HAL
open science

Stereo Calibration from Rigid Motions

Radu Horaud, Gabriella Csurka, David Demirdjian

To cite this version:

Radu Horaud, Gabriella Csurka, David Demirdjian. Stereo Calibration from Rigid Motions. RR-3467, INRIA. 1998. inria-00073223

HAL Id: inria-00073223

https://inria.hal.science/inria-00073223v1

Submitted on 24 May 2006

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.

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Stereo Calibration from Rigid Motions

Radu Horaud, Gabriella Csurka, and David Demirdjian

No 3467

Août 1998

THÈME 3

Rapport de recherche

ISSN 0249-6399

Stereo Calibration from Rigid Motions

Radu Horaud, Gabriella Csurka, and David Demirdjian

Thème 3 — Interaction homme-machine,
images, données, connaissances

Projet Movi

Rapport de recherche n° 3467 — Août 1998 — 30 pages

Abstract: This paper describes a method to upgrade projective reconstruction to affine and to metric reconstructions using rigid general or planar motions of a stereo rig. We make clear the algebraic relationships between projective reconstruction, the plane at infinity (affine reconstruction), camera calibration, and metric reconstruction when a 3-D scene is observed with a moving stereo rig. Based on an in-depth algebraic analysis we show that all the computations can be carried out using standard linear resolution methods. We carry out a theoretical error analysis which quantify the relative importance of the accuracies of projective-to-affine conversion and affine-to-Euclidean conversion. Extensive experiments performed with calibrated and natural data confirm the theoretical error analysis and are consistent with a sensitivity analysis performed with simulated data.

Key-words: self-calibration, projective reconstruction, metric reconstruction, rigid motion, stereo vision, affine calibration, epipolar geometry.

(Résumé : tsvp)

This work was supported by Société Aérospatiale

Unité de recherche INRIA Rhône-Alpes
655, avenue de l'Europe, 38330 MONTBONNOT ST MARTIN (France)
Téléphone : 04 76 61 52 00 - International: +33 4 76 61 52 00
Télécopie : 04 76 61 52 52 - International: +33 4 76 61 52 52

Calibration stéréo avec des mouvements rigides

Résumé : Cet article présente une méthode qui permet de convertir une reconstruction projective en une reconstruction affine ou en une reconstruction métrique utilisant les mouvements rigides d'une tête stéréo. Nous rendons explicite les relations algébriques existant entre une reconstruction projective, le plan à l'infini, la calibration d'une caméra et une reconstruction métrique lorsque la scène tri-dimensionnelle est observée avec une tête stéréo en mouvement. Sur la base de cette analyse nous sommes capables de montrer que tous les calculs peuvent être faits avec des techniques de résolution linéaire. Nous analysons l'erreur et quantifions l'importance relative de la calibration affine par rapport à la calibration métrique. De nombreuses expériences effectuées avec des données simulées, calibrées et naturelles confirment l'analyse de l'erreur et sont cohérentes avec une analyse de sensibilité au bruit.

Mots-clé : auto-calibration, reconstructio projective, reconstruction métrique, mouvement rigide, vision stéréo, calibration affine, géométrie épipolaire.

1 Introduction, background, and contribution

In this paper we address the following problem: An uncalibrated stereo rig observes an unknown 3-D scene while it performs a set of rigid motions. A 3-D Euclidean reconstruction of the scene is desired. In the general case, 3-D structure can be recovered only up to a 3-D projective transformation. However, if the stereo rig undergoes rigid motions, the projective ambiguity can be reduced to affine or to Euclidean. It is well known that the process of converting projective reconstruction into Euclidean reconstruction is equivalent to camera calibration.

The relationship between projective, affine, metric spaces and camera calibration has been thoroughly investigated both in the case of a moving unique camera and of a moving stereo rig. The Kruppa equations [18], [7], [16], [12] consists of a system of polynomial equations relating the intrinsic camera parameters to the epipolar geometry between two views taken with the camera. However, solving the Kruppa equations requires non-linear resolution methods. An alternative solution consists to, first, recover affine structure and, second, solve for camera calibration using the affine structure. This stratified approach [8] can be applied to a single camera in motion [9], [17] or to a stereo rig in motion [24], [6].

Affine calibration amounts to recover the position of the plane at infinity or, equivalently, the infinite homography between two views [22]. In practice this may be done using three classes of methods:

(i) *special camera motions* such as pure translations of a stereo rig [19], [21], rotations around the camera's center of projection [10], [20], planar motions of a stereo rig [2], [3] or of a single camera [1];

(ii) *exploiting special scene structure* such as parallel lines, or

(iii) *using fixed entities* under rigid motion [24].

In this paper we investigate linear algebraic methods for recovering metric structure, affine calibration, and intrinsic camera parameters with an uncalibrated stereo rig, by performing a set of rigid motions. 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 collineation $\mathbf{H}_{12}$ which is related to the rigid motion $\mathbf{D}_{12}$ by ([24], [6]):

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

where "$\simeq$" designates projective equality (defined up to a scale factor) and $\mathbf{H}_{PE}$ is a 4×4 collineation allowing the projective reconstruction to be upgraded to an Euclidean one. It will be shown that this collineation encapsulates affine calibration of the stereo rig and the intrinsic parameters of the left camera. 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 $\boldsymbol{N}_2 \simeq \mathbf{D}_{12}\boldsymbol{N}_1$.

Zisserman et al. [24] showed that the plane at infinity can be recovered from one eigenvector of matrix $\mathbf{H}_{12}^{-\top}$ and that intrinsic parameters of the left or right cameras can be recovered from three *virtual image points* that are fixed under Euclidean motion. Their method necessitates the computation of the epipolar geometries associated with the left and right camera motions in addition to the epipolar geometry associated with the stereo rig itself. The method described in this paper only needs the epipolar geometry of the stereo rig.

Devernay and Faugeras [6] showed that one possible factorization of $\mathbf{H}_{12}$ in eq. (1) is such that $\mathbf{H}_{PE}$ becomes a lower triangular matrix and the fourth row vector of this matrix is the plane at infinity. The authors propose a non-linear minimization method to directly estimate Euclidean upgrading, i.e., the entries of $\mathbf{H}_{PE}$, from point correspondences between two stereo image pairs (before and after the motion). The method of Devernay and Faugeras gives interesting algebraic insights, although the algebraic properties associated with $\mathbf{H}_{PE}$ are not used in practice. Moreover, the intrinsic camera parameters do not appear explicitly. In practice it is sometimes useful to assume that some of the intrinsic camera parameters are known (such as image skew) but this type of constraint cannot be used with [6].

This paper has the following contributions. We show that with an appropriate choice for the Cartesian reference frame associated with the rigid motion, the matrix $\mathbf{H}_{PE}$ is parameterized by the plane at infinity and by the intrinsic parameters of the left camera. So, the homography $\mathbf{H}_{PE}$ in eq. (1) directly encapsulates projective to Euclidean upgrading, affine calibration and left-camera calibration. This particular parameterization of $\mathbf{H}_{PE}$ allows for an error analysis which determines the relative importance of affine calibration and metric calibration as well as the relative importance of the various intrinsic camera parameters. This error analysis reveals that the error generated by the projective-to-affine upgrade is considerably larger than the errors associated with projective reconstruction and with the affine-to-Euclidean upgrade. Therefore, projective-to-affine calibration must be thoroughly studied.

The advantage of using a moving stereo rig rather than a moving single camera is that the plane at infinity is an eigenvector of $\mathbf{H}_{12}^{-\top}$ or of $\mathbf{H}_{12}^{\top}$. We show that for a sequence of general rigid motions the corresponding collineations *have a common eigenvector* associated with the double eigenvalue 1. This eigenvector is the plane at infinity of the stereo rig and therefore it is an intrinsic property of the rig. We extend this result to a sequence of distinct planar motions.

This property allows us to estimate the plane at infinity from any number of motions (general, planar, or a combination of both), the eigenvector being the common root of a set of linear equations. Once this eigenvector (the plane at infinity) has been recovered, the parameterization of $\mathbf{H}_{12}$ in terms of $\mathbf{H}_{PE}$ and $\mathbf{D}_{12}$ provides a simple algebraic expression for the infinite homography between the images associated with the left camera before and after

a motion. This means that, unlike the Kruppa equations and unlike the method described in [24] it is not necessary to determine the epipolar geometry associated with the left (or right) camera motion.

We describe extensive experiments done with simulated data, calibrated data, and real data. The noise sensitivity analysis performed with simulated data allows to determine the optimal experimental conditions under which the method is expected to yield reliable camera calibration and metric reconstruction. The experiments performed with calibrated data allows to compare this self-calibration method with more classical off-calibration methods where the 3-D Euclidean geometry of the calibrating object is known in advance. The results obtained with natural data confirm both the error analysis and the noise sensitivity analysis, namely that some calibration parameters are more critical than others and that image point localization with sub-pixel accuracy is crucial.

Aside from the theoretical contributions cited above which make self-calibration of a stereo rig to be very attractive, there is another major advantage of a camera pair over a single camera. When a single camera moves the epipolar geometry between the first, second, third positions an so forth changes with camera motion and therefore a new fundamental matrix has to be estimated each time a new camera motion is performed. When a stereo rig moves the epipolar geometry between the left and right cameras of the stereo pair remains unchanged. Therefore one can use all the image pairs available along a sequence to estimate the same fundamental matrix. Moreover, the stereo rig motion can compensate for flat scenes which are known to be an important source of numerical instability in fundamental matrix estimation.

1.1 Paper organization

The remainder of the paper is organized as follows. Section 2 recalls the projective geometry associated with a stereo rig: projective reconstruction in a sensor-centered projective basis. Section 3 analyses in detail the algebraic properties of an uncalibrated stereo rig undergoing rigid motion or, equivalently, a moving rigid scene being observed by a stereo rig. Section 4 establishes a parameterization for the projective to Euclidean upgrade transformation (proposition 1) and describes how to compute the plane at infinity of a stereo rig either from a sequence of general motions (corollary 1.1) or from a sequence of distinct planar motions (corollary 1.2). It is shown that the infinite homography associated with the left (or right) camera motion can be obtained in closed form. Section 5 provides an algebraic expression for the metric reconstruction error as a function of projective reconstruction, affine calibration, and metric calibration. Implementation, simulations and experimental results are described in detail in section 6. Finally section 7 provides a discussion and directions for future work.

2 Preliminaries

2.1 Notations

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. It is important two distinguish between homogeneous 4-vectors which are designated by upper case letters and homogeneous 3-vectors which are designated by lower case letters. A 4-vector will often be designated as the concatenation of a 3-vector with a scalar: $\boldsymbol{A} = \begin{pmatrix} \boldsymbol{a} \\ a \end{pmatrix}$. 3-vectors will sometime be designated by over-lined lower case letters ($\overline{\boldsymbol{a}}$). $\mathbf{H}^\top$ is the transpose of $\mathbf{H}$.

2.2 Camera models

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:

$$\boldsymbol{m} \simeq \mathbf{P}\boldsymbol{M} \tag{2}$$

The equal sign designates the projective equality – equality up to a scale factor. 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):

$$\mathbf{P}_E \simeq \mathbf{K} \begin{pmatrix} \mathbf{R} & \boldsymbol{t} \end{pmatrix} \simeq \begin{pmatrix} \mathbf{KR} & \mathbf{K}\boldsymbol{t} \end{pmatrix} \tag{3}$$

If we choose the standard camera frame as the 3-D Euclidean frame (the origin is the center of projection, the xy-plane is parallel to the image plane and the z-axis points towards the visible scene), the rotation matrix $\mathbf{R}$ is equal to the identity matrix and the translation vector $\boldsymbol{t}$ is the null vector. The projection matrix becomes:

$$\mathbf{P}_E \simeq \begin{pmatrix} \mathbf{K} & \mathbf{0} \end{pmatrix} \tag{4}$$

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

$$\mathbf{K} = \begin{pmatrix} \alpha & r\alpha & u_0 \\ 0 & k\alpha & v_0 \\ 0 & 0 & 1 \end{pmatrix} \tag{5}$$

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 $k = 1$

In practice the scale factor can be obtained from the cameras' and frame grabbers' technical sheets. In general it is not equal to 1 but the image coordinates can be rescaled such that we get $k = 1$.

2.3 The geometry of a stereo rig

A stereo rig is composed of two cameras fixed together. Let $\mathbf{P}$ and $\mathbf{P}'$ be the projection matrices of the left and right cameras. We can write these 3×4 matrices as:

$$\mathbf{P} \simeq \begin{pmatrix} \overline{\mathbf{P}} & \boldsymbol{p} \end{pmatrix}$$

$$\mathbf{P}' \simeq \begin{pmatrix} \overline{\mathbf{P}}' & \boldsymbol{p}' \end{pmatrix}$$

It is useful to recall the expressions of the infinite homography between the left and right images as well as the left and right epipoles:

$$\mathbf{H}_\infty \simeq \overline{\mathbf{P}}'\overline{\mathbf{P}}^{-1} \tag{6}$$

and

$$\boldsymbol{e} \simeq -\mathbf{H}_\infty^{-1}\boldsymbol{p}' + \boldsymbol{p} \tag{7}$$

$$\boldsymbol{e}' \simeq -\mathbf{H}_\infty\boldsymbol{p} + \boldsymbol{p}' \tag{8}$$

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

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

$$\mathbf{P}' \simeq \begin{pmatrix} \overline{\mathbf{P}}' & \boldsymbol{p}' \end{pmatrix} \tag{10}$$

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):

$$\begin{aligned} \mathbf{P}_E &\simeq \begin{pmatrix} \mathbf{K} & \mathbf{0} \end{pmatrix} \\ \mathbf{P}'_E &\simeq \begin{pmatrix} \mathbf{K}'\mathbf{R} & \mathbf{K}'\boldsymbol{t} \end{pmatrix} \end{aligned} \tag{11}$$

With these expressions for $\mathbf{P}$ and $\mathbf{P}'$ we obtain the following parameterizations:

$$\mathbf{H}_\infty \simeq \mathbf{K}'\mathbf{R}\mathbf{K}^{-1} \tag{12}$$

$$\boldsymbol{e} \simeq -\mathbf{K}\mathbf{R}^\top\boldsymbol{t} \tag{13}$$

$$\boldsymbol{e}' \simeq \mathbf{K}'\boldsymbol{t} \tag{14}$$

2.4 Projective reconstruction with a stereo rig

Given a stereo rig with two projection matrices $\mathbf{P}$ and $\mathbf{P}'$, it is possible to compute the 3-D projective coordinates of a point $\boldsymbol{M}$ from the equations $\mu\boldsymbol{m} = \mathbf{P}\boldsymbol{M}$ and $\mu'\boldsymbol{m}' = \mathbf{P}'\boldsymbol{M}$, where $\boldsymbol{m}$ and $\boldsymbol{m}'$ are the projections of $\boldsymbol{M}$ onto the left and right images and μ and μ' are two unknown scale factors:

$$\begin{pmatrix} \mathbf{P} & \boldsymbol{m} & \mathbf{0} \\ \mathbf{P}' & \mathbf{0} & \boldsymbol{m}' \end{pmatrix} \begin{pmatrix} \boldsymbol{M} \\ \mu \\ \mu' \end{pmatrix} = \mathbf{0} \tag{15}$$

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 [23], [11]. Several authors proved that the two projection matrices can be obtained from the epipolar geometry up to a 4-parameter projective mapping [17]:

$$\mathbf{P} \simeq \begin{pmatrix} \mathbf{I} & \mathbf{0} \end{pmatrix} \tag{16}$$

$$\mathbf{P}' \simeq \begin{pmatrix} \mathbf{H}_\pi & a\,\boldsymbol{e}' \end{pmatrix} \tag{17}$$

with

$$\mathbf{H}_\pi \simeq \mathbf{H}_\infty + \boldsymbol{e}'\boldsymbol{a}^\top \tag{18}$$

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.

2.5 Projective basis associated with a stereo rig

Equation (15) allows one to compute the projective coordinates of a 3-D point M in a *sensor-centered projective basis.* A projective basis is defined by 5 points in general position and let us make explicit the physical positions of these points. We show that these points are determined from $\mathbf{P}$ and $\mathbf{P}'$ defined above (equations (16) and (17)) and therefore they are linked to the stereo camera pair, Figure 1.

The first point is the center of projection of the left camera, denoted by C and defined as the null vector of the projection matrix:

$$\mathbf{P}\boldsymbol{C} \simeq \mathbf{0}$$

We obtain: $\boldsymbol{C} \simeq \begin{pmatrix} 0 & 0 & 0 & 1 \end{pmatrix}^\top$. The second point is the center of projection of the right camera whose projection onto the left camera is the left epipole:

$$\mathbf{P}\boldsymbol{C}' \simeq \boldsymbol{e} = \lambda\boldsymbol{e}$$

We obtain:

$$\boldsymbol{C}' \simeq \begin{pmatrix} \lambda \boldsymbol{e} \\ c' \end{pmatrix}$$

$\boldsymbol{C}'$ is the null vector of the projection matrix $\mathbf{P}'$: $\mathbf{P}\boldsymbol{C}' = \mathbf{0}$ from which we get:

$$\lambda \mathbf{H}_\pi \boldsymbol{e} + c' \boldsymbol{e}' = \mathbf{0}$$

With $\boldsymbol{e}' \simeq \mathbf{H}_\pi \boldsymbol{e}$ we obtain $c' = -1$ and finally we get:

$$\boldsymbol{C}' \simeq \begin{pmatrix} -\boldsymbol{e} \\ 1 \end{pmatrix}$$

The three remaining points can be chosen to lie in the space plane associated with the plane homography $\mathbf{H}_\pi$. In theory, there is a 3-parameter family of plane homographies satisfying the same epipolar geometry, i.e., $\mathbf{H}_\pi + \boldsymbol{e}'\boldsymbol{v}^\top$ for any 3-vector $\boldsymbol{v}$. In practice one can choose one such vector and without loss of generality we take $\boldsymbol{v} = \mathbf{0}$. Let A_1, A_2, and A_3 be three points belonging to the space plane π. Clearly the left-image and right-image projections of these points must simultaneously satisfy $\mathbf{P}\boldsymbol{A}_i \simeq \boldsymbol{a}_i$, $\mathbf{P}'\boldsymbol{A}_i \simeq \boldsymbol{a}'_i$, and $\boldsymbol{a}'_i \simeq \mathbf{H}_\pi \boldsymbol{a}_i$. Finally we obtain the following projective basis:

$$\boldsymbol{C} \simeq \begin{pmatrix} \mathbf{0} \\ 1 \end{pmatrix} \quad \boldsymbol{A}_1 \simeq \begin{pmatrix} \boldsymbol{a}_1 \\ 0 \end{pmatrix} \quad \boldsymbol{A}_2 \simeq \begin{pmatrix} \boldsymbol{a}_2 \\ 0 \end{pmatrix} \quad \boldsymbol{A}_3 \simeq \begin{pmatrix} \boldsymbol{a}_3 \\ 0 \end{pmatrix} \quad \boldsymbol{C}' \simeq \begin{pmatrix} -\boldsymbol{e} \\ 1 \end{pmatrix}$$

Moreover, if the left image points $\boldsymbol{a}_1$, $\boldsymbol{a}_2$, $\boldsymbol{a}_3$, and $-\boldsymbol{e}$ are given the coordinates of the canonical basis of the projective plane, we obtain the canonical basis of the projective space, (0001), (1000), (0100), (0010), and (1111).

3 Rigid motion

3.1 Sensor motion

The projective basis just defined is physically attached to the camera pair. If the camera pair undergoes a 3-D rigid motion (the cameras are rigidly attached to each other and their internal parameters remain unchanged) then the projective basis undergoes a rigid motion as well. We consider two positions of the sensor, i.e., the stereo pair, position 1 and position 2, before and after such a motion. The projective coordinates of a 3-D point M are related to its Euclidean coordinates by the formula:

$$\boldsymbol{N}_i \simeq \mathbf{H}_{PE} \boldsymbol{M}_i$$

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 camera pair