions: $\boldsymbol{y} = \boldsymbol{n}$ and $\boldsymbol{y} = \mathbf{0}$ where $\boldsymbol{n}$ is the rotation axis. For a general motion the direction of translation is such that $\boldsymbol{t}_{12}^\top \boldsymbol{n} \neq 0$. For a planar motion we necessarily have $\boldsymbol{t}_{12}^\top \boldsymbol{n} = 0$. Therefore we must distinguish between these two motions:

- *general motion*: the only possible solution is $\boldsymbol{y} = \mathbf{0}$ and therefore the eigenspace has dimension 1 with: $\boldsymbol{Y}_1 = (0\ 0\ 0\ 1)^\top$. This is the plane at infinity.
- *planar motion*: the eigenspace has dimension 2 because both $\boldsymbol{y} = \boldsymbol{n}$ and $\boldsymbol{y} = \mathbf{0}$ are admissible solutions of $\mathbf{R}_{12}\boldsymbol{y} = \boldsymbol{y}$ and $\boldsymbol{t}_{12}^\top \boldsymbol{y} = 0$. $\boldsymbol{Y}$ spans a two dimensional vector space and one possible basis of this space is provided by $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2^\top = \begin{pmatrix} \boldsymbol{n}^\top & 0 \end{pmatrix}$. Both $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$ are planes. $\boldsymbol{Y}_1$ is the plane at infinity and $\boldsymbol{Y}_2$ describes a plane orthogonal to the axis of rotation and passing through the origin. In the case of planar motion the eigenspace spanned by $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$ is a pencil of planes, $\boldsymbol{Y} = n\boldsymbol{Y}_1 + \boldsymbol{Y}_2 = \begin{pmatrix} \boldsymbol{n}^\top & n \end{pmatrix}^\top$.

From eq. (16) it follows that the eigenspace of $\mathbf{H}_{12}^{-\top}$ associated with $\lambda = 1$ is spanned by $\mathbf{H}_{PE}^\top \boldsymbol{Y}$:

$$\boldsymbol{X}_1 = \mathbf{H}_{PE}^\top \boldsymbol{Y}_1 = \begin{pmatrix} \mathbf{K}^{-\top} & \boldsymbol{a} \\ \mathbf{0}^\top & a \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} \boldsymbol{a} \\ a \end{pmatrix} \tag{18}$$

for the general-motion case, and by $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ in the planar-motion case with:

$$\boldsymbol{X}_2 = \mathbf{H}_{PE}^\top \boldsymbol{Y}_2 = \begin{pmatrix} \mathbf{K}^{-\top} & \boldsymbol{a} \\ \mathbf{0}^\top & a \end{pmatrix} \begin{pmatrix} \boldsymbol{n} \\ 0 \end{pmatrix} = \begin{pmatrix} \mathbf{K}^{-\top}\boldsymbol{n} \\ 0 \end{pmatrix} \tag{19}$$

Therefore $\boldsymbol{X}_1^\top = (\boldsymbol{a}^\top\ a)$ represents the plane at infinity in the projective basis associated with the stereo rig. This proves the second part of proposition 1. ■

A. Affine calibration

We derive two corollaries which allow for affine calibration of a stereo rig either from general motions (corollary 1.1) or from planar motions (corollary 1.2):

Corollary 1.1: Let us consider several rigid **general motions** being observed with a stereo rig. The normalized matrices $\mathbf{H}_{12}^\top$, $\mathbf{H}_{23}^\top$, $\mathbf{H}_{n\,n+1}^\top$ have the same eigenvector associated with the double eigenvalue 1. This eigenvector is solution of the following set of linear homogeneous equations:

$$\begin{pmatrix} \mathbf{H}_{12}^\top - \mathbf{I} \\ \vdots \\ \mathbf{H}_{n\,n+1}^\top - \mathbf{I} \end{pmatrix} \begin{pmatrix} \boldsymbol{a} \\ a \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} \tag{20}$$

Proof: First, notice that $\mathbf{H}_{12}^\top$ and $\mathbf{H}_{12}^{-\top}$ admit $\lambda = 1$ as eigenvalue and the associated eigenvector is the same for both matrices. Therefore and unlike it is suggested in [13], it is not necessary to invert the homographies in order to extract affine structure. From the proof of proposition 1 it is straightforward to notice that the eigenvector of $\mathbf{H}_{12}^\top$ associated with the unit eigenvalue is not a function of the rigid motion $\mathbf{D}_{12}$ of the stereo rig. Hence this eigenvector can be estimated as the common root of equations $(\mathbf{H}_{ij}^\top - \mathbf{I})\boldsymbol{X} = 0$ for all motions 12, 23, etc. ∎

Corollary 1.2: Let us consider $n \geq 2$ planar motions with **distinct motion planes** of a stereo rig. The normalized matrices $\mathbf{H}_{12}^\top$, $\mathbf{H}_{23}^\top$, $\mathbf{H}_{n\,n+1}^\top$ have a common eigenvector associated with the double eigenvalue 1. This eigenvector is the plane at infinity and is the solution of the set of linear homogeneous equations given by eq. (20).

Proof: Without loss of generality we consider two distinct planar motions and let $\mathbf{H}$ and $\mathbf{H}'$ be the two 4×4 homographies associated with these two motions. Moreover, let $\boldsymbol{X}$ and $\boldsymbol{X}'$ be the eigenspaces of $\mathbf{H}^\top$ and $\mathbf{H}'^\top$ associated with $\lambda = 1$. We have $\left(\mathbf{H}^\top - \mathbf{I}\right)\boldsymbol{X} = \mathbf{0}$ and $\left(\mathbf{H}'^\top - \mathbf{I}\right)\boldsymbol{X}' = \mathbf{0}$. Both $\boldsymbol{X}$ and $\boldsymbol{X}'$ are pencils of planes and each-one of these pencils can be parameterized by two planes which have the structure given by eq. (18) and eq. (19): $\boldsymbol{X} = \alpha_1 \boldsymbol{X}_1 + \alpha_2 \boldsymbol{X}_2$ and $\boldsymbol{X}' = \alpha_1' \boldsymbol{X}_1 + \alpha_2' \boldsymbol{X}_2'$. The condition that $\mathbf{H}^\top$ and $\mathbf{H}'^\top$ have a common eigenvector writes $\alpha_1 \boldsymbol{X}_1 + \alpha_2 \boldsymbol{X}_2 = \alpha_1' \boldsymbol{X}_1 + \alpha_2' \boldsymbol{X}_2'$ which implies that $\boldsymbol{X}_2$ and $\boldsymbol{X}_2'$ belong to the same pencil of planes. This is in contradiction with the fact that the two motions have two distinct axes of rotation. Therefore the common solution must satisfy $\alpha_2 = \alpha_2' = 0$ and we obtain:

$$\boldsymbol{X} = \boldsymbol{X}' \simeq \boldsymbol{X}_1 = \begin{pmatrix} \boldsymbol{a} \\ a \end{pmatrix}$$

■

In practice, the computation of the plane at infinity is a straightforward linear estimation method. It is worthwhile to notice that one can use (i) a single general motion, (ii) two planar motions, or (iii) sum up several general and/or planar motions in order to recover the sought plane.

B. Metric calibration

In order to compute the intrinsic parameters of the left (or right) camera one must determine the infinite homography associated with the camera motion and with two views. The following proposition establishes the link between the intrinsic parameters on one side and the affine calibration obtained from projective motion on the other side.

Proposition 2: Let a projective displacement $\mathbf{H}_{12}$ be written as:

$$\mathbf{H}_{12} = \begin{pmatrix} \overline{\mathbf{H}} & \boldsymbol{h} \\ \boldsymbol{k}^\top & h \end{pmatrix} \tag{21}$$

The relationship between the intrinsic parameters of the left camera, matrix $\mathbf{H}_{12}$ and the affine calibration of the stereo rig is:

$$\mathbf{K}\mathbf{R}_{12}\mathbf{K}^{-1} = \overline{\mathbf{H}} - \frac{1}{a}\boldsymbol{h}\boldsymbol{a}^\top \tag{22}$$

Proof: The structure of $\mathbf{H}_{PE}$ given by eq. (15) allows us to write matrix $\mathbf{H}_{12}$ as a function of $\mathbf{K}$, $(\boldsymbol{a}^\top\ a)$, $\mathbf{R}_{12}$, and $\boldsymbol{t}_{12}$:

$$\begin{aligned} \mathbf{H}_{12} &= \mathbf{H}_{PE}^{-1}\mathbf{D}_{12}\mathbf{H}_{PE} && (23) \\ &= \begin{pmatrix} \mathbf{K}\mathbf{R}_{12}\mathbf{K}^{-1} + \mathbf{K}\boldsymbol{t}_{12}\boldsymbol{a}^\top & a\mathbf{K}\boldsymbol{t}_{12} \\ \frac{1}{a}\left(-\boldsymbol{a}^\top\mathbf{K}\mathbf{R}_{12}\mathbf{K}^{-1} - \boldsymbol{a}^\top\mathbf{K}\boldsymbol{t}_{12}\boldsymbol{a}^\top + \boldsymbol{a}^\top\right) & -\boldsymbol{a}^\top\mathbf{K}\boldsymbol{t}_{12} + 1 \end{pmatrix} && (24) \end{aligned}$$

By identifying eq. (21) with eq. (24) we obtain eq. (22). ■

Notice that:

$$\mathbf{G}_{\infty} = \mathbf{K}\mathbf{R}_{12}\mathbf{K}^{-1} \tag{25}$$

is the 3×3 infinite homography relating two views of the left camera before and after a motion and it is not the left-view-to-right-view infinite homography $\mathbf{H}_{\infty}$ defined above. Therefore, estimating the vector $\boldsymbol{A} = (\boldsymbol{a}^{\top}\ a)^{\top}$ from a number of projective motions $\mathbf{H}_{ij}$ is equivalent to performing an affine calibration of the left (or of the right) camera. The internal camera parameters (matrix $\mathbf{K}$) can be determined from equation (22) either by using the constraint $\mathbf{R}_{12}\mathbf{R}_{12}^{\top} = \mathbf{I}$ or by performing a real-Jordan factorization of $\mathbf{G}_{\infty}$ [2].

V. Error analysis

In this section we establish an expression for the error in Euclidean reconstruction as a function of errors associated with projective reconstruction, affine, and Euclidean upgrades. We show that, independently of the calibration method being used, projective reconstruction and affine calibration are tight together and they have stronger impact onto the final result than than affine to Euclidean upgrade.

We consider again the relationship between the projective and Euclidean homogeneous coordinates of the same 3-D point, $\boldsymbol{N} = \mu\mathbf{H}_{PE}\boldsymbol{M}$, where μ is an arbitrary scale factor and we write the 4-vector $\boldsymbol{M}^{\top} = \begin{pmatrix} \overline{\boldsymbol{M}} & m \end{pmatrix}^{\top}$. Moreover, the homogeneous coordinates of $\boldsymbol{M}$ are normalized such that $\|\boldsymbol{M}\|^2 = 1$. If the coordinates of $\boldsymbol{M}$ verify $\boldsymbol{a}^{\top}\overline{\boldsymbol{M}} + am \neq 0$ then the point in question is not a point at infinity; we can write $\boldsymbol{N}^{\top} = \begin{pmatrix} \overline{\boldsymbol{N}} & 1 \end{pmatrix}^{\top}$ and we obtain:

$$\overline{\boldsymbol{N}} = \frac{1}{\boldsymbol{A}^{\top}\boldsymbol{M}}\mathbf{K}^{-1}\overline{\boldsymbol{M}} \tag{26}$$

where $\boldsymbol{A}^{\top} = (\boldsymbol{a}^{\top}\ a)$ denotes the fourth row of matrix $\mathbf{H}_{PE}$, i.e., the plane at infinity.

Let $\hat{\mathbf{H}}_{PE}$ be an estimation of $\mathbf{H}_{PE}$. We obtain a similar expression for $\hat{\overline{\boldsymbol{N}}}$ – the estimated Euclidean coordinates. With $\hat{\overline{\boldsymbol{M}}} = \overline{\boldsymbol{M}} + d\overline{\boldsymbol{M}}$ and after some simple algebraic manipulations we obtain:

$$\hat{\overline{\boldsymbol{N}}} = \frac{\boldsymbol{A}^{\top}\boldsymbol{M}}{\hat{\boldsymbol{A}}^{\top}\hat{\boldsymbol{M}}}\,\hat{\mathbf{K}}^{-1}\mathbf{K}\,\overline{\boldsymbol{N}} + \frac{1}{\hat{\boldsymbol{A}}^{\top}\hat{\boldsymbol{M}}}\hat{\mathbf{K}}^{-1}d\overline{\boldsymbol{M}}$$

The matrix $\mathbf{K}$ of intrinsic parameters is the one given by eq. (2) with $\beta = k\alpha$ and $r = 0$. The estimated camera matrix $\hat{\mathbf{K}}$ has as entries $k_{ij} + dk_{ij}$. By using the first order Taylor expansion we obtain: $\hat{\mathbf{K}}^{-1}\mathbf{K} = \mathbf{I} + \mathbf{K}_\varepsilon$ with:

$$\mathbf{K}_\varepsilon = \begin{pmatrix} -\frac{d\alpha}{\hat{\alpha}} & 0 & -\frac{du_0}{\hat{\alpha}} \\ 0 & -\frac{d\beta}{\hat{\beta}} & -\frac{dv_0}{\hat{\beta}} \\ 0 & 0 & 0 \end{pmatrix} \tag{27}$$

With $\hat{\boldsymbol{A}}^\top \hat{\boldsymbol{M}} = \boldsymbol{A}^\top \boldsymbol{M} + d(\boldsymbol{A}^\top \boldsymbol{M})$ we have:

$$\frac{\boldsymbol{A}^\top \boldsymbol{M}}{\hat{\boldsymbol{A}}^\top \hat{\boldsymbol{M}}} = 1 - \frac{d(\boldsymbol{A}^\top \boldsymbol{M})}{\hat{\boldsymbol{A}}^\top \hat{\boldsymbol{M}}}$$

Finally, $\hat{\overline{\boldsymbol{N}}}$ becomes:

$$\begin{aligned} \hat{\overline{\boldsymbol{N}}} &= \overline{\boldsymbol{N}} + d\overline{\boldsymbol{N}} \\ &= \overline{\boldsymbol{N}} + \underbrace{\mathbf{K}_\varepsilon \overline{\boldsymbol{N}}}_{1} - \underbrace{\frac{d(\boldsymbol{A}^\top \boldsymbol{M})}{\hat{\boldsymbol{A}}^\top \hat{\boldsymbol{M}}} \overline{\boldsymbol{N}}}_{2} + \underbrace{\frac{1}{\hat{\boldsymbol{A}}^\top \hat{\boldsymbol{M}}} \hat{\mathbf{K}}^{-1} d\overline{\boldsymbol{M}}}_{3} \end{aligned} \tag{28}$$

Let us analyze one by one these three error terms:

- *The first term depends on matrix* $\boldsymbol{K}_\varepsilon$ defined by eq. (27). Numerically, α and β are one order of magnitude greater than the image center coordinates, u_0 and v_0. Therefore, camera calibration may tolerate large errors in the estimation of the position of the optical center. In practice, as it can be observed in tables I and II, the entries $\frac{du_0}{\hat{\alpha}}$, $\frac{dv_0}{\hat{\beta}}$, $\frac{d\alpha}{\hat{\alpha}}$, and $\frac{d\beta}{\hat{\beta}}$ have the same order of magnitude.
- *The second term depends on the projective-to-affine error* $\frac{d(\boldsymbol{A}^\top \boldsymbol{M})}{\hat{\boldsymbol{A}}^\top \hat{\boldsymbol{M}}}$ which is a function of the accuracies with which the projective coordinates and the plane at infinity are estimated.
- *The third term is a combination of the projective error*[1] $(d\overline{\boldsymbol{M}})$, *the projective-to-affine estimated upgrade* $(\hat{\boldsymbol{A}}^\top \hat{\boldsymbol{M}})$, and *the affine-to-Euclidean estimated upgrade* $(\hat{\mathbf{K}}^{-1})$.

The above analysis reveals that while the terms 1 and 2 above are bounded errors, the projective-to-affine upgrade error (term 3) is not bounded because it's value depends on

[1]Remember that the projective coordinates are normalized such that the coordinates $\boldsymbol{M}$ of a point M satisfy $\|\boldsymbol{M}\|^2 = 1$.

the position in space of the reconstructed points. Therefore we can conclude that the projective-to-affine error dominates the self-calibration process.

VI. Implementation

The method described above was implemented and tested with both simulated and experimental data. Simulated data allows us to systematically study the sensitivity of the method with respect to image noise and to assess the conditions under which reliable results are expected.

We used two types of experimental data: "calibrated data" and natural data. Calibrated data are obtained with a specially designed 3-D object which consists of 150 circular features evenly distributed onto three mutually orthogonal planes. The Cartesian coordinates these features are known with an accuracy of 0.01mm. The images of these circular features are processed using standard photogrammetric techniques and they are localized within 0.05 pixels. These highly precise 3-D and 2-D data allow us to calibrate the cameras using standard off-line calibration techniques and use the internal camera parameters thus obtained as the ground truth. When the calibrated data are used in conjunction with the method described in this paper, only the 2-D data are necessary (the 3-D coordinates of the circular features are not used).

Natural data consist of stereo pairs of various 3-D objects with no circular features on them – therefore their features are not localized with the accuracy with which the circular features are localized. The difference in accuracy between calibrated and natural data make the self-calibration process more difficult to achieve.

A. Noise sensitivity analysis

We studied the behavior of the method as a function of image noise and as a function of the number of motions of the stereo rig. According to section II three camera models are studied: a camera with three parameters (P3), a camera with four parameters (P4) and a camera with five parameters (P5).

In order to have a global view we plotted the values obtained for camera calibration over all trials: both the camera model (P3, P4, and P5) and the number of motions vary but, for each plot, the standard deviation of the added noise is fixed. The distributions that we obtained for the camera parameters are shown on Figure 1 (top and middle).

B. Experiments with natural data

The experiments that we carried out with natural data were performed with two types of stereo sensors:

- A camera pair mounted onto a metal slide and which can perform free motions and
- A stereo head mounted onto a 5 degrees of freedom robot with three prismatic joints and two revolute joints.

The first camera pair was used to perform general motions while the second camera pair was used to perform planar motions. In both cases interest points were extracted using an improved version of the Harris operator and these points are tracked over time using a stereo tracker. Figure 1 (bottom) shows the result of tracking interest points for one planar motion.

The results of metric calibration using the first stereo rig are shown on Table I. The first row shows the results of off-line calibration which is taken as the ground-truth, as already explain. The second row shows the result of calibration using 5 general motions (6 image pairs). The last row shows the reconstruction error that results from this calibration data using the error formulation of eq. (27).

The second stereo rig was used to perform metric calibration using several distinct planar motions. Table II summarizes this experiment. Again, the stereo rig was calibrated using an off-line calibration technique (first row) and these data are taken as the ground truth for assessing the quality of self calibration. The second row shows the results of metric calibration of the left and right cameras from 6 planar motions in 6 distinct planes. As in the previous experiment, the third row shows the reconstruction error.

VII. Discussion

In this paper we described a method for recovering camera calibration from rigid motions of an uncalibrated stereo rig. We showed how to estimate the plane at infinity from several

TABLE I

THE RESULTS OF CALIBRATION USING 5 GENERAL MOTIONS OF THE STEREO RIG ARE COMPARED WITH A STANDARD OFF-LINE METHOD WHICH IS TAKEN AS THE GROUND-TRUTH.

	left camera				right camera			
Method	α	$k\alpha$	u_0	v_0	α	$k\alpha$	u_0	v_0
Off-line	1534	1528	270	265	1520	1514	264	271
5 general motions	1544	1536	262	295	1539	1535	273	292
Reconstruction error	0.65%	0.52%	0.52%	1.9%	1.25%	1.4%	0.59%	1.4%

TABLE II

SAME AS THE TABLE ABOVE BUT WITH 6 PLANAR MOTIONS IN 6 DISTINCT PLANES.

	left camera				right camera			
Method	α	$k\alpha$	u_0	v_0	α	$k\alpha$	u_0	v_0
Off-line	1534	1539	255	267	1516	1521	260	253
6 planar motions	1513	1507	250	255	1506	1502	261	246
Reconstruction error	1.37%	2.08%	0.32%	0.84%	1.25%	1.25%	0.06%	0.46%

general or planar motions and how to convert the affine calibration thus obtained into metric calibration. The method relies on linear algebraic techniques. Moreover, the method requires the epipolar geometry between the left and right cameras but does not require the epipolar geometry associated with the motion of each one of these cameras. This is both an improvement and an advantage over existing techniques because the epipolar geometry associated with the stereo rig remains unchanged during the motion and therefore it can be estimated using all the image pairs over the sequence.

An error analysis emphasizes the importance of projective-to-affine upgrade over affine-to-metric upgrade. Hence, the fact that the plane at infinity is estimated as the common root of a set of linear equations, as explained and experimented in this paper, considerably improves the quality of the final metric calibration and reconstruction.

Extensive experiments with simulated data allow both a statistical characterization of the behavior of the method and a noise sensitivity analysis. Based on this analysis one can conclude that the method tolerates Gaussian noise with standard deviation up to 0.5 pixels.

The method has been extensively evaluated with three camera models. Indeed, the question of weather one should use a 5-, 4-, or a 3-parameter camera was somehow open. The statistical analysis does not reveal that one model is more resistant to noise than another. In practice we believe that a 4-parameter camera is the most suited model.

References

[1] P.A. Beardsley, I.D. Reid, A. Zisserman, and D.W. Murray. Active visual navigation using non-metric structure. In E. Grimson, editor, *Proceedings of the 5th International Conference on Computer Vision, Cambridge, Massachusetts, USA*, pages 58–64. IEEE Computer Society Press, June 1995.

[2] D. Demirdjian, G. Csurka, and R. Horaud. Autocalibration in the presence of critical motions. In *Proceedings British Machine Vision Conference*, pages 751–759, Southampton, Great Britain, September 1998. Springer-Verlag.

[3] F. Devernay and O. Faugeras. From projective to Euclidean reconstruction. In *Proceedings Computer Vision and Pattern Recognition Conference*, pages 264–269, San Francisco, CA., June 1996.

[4] R. I. Hartley. In defence of the 8-point algorithm. In *Proceedings Fifth International Conference on Computer Vision*, pages 1064–1070, Cambridge, Mass., June 1995. IEEE Computer Society Press, Los Alamitos, Ca.

[5] R. I. Hartley and P. F. Sturm. Triangulation. *Computer Vision and Image Understanding*, 68(2):146–157, November 1997.

[6] R. A. Horn and C. A. Johnson. *Matrix Analysis*. Cambridge University Press, Cambridge, 1994.

[7] Q-T. Luong. *Matrice Fondamentale et Autocalibration en vision par ordinateur*. PhD thesis, Université de Paris Sud, Orsay, December 1992.

[8] 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.

[9] T. Moons, L. Van Gool, M. Proesmans, and E Pauwels. Affine reconstruction from perspective image pairs with a relative object-camera translation in between. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(1):77–83, January 1996.

[10] A. Ruf, G. Csurka, and R. Horaud. Projective translations and affine stereo calibration. In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, pages 475–481, Santa Barbara, CA, June 1998. IEEE Computer Society Press, Los Alamitos, Ca.

[11] A. Ruf and R. Horaud. Visual servoing of robot manipulators, part I: Projective kinematics. *International Journal of Robotics Research*, 18(11):1101–1118, November 1999.

[12] B. Triggs. Autocalibration and the absolute quadric. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Puerto Rico, USA*, pages 609–614. IEEE Computer Society Press, June 1997.

[13] 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, Cambridge, Mass., June 1995.

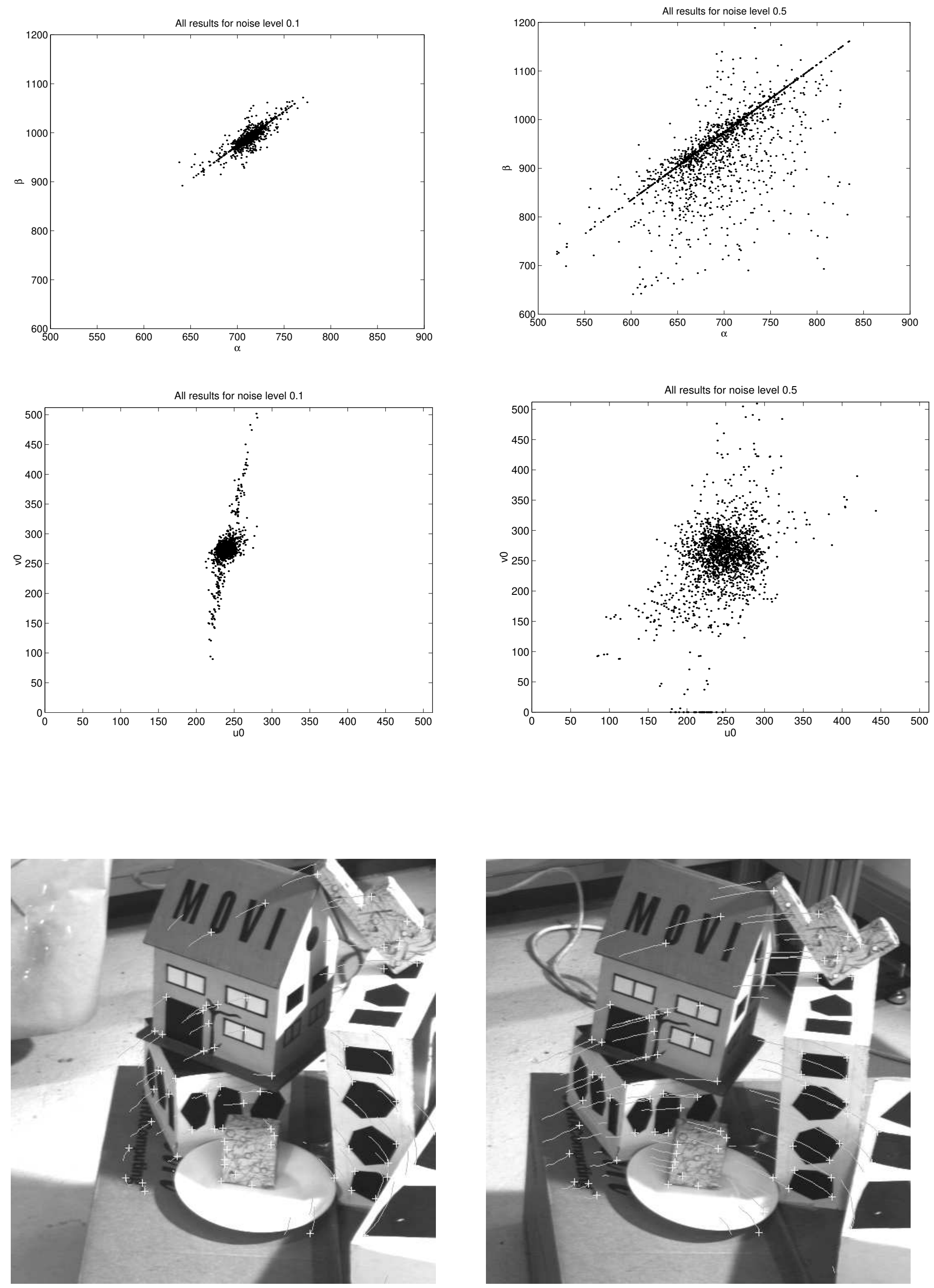

Fig. 1. Top: the statistical distribution of the horizontal and vertical image scale factors computed over a large number of experiments and for 2 levels of image noise: 0.1 and 0.5 pixels. Middle: the optical center. Bottom: Tracking points over image-pairs while the rig performs a planar motion.