INRIA
INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Using Singular Displacements for Uncalibrated Monocular Visual Systems

Thierry Viéville and Diane Lingrand

N° 2678

Octobre 1995

PROGRAMME 4

ISSN 0249-6399

SOPHIA ANTIPOLIS

Using Singular Displacements for Uncalibrated Monocular Visual Systems

Thierry Viéville and Diane Lingrand

Programme 4 — Robotique, image et vision
Projet RobotVis

Rapport de recherche n ° 2678 — Octobre 1995 — 32 pages

Abstract:

In the present paper, we review and complete the equations and the formalism which allow to achieve a minimal parameterization of the retinal displacement for a monocular visual system without calibration.

Considering the emergence of active visual systems for which we **can not** consider that the calibration parameters are either known or fixed, we develop an alternative strategy using the fact that certain class of special displacements induces enough equations to evaluate the calibration parameters, so that we can recover the affine or Euclidean structure of the scene when needed.

A synthesis of what can be recovered for singular displacements in terms of camera calibration, scene geometry and kinematics is proposed. We give, for the different levels of calibration, an exhaustive list of the geometric and kinematic information which can be recovered. Following a strategy based on special kind of displacements, such as fixed axis rotations or pure translations for instance, we describe how to detect this particular classes of displacement.

The implementation of these equations is analyzed here and the realization is reported as a single module, to be embedded in a high-level Image Understanding Environment.

Key-words: Structure and Motion, Singular Displacements, Self-Calibration

(Résumé : tsvp)

Unité de recherche INRIA Sophia-Antipolis
2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex (France)
Téléphone : (33) 93 65 77 77 – Télécopie : (33) 93 65 77 65

Utilisation de Déplacements Singuliers pour les Systèmes Visuels non Calibrés

Résumé :

Ce papier est une revue, complétée de développements nouveaux, des équations et du formalisme qui permettent d'obtenir une paramétrisation minimale du mouvement rétinien pour un système visuel monoculaire non calibré.

Considérant l'émergence de systèmes visuels actifs pour lesquels on ne peut plus considérer que les paramètres de calibration sont connus ni même constants, on développe ici une stratégie alternative basée sur le fait que des classes particulières de mouvement permettent de générer assez d'équations pour évaluer les paramètres de calibration, permettant ainsi de récupérer -si besoin- la structure euclidienne de la scène observée.

Dans ce papier, une synthèse de ce qui peut être récupéré, lors de mouvements singuliers, en terme de calibration, d'attributs géométriques et cinématiques de la scène est proposée. On décrit les différents niveaux de calibration et donne une liste exhaustive des différentes informations obtenues à chaque niveau.

En suivant une stratégie basée sur certains types de déplacements, tels que -par exemple- des rotations autour d'axe fixe ou des translations pures, on décrit comment détecter de tels mouvements.

L'implémentation de ces équations est analysée dans ce papier et une réalisation sous forme d'un module logiciel à intégrer au sein d'un Environnement d'Analyse de Séquences d'Images est expérimentée.

Mots-clé : Structure et Mouvement, Mouvements Singuliers, Self-Calibration

Contents

1 Introduction 4

2 Reviewing the theory of motion when no calibration. 4
- 2.1 Setting the equations . . . 5
  - 2.1.1 Camera model and frame of reference. . . . 5
  - 2.1.2 Using points as primitives. . . . 5
  - 2.1.3 A suitable model of the intrinsic parameters of the camera. . . . 5
  - 2.1.4 Representation of rigid displacements. . . . 6
- 2.2 Parameterization of motion when no calibration. . . . 6
  - 2.2.1 The Qs-representation and the F-matrix. . . . 6
  - 2.2.2 The case of a pure rotation, and the planar case. . . . 7

3 Using specific displacements for motion analysis. 8
- 3.1 General rigid displacement. . . . 9
- 3.2 General planar rigid displacement. . . . 11
- 3.3 Elementary displacements of the camera: no translation. . . . 12
- 3.4 Elementary displacements of the camera: pure rotation. . . . 12
- 3.5 Elementary displacements of the camera: retinal translation. . . . 13
- 3.6 Elementary displacements of the camera: pure translation. . . . 14
- 3.7 Elementary displacements of the camera: pure planar translation. . . . 14
- 3.8 Elementary displacements of the camera: retinal displacement. . . . 15
- 3.9 Elementary displacements of the camera: retinal planar displacement. . . . 17
- 3.10 Elementary displacements of the camera: fixed axis rotation. . . . 18
- 3.11 Elementary displacements of the camera: zoom displacement. . . . 19

4 Defining a hierarchical motion module 21
- 4.1 Combining different models of displacements . . . 21
- 4.2 Using a statistical framework. . . . 23
  - 4.2.1 The Early-vision module. . . . 23
  - 4.2.2 Eliminating outliers. . . . 24
  - 4.2.3 Refining the estimation. . . . 24
  - 4.2.4 Computing covariances and comparing different models. . . . 24
  - 4.2.5 Implementing a motion module . . . 25
- 4.3 Experimental results . . . 26
  - 4.3.1 Using synthetic data . . . 26
  - 4.3.2 An example with real data : grid scene. . . . 29
  - 4.3.3 An example with real data : external scene. . . . 29
  - 4.3.4 An example with real data : general displacement. . . . 30

5 Conclusion 31

1 Introduction

The analysis of motion in the case of an uncalibrated monocular image sequence has already been developed by several authors, considering point and/or line correspondences or correspondences between planar patches and using either a discrete or a continuous representation of the rigid displacement between two or more frames.

These studies are motivated by the fact that *we must not consider an active visual system is calibrated* [10]. However, it has been demonstrated that, in the general case, it is not possible to self-calibrate the camera when zooming or modifying the intrinsic calibration parameters.

Considering this fact, the key idea of the present study is that **several singular displacements induce enough equations to evaluate the calibration parameters.**

For instance, fixed axis rotations of known angles or pure rotations [9] allow to estimate the calibration parameters, their uncertainty and, for a given kind of displacement, which parameters are optimally estimated, so that active visual strategies can be developed. On the other hand, pure translations do not allow to calibrate the Euclidean geometry of the scene [11], but its affine geometry [14].

Collecting all this information and considering a suitable statistical framework as in [1], it is then possible to infer which kind of displacement will increase at most the information (usually represented by the inverse of a covariance matrix) on the scene geometry, object kinematics and calibration parameters.

This is the goal of this paper.

In order to attain this objective, we are first going to review the theory of motion when no calibration: equations, parameterization of motion, etc...

We then are going to propose a synthesis of what can be recovered in terms of scene geometry and kinematics when calibration is not given as an input: describe the different forms of calibration, the different levels of calibration and give an exhaustive list of the different geometric and kinematic information to be recovered, depending the chosen geometry.

2 Reviewing the theory of motion when no calibration.

Notations.

We write vectors and matrices using bold letters, matrices being written with capital letters. The duals of vectors are represented as the transpose of a vector and scalars in italic. The notation $\mathbf{x} \wedge \mathbf{y} = \tilde{\mathbf{x}}\mathbf{y}$ corresponds to the cross-product, the dot-product being written as $\mathbf{x}^T\mathbf{y}$. $\tilde{\mathbf{x}}$ is a 3×3 skew-symmetric matrix[1] . The identity matrix is written $\mathbf{I}$. Geometric objects such as points, lines, planes are written with capital letters in 3D, and small letters in 2D. We represent the components of a matrix or a vector using superscripts from 0 to 2, e.g.: $\mathbf{x} = (x^0, x^1, x^2)^T$.
We write $\mathbf{a} \equiv \mathbf{b}$ if $\mathbf{a}$ is equal to $\mathbf{b}$ up to a scale factor, i.e. $\exists k, \mathbf{a} = k\,\mathbf{b}$.

[1] Remember that a 3×3 skew-symmetric matrix has 3 parameters and can always be represented by the crossproduct of a vector, i.e. is of the form $\tilde{\mathbf{x}}$ for some $\mathbf{x}$.

2.1 Setting the equations

2.1.1 Camera model and frame of reference.

We use *the standard pinhole model* for a camera, assuming the camera performs a perfect perspective transform with center C (the camera optic center) at a distance f (the focal length) of the retinal plane. The pinhole model can still be used for a zoom lens if the object-to-image distance is not considered as fixed.

All coordinates are related to an affine frame of reference $\mathcal{R} = (C, \mathbf{x}, \mathbf{y}, \mathbf{z})$ *attached to the retina*, $\mathbf{z}$ being aligned with the optical axis, $\mathbf{x}$ and $\mathbf{y}$ being aligned with the horizontal and vertical axis in the image. The retinal plane is thus perpendicular to the optical axis Cz, as shown in figure 1.

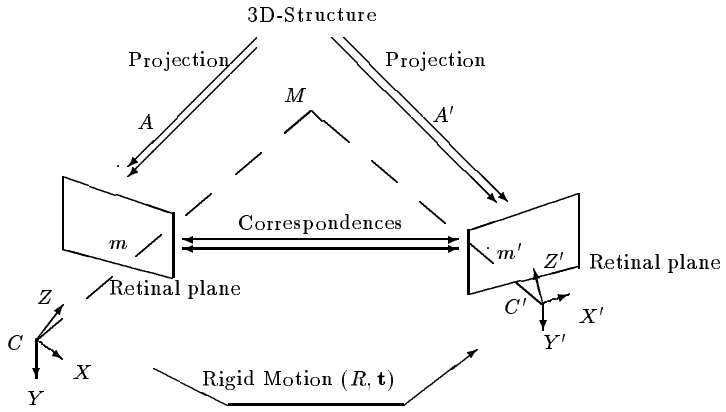


Figure 1: Elements used in the definition of the problem

2.1.2 Using points as primitives.

We represent a 3D-point M by the vector $\mathbf{M} = C\vec{M} = (X, Y, Z)^T$ using Euclidean coordinates. Points in the retina, with pixel coordinates coordinates (u, v) will be represented as homogeneous 3-D vectors: $\lambda\ \mathbf{m} = \lambda\ C\vec{m} = \lambda\ (u, v, 1)^T$, corresponding to lines of a given direction passing through the optical center (2-D projective space).

Other primitives will be represented using set of points. This will be discussed in the sequel.

2.1.3 A suitable model of the intrinsic parameters of the camera.

In this study, *we do not assume the system is calibrated.* However, we are in a specific situation because we have chosen a "canonical" frame attached to the retina. Therefore, we consider only the matrix of the intrinsic parameters (called A-matrix) in the projection and

write:

$$Z\,\mathbf{m} = \mathbf{A}\,\mathbf{M} \quad , \quad \mathbf{A} = \begin{pmatrix} f & 0 & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{pmatrix} \tag{1}$$

A complete review can be found in [1].

In the present model, (u_0, v_0) is the principal point, and f the focal length; following [11], we assume that we know the ratio between the horizontal and vertical focal length and that we assume that the two retinal coordinates are orthogonal. It has been shown experimentally that these assumptions are valid for standard cameras [11] and also for high-level visual sensors [13]. Using this simple model will allow us to improve the obtained results.

We also assume that the intrinsic parameters are different for each camera position, as during a zoom. In the consecutive frame $\mathcal{R}' = (C', \mathbf{x}', \mathbf{y}', \mathbf{z}')$ we write:

$$Z'\,\mathbf{m}' = \mathbf{A}'\mathbf{M}' \tag{2}$$

2.1.4 Representation of rigid displacements.

We consider motion of rigid objects and the ego-motion of the camera, *in the discrete case*. We thus represent motion through rigid displacements.

It means that the tokens in the scene are undergoing a rigid displacement parameterized by a rotation matrix R and a translation vector $\mathbf{t}$:

$$\mathbf{M}' = \mathbf{R}\,\mathbf{M} + \mathbf{t} \tag{3}$$

2.2 Parameterization of motion when no calibration.

The goal of the parameterization of motion is the following: given a set of points in correspondence between two views, i.e. a set of matches $\{m.m'\}$ we want to analyze all constraints which relate the two points i.e. find the equations of the form $\forall\{m.m'\}, f(m, m') = 0$. In particular, we would like to predict the location of a point given its correspondent, i.e. a relation of the form $\forall\{m.m'\}, m' = g(m)$. Having such parameterization allows to exact all information available from the retinal displacements, which is measured through the set of matches.

2.2.1 The Qs-representation and the F-matrix.

Considering the 2D correspondences between two points m and m' in two different frames, we obtain, combining equations (1),(2) and (3):

$$Z'\,\mathbf{m}' = Z\,\underbrace{\mathbf{A}'\mathbf{R}\mathbf{A}^{-1}}_{\mathbf{H}_\infty}\mathbf{m} + \underbrace{\mathbf{A}'\mathbf{t}}_{\mathbf{s}} \tag{4}$$

where the Q-matrix $\mathbf{H}_\infty$ corresponds to the "uncalibrated rotational component of the rigid displacement", or more geometrically *the collineation of the plane at infinity*, while the s-vector corresponds to the "uncalibrated translational component of the rigid displacement", also called "focus of expansion" by some authors, and more geometrically *the epipole*. These

notations have been introduced in [11] to analyze the motion of points and lines in the general case.

If we eliminate Z and Z' in equation (4) (by taking the cross-product with $\mathbf{s}$ and multiplying by $\mathbf{m}'^T$) we obtain:

$$\mathbf{m}'^T \underbrace{[\tilde{\mathbf{s}}\,\mathbf{H}_\infty]}_{\mathbf{F}}\,\mathbf{m} = 0 \tag{5}$$

The matrix $\mathbf{F} = \tilde{\mathbf{s}}\,\mathbf{H}_\infty$ is the *Fundamental matrix* and is also called the "essential matrix in the uncalibrated case". If we consider that the only information available is related to the retinal correspondences between points, without any knowledge about the depths Z, equation (5) is the only equation that can be derived [11].

Considering a set of matches related by equation (3) the equation (5) is well defined if and only if (i) $\mathbf{s} \neq 0$ and (ii) there is no linear relations between all m' and m. The degenerated cases occur only if the translation is zero, or if all points belong to the same plane [11][2]. This particular case will be analyzed in detail.

As discussed in [15] an efficient criterion is the average retinal Euclidean distances between each point and its epipolar line. The following symmetric least-square sum is minimized:

$$\mathbf{F}_\bullet = argmin_{\mathbf{F}} \underbrace{\left[\sum_{\{\mathbf{m}\}} w_{\mathbf{m}} \underbrace{\left[d(\mathbf{m}', \mathbf{F}\mathbf{m})^2 + d(\mathbf{m}, \mathbf{F}^{\mathbf{T}}\mathbf{m}')^2\right]}_{f_{\mathbf{m}}(\mathbf{F})^2}\right] / \left[2 \sum_{\{\mathbf{m}\}} w_{\mathbf{m}}\right]}_{[\epsilon_{\mathbf{F}}(\mathbf{F})]^2} \tag{6}$$

where $w_{\mathbf{m}}$ is a weighted corresponding to the precision of the match, in fact the inverse of the variance of the precision of the match. The quantity $w_{\mathbf{m}}$ is given in $pixel^{-2}$, while $\epsilon_{\mathbf{F}}(\mathbf{F})$, the *average distance to the epipolar*, is in pixel.

A camera for which F has been computed is called *a weak calibrated camera*.

The vector $\mathbf{s}$ is defined as the basis vector of the kernel of $\mathbf{F}^T$.

2.2.2 The case of a pure rotation, and the planar case.

As pointed out previously, in the case of a pure rotation or if the set of points belongs to a unique planar structure, we cannot estimate the F-matrix because all points in one view are related to points in the other view by a relation of the form:

$$\mathbf{m}' \equiv \mathbf{H}\,\mathbf{m} \tag{7}$$

which corresponds to two equations for each match.

There, if the matrix $\mathbf{F}$ is undefined, we still can estimate the matrix $\mathbf{H}$ as in [11], using $\mathbf{H} = \mathbf{I}$ as initial value.

[2]From algebraic point of view, equation (5) has singular solutions if and only if there exist a linear relation between $\mathbf{m}$ and $\mathbf{m}'$, i.e. a relation of the form $\mathbf{m}' = \mathbf{H}\,\mathbf{m}$. This situation corresponds to the case were the points are related by a collineation, i.e. correspond to a planar structure as reviewed in the sequel.

Following the same method as for the F-matrix, an efficient criterion is to minimize the residual disparity again, as in [11] and obtain $\mathbf{H}$ through:

$$\mathbf{H}_\bullet = argmin_{\mathbf{H}} \underbrace{\left[\sum_{\{\mathbf{m}\}} w_{\mathbf{m}} \underbrace{||\mathbf{m}' - \frac{\mathbf{H}\,\mathbf{m}}{((\mathbf{h}^2)^T\,\mathbf{m})}||^2}_{f_{\mathbf{m}}(\mathbf{H})^2} \right] / \left[\sum_{\{\mathbf{m}\}} w_{\mathbf{m}} \right]}_{[\epsilon_{\mathbf{H}}(\mathbf{H})]^2} \tag{8}$$

where we write $\mathbf{H} = (\mathbf{h}^0, \mathbf{h}^1, \mathbf{h}^2)$ in order to have a compact notation[3].

We need at least 4 non-collinear points, since a H-matrix is a 3x3 matrix up to a scale factor thus defined by 8 parameters, while each match provides 2 equations.

The error $\epsilon_{\mathbf{H}}(\mathbf{H})$, given in pixel, will be called *residual disparity after motion reduction* in the sequel.

Reciprocally, as soon as the points belongs to at least two planes, we can defined a F-matrix [4]. This degenerated situation thus only exists in the case of an unique plane.

3 Using specific displacements for motion analysis.

Let us now discuss situations for which the F-matrix or the H-matrix have a particular form. Considering a robotic system, it is very often that a displacement is not a general displacement but a constrained motion such as a pure translation, a fixed axis rotation, etc... as illustrated in figure 2.

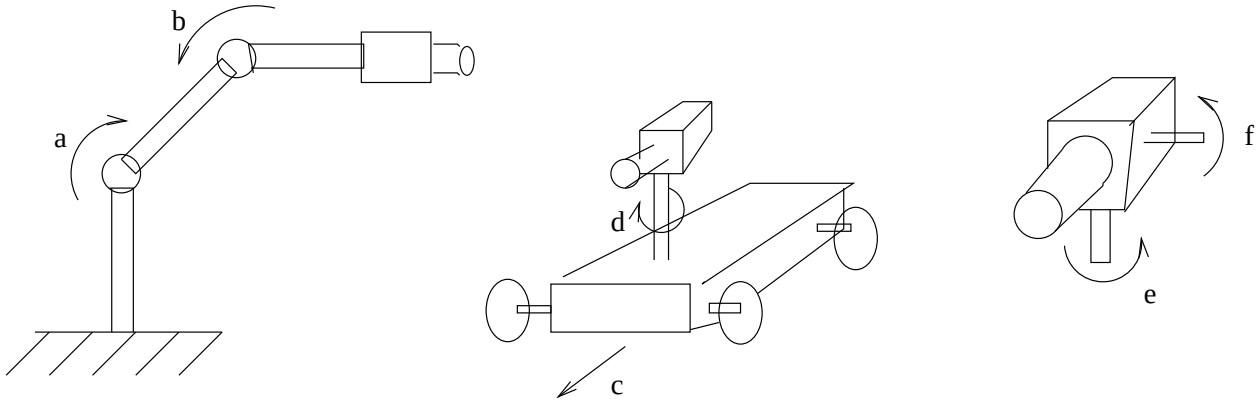
Moreover we can make the following assumptions about this kind of hardware [10]:

- The displacements are reproducible.
- We can measure the angles of the rotation. For technical reasons we do not assume the same thing for translations (it is not sure that we can estimate the norm of the linear translation of a zoom for instance [10] while the precision of the translation of a mobile robot is not very high).
- All extrinsic parameters are unknown, and equations are expected to provide unstable estimates of them [1].

These particular constraints are far from having negative properties. On the contrary, they induce additional equations which help solving the reconstruction or calibration problem.

[3] The relation $\mathbf{m}' = \frac{\mathbf{H}\,\mathbf{m}}{((\mathbf{h}^2)^T\,\mathbf{m})}$ is a vectorial form for :

$$\begin{cases} u' &= \frac{H^{00}\,u + H^{01}\,v + H^{02}}{H^{20}\,u + H^{21}\,v + H^{22}} \\ v' &= \frac{H^{10}\,u + H^{11}\,v + H^{12}}{H^{20}\,u + H^{21}\,v + H^{22}} \\ 1 &= 1 \end{cases}$$



A robotic arm A mobile robot A turret

Figure 2: Examples of robotic mechanism which generates pure translations, pure rotations or fixed axis rotations. *If, on the robotic arm, (a) and (b) have opposite values a pure translation occurs. Applying the same command (c) on both wheels of a mobile robot induces also a translation. A motion of (a) alone, or (b) alone, on the robotic arm induces a fixed axis rotation. A displacement (c) applying the opposite commands on both wheels of a mobile robot also induces a fixed axis rotation. Turrets for camera can induce pure rotations in pan (e) or tilt (f) around the optical center.*

Furthermore, the estimation of the displacement are easier in these cases, because we have to evaluate less parameters.

However, the system must be able to recognize if the displacement corresponds to such a particular case, so that we must characterize the situation in each case.

Finally, the different class of displacements might have several implications on the perception strategy, which is also to be discussed.

This is the goal of this section.

3.1 General rigid displacement.

If we consider a general rigid displacement for which we have the following decomposition of the F-matrix:

$$\begin{array}{cccccc} \mathbf{F} = & \tilde{\mathbf{s}} & \mathbf{A}' & \mathbf{R} & \mathbf{A}^{-1} & \\ & 2 & 3 & 3 & 3 & = 11 \; d.o.f. \end{array} \tag{9}$$

The F-matrix is thus related to, 2 parameters for the projection of the translation $\mathbf{s} = \mathbf{A}'\,\mathbf{t}$ (defined up to a scale factor), 3 parameters for the rotation and twice 3 parameters for the intrinsic calibration parameters. It is thus clear that we can not recover all these parameters from the 7 parameters of the F-matrix.

Parameterization of the F-matrix We can derive a parameterization of $\mathbf{F}$ as follows. We write $\mathbf{F}$ in the form :

$$\mathbf{F} = \tilde{\mathrm{s}} \underbrace{\begin{pmatrix} a & b & 0 \\ c & d & 0 \\ 0 & 0 & 0 \end{pmatrix}}_{\mathbf{L}} \tilde{\mathrm{s}'} \tag{10}$$

with $||\mathrm{s}||^2 = ||\mathrm{s}'||^2 = a^2 + b^2 + c^2 + d^2 = 1$, while s' is defined as the basis vector of the kernel of $\mathbf{F}$.

This parameterization is singular in the case where $s^2 = 0$ or $s'^2 = 0$. However, this case corresponds to a retinal translation or a configuration where the epipoles are at infinity, which is of practical interest as discussed in this paper, and thus analyzed using a specific parameterization. Else we can write $\mathrm{s} = (s_0, s_1, 1)$ and $\mathrm{s}' = (s'_0, s'_1, 1)$ to obtain a parameterization of the epipole and similarly, distinguish if either a, b, c or d can be set to 0 or 1. In the case of a retinal translation i.e. $s_2 = 0$ we can use the decomposition of F with a permutation on the lines and rows of L.

This parameterization allows to directly estimate the epipoles s and s′. Each unary vector is defined by 2 parameters, while the matrix $\mathbf{L}$ is defined by 3 parameters. The elements of the matrix $\mathbf{L}$ have a geometric interpretation, since they define the homographic relation between the two pencils of epipolar lines [4].

As a consequence, with this parameterization, even if s and $\mathbf{F}$ are defined up to a scale factor only, the relative scale between these two quantities can be fixed, when using this definition.

Moreover, given an estimation $\bar{\mathbf{F}}$ of F-matrix for which we might have $det(\bar{\mathbf{F}}) \neq 0$ and its singular value decomposition $\bar{\mathbf{F}} = \mathbf{U} \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{pmatrix} \mathbf{V}$ we can easily estimate an "optimal" matrix $\mathbf{F}$ such that $det(\mathbf{F}) = 0$ and the two related epipole[4]. As a consequence, we always can estimate an unbiased value of the F-matrix from an estimate $\bar{\mathbf{F}}$.

Self-calibration from the F-matrix Let us now consider the problem of recovering the intrinsic calibration parameters.

[4] Let us write :

$$\bar{\mathbf{F}} = \left[\underbrace{(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3)}_{\mathbf{U}}\right] \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{pmatrix} \left[\underbrace{(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)}_{\mathbf{V}}\right]^T \tag{11}$$

with $\mathbf{U} \cdot \mathbf{U}^T = \mathbf{U}^T \cdot \mathbf{U} = \mathbf{I}$, $\mathbf{V} \cdot \mathbf{V}^T = \mathbf{V}^T \cdot \mathbf{V} = \mathbf{I}$ and $\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq 0$. This decomposition, called *singular value decomposition*, is unique and can be efficiently estimated numerically [5]. The reader can easily check that the matrix $\mathbf{F} = \mathbf{U} \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & 0 \end{pmatrix} \mathbf{V}^T$ verifies $det(\mathbf{F}) = 0$ and minimizes $||\bar{\mathbf{F}} - \mathbf{F}||$ considering the norm $||\mathbf{F}|| = sup_{||\mathbf{u}||=1} ||\mathbf{F}\,\mathbf{u}||$. Moreover $\mathbf{u}_3$ is the vector generating the kernel of $\mathbf{F}^{*T}$, i.e. $\mathbf{u}_3^T\,\mathbf{F} = 0$ and $\mathbf{v}_3$ is the vector generating the kernel of $\mathbf{F}$, i.e. $\mathbf{F} \cdot \mathbf{v}_3 = 0$

We write $\mathbf{K} = \mathbf{A}\mathbf{A}^T$ and $\mathbf{K}' = \mathbf{A}'\mathbf{A}'^T$. Each K-matrix is in one-to-one correspondence with the A-matrix[5] [11]. If we explicit the fact that $\mathbf{R}$ is an orthogonal matrix in the definition $\mathbf{H}_\infty$ of equation (4), we obtain:

$$\mathbf{K}' \equiv \mathbf{H}_\infty \, \mathbf{K} \, \mathbf{H}_\infty^T \tag{13}$$

which leads to (multiplying left and right by $\tilde{\mathbf{s}}$):

$$\tilde{\mathbf{s}} \, \mathbf{K}' \, \tilde{\mathbf{s}} \equiv \mathbf{F} \, \mathbf{K} \, \mathbf{F}^T \tag{14}$$

which corresponds to *two independent* equations known as the Kruppa equations. These equations can be easily made explicit considering the singular-value decomposition of $\mathbf{F}$ given in equation (11). After some algebra, equation (13) reduces to:

$$\begin{bmatrix} \mathbf{u}_1^T \, \mathbf{K}' \mathbf{u}_1 \\ \mathbf{u}_1^T \, \mathbf{K}' \mathbf{u}_2 \\ \mathbf{u}_2^T \, \mathbf{K}' \mathbf{u}_2 \end{bmatrix} \wedge \begin{bmatrix} (\sigma_1)^2 & \mathbf{v}_1^T \, \mathbf{K} \mathbf{v}_1 \\ \sigma_1 \, \sigma_2 & \mathbf{v}_1^T \, \mathbf{K} \mathbf{v}_2 \\ (\sigma_2)^2 & \mathbf{v}_2^T \, \mathbf{K} \mathbf{v}_2 \end{bmatrix} = 0 \tag{15}$$

which, as a cross-product, is equivalent to 2 equations. They yield quartic equations in the intrinsic parameters not easily calculable [4] which are the only equations we can obtain for self-calibration in the general case.

These equations have an important negative consequence : *in the general case (non-constant calibration parameters and only projective data) it is not possible to self-calibrate a system*, even with our simple model of equation (1). We thus must introduce another assumption, such as the fact that these parameters are constant [4] or discuss the calibration using another mechanisms, as developed in this paper.

3.2 General planar rigid displacement.

Let us assume that all points belongs to a 3D plane $\mathcal{P}$, with: $M \in \mathcal{P} \Leftrightarrow \mathbf{n}^T \, M = d$ where $\mathbf{n}$, $||\mathbf{n}|| = 1$, is the plane normal and $d > 0$ its distance to the origin.

If we now consider a general planar rigid displacement for which we have the following decomposition [11]:

$$\begin{array}{rccccccc} \mathbf{H} = & \mathbf{A}' & \mathbf{R} & \mathbf{A}^{-1} & + & \mathbf{s} & \nu^T & \\ & 3 & 3 & 3 & & 2 & 3 & = 14 \; d.o.f. \end{array} \tag{16}$$

we have 2 parameters for the projection of the translation $\mathbf{s}$, 3 parameters for the rotation and twice 3 parameters for the intrinsic calibration parameters, plus 3 parameters for the planar structure represented by $\nu = \mathbf{A}^{-1}\mathbf{n}/d$, with $\nu^T m = 1$. It is also clear that we cannot recover all these parameters in the general case.

[5] In our case we have:

$$\left\{ \begin{array}{ccc} u_0 = K_{02} & v_0 = K_{12} & f^2 = K_{00} - K_{02}^2 = K_{11} - K_{12}^2 \\ with & K_{22} = 1 & K_{01} = K_{02} K_{12} \end{array} \right. \tag{12}$$

so that the six elements of the *symmetric* matrix $\mathbf{K}$ verifies 2 quadratic constraints, 1 linear constraint and allow to estimate the parameters of the corresponding A-matrix.

Furthermore, even if either (i) the calibration parameters remain constant ($\mathbf{A}' = \mathbf{A}$) [11], or (ii) there is no rotation ($\mathbf{R} = \mathbf{I}$) with calibration parameters variations we can not recover these parameters. However, several more interesting situations can occur as detailed now.

The parameterization of the H-matrix is very simply $H = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & 1 \end{pmatrix}$ since for realistic displacements H_{22} is never null [11].

3.3 Elementary displacements of the camera: no translation.

Let us consider that we have a displacement of the camera with no translation, i.e. only a rotation and a variation of the intrinsic parameters. In that case,as discussed previously, we have:

$$\mathbf{F} = \mathrm{s} = 0 \quad and \quad \mathbf{H}_\infty \equiv \mathbf{H}_\bullet \tag{17}$$

This situation is thus characterized by the fact that all the retinal disparity can be predicted by a unique collineation. As discussed before, this situation
cannot distinguish from the fact that all points belong to a unique plane, unless an additional information is added.

In order to be sure to be in the case of a pure rotation we must:

- either get the information from an external source, such as the robotic system itself,
- or assume that the scene is made of at least two planes, so that if we see a "unique" plane, it corresponds to the fact we have a pure rotation.

In any case, the main property is that the correspondences between the two images does not depend on the depth of objects, so that the system can relate the visual information before and after displacement, as given in equation (4).

Moreover, if we are in the case of a pure rotation using equation (13) we can estimate the new intrinsic calibration parameters from the previous one [11].

Furthermore, if this motion is very quick, the additional variations in the image due to non-stationary objects between the beginning and the end of the motion are expected to be small.

Considering active vision, this yields to the following *saccadic* mechanism : *rotate the camera as quick as possible such that the angular position of the target is aligned with the optical axis.*

3.4 Elementary displacements of the camera: pure rotation.

In the case where the calibration parameters are constant, and knowing $\mathbf{H}_\bullet$ we can easily calculate the intrinsic calibration parameters since we have:

$$\mathbf{Q} = \frac{\mathbf{H}_\bullet}{det(\mathbf{H}_\bullet)} \quad and \quad \rho = \frac{\mathbf{A}\mathbf{u}}{det(\mathbf{A})} \quad \Rightarrow \mathbf{O} = \frac{\mathbf{Q} - \mathbf{Q}^{-1}}{2} = \sin(\theta)\ \mathbf{K}\ \tilde{\rho} \tag{18}$$

where $\mathbf{K} = \mathbf{A}\mathbf{A}^T$, $\mathbf{u}$ is the unary vector of the axis of the rotation and θ the angle of the rotation. In the general case we have $det(\mathbf{O}) = trace(\mathbf{O}) = 0$ and we can recover the intrinsic parameters only after two rotations not around the same axis [11].

We also recover ρ up to a scale factor, as the null vector of $\mathbf{O}$ since we have : $\mathbf{O}\,\rho = 0$.

With the model of equation (1) we obtain the following equations:

$$\left\{\begin{array}{rcl} O_{00} - u_0\,O_{20} & = & 0 \\ O_{11} - v_0\,O_{21} & = & 0 \\ (O_{01} + O_{10}) - u_0\,O_{21} - v_0\,O_{20} & = & 0 \\ O_{20}\,f^2 + O_{00}\,u_0 + O_{01}\,v_0 + O_{02} & = & 0 \\ O_{21}\,f^2 + O_{10}\,u_0 + O_{11}\,v_0 + O_{12} & = & 0 \end{array}\right. \tag{19}$$

which allow[6] to recover (u_0, v_0) from the 3 first linear equations and, knowing (u_0, v_0), f except if:

$$O_{20} = -\frac{sin(\theta)}{f}u_y = 0 \quad and \quad O_{21} = \frac{sin(\theta)}{f}u_x = 0 \tag{20}$$

This singular configuration corresponds to a rotation around the optical axis only which will be identified in the sequel.

As a consequence, *performing a pure rotation allows to directly calibrate the system, with our simple model.* This is not the case with another calibration model [11].

3.5 Elementary displacements of the camera: retinal translation.

This other situation is simply characterized by the fact that the translation is parallel to the image plane, i.e. $t^2 = 0$. This is equivalent to:

$$s^2 = 0 \tag{21}$$

since from equation (1), $t^2 \equiv s^2$

In that case, the orientation θ of the translation is $\theta = arctan(s^1/s^0)$. This is due to that fact that from equation (1), $s^0 = \lambda(f\ t^0 + u_0\ t^2)$ and $s^1 = \lambda(f\ t^1 + v_0\ t^2)$ so that if $t^2 = 0$, we have $s^1/s^0 = t^1/t^0$.

If we are not in this situation, we can parameterize the s-vector using s^0 and s^1 and assuming that $s^2 = 0$ whereas if we are not in this situation, we can parameterize the s-vector using s^0 and s^1 and assuming that $s^2 = 1$.

[6] The explicit equations are:

$$\left\{\begin{array}{rcl} u0 & = & \frac{{O_{20}}^3 O_{00} + (O_{00} - O_{11}) {O_{21}}^2 O_{20} + (O_{01} + O_{10}) {O_{21}}^3}{{O_{20}}^2 {O_{21}}^2 + {O_{20}}^4 + {O_{21}}^4} \\ v0 & = & -\frac{(-O_{01} - O_{10}) {O_{20}}^3 + (O_{00} - O_{11}) O_{21}\ {O_{20}}^2 - {O_{21}}^3 O_{11}}{{O_{20}}^2 {O_{21}}^2 + {O_{20}}^4 + {O_{21}}^4} \\ f^2 & = & -\frac{(O_{21}\ O_{11} + O_{20}\ O_{01}) v0}{{O_{20}}^2 + {O_{21}}^2} - \frac{(O_{20}\ O_{00} + O_{21}\ O_{10}) u0}{{O_{20}}^2 + {O_{21}}^2} - \frac{O_{21}\ O_{12} + O_{20}\ O_{02}}{{O_{20}}^2 + {O_{21}}^2} \end{array}\right.$$

3.6 Elementary displacements of the camera: pure translation.

Let us now consider that we have a displacement of the camera which is a pure translation, i.e. there is no rotation and no variation of the intrinsic parameters. In that case, from equations (4), (5):

$$\mathbf{F} = \tilde{\mathbf{s}} \quad and \quad \mathbf{H}_\infty \equiv \mathbf{I} \tag{22}$$

This situation is simply characterized by the fact that the F-matrix is skew-symmetric.

Considering the reconstruction problem, we see from equation (4) that some translational motion is required to infer structure from motion and to relate the 3D structure parameters to the projected displacement.

Looking for optimal translational displacement, it is already known that the orientation of the translation is better "orthogonal to the 2D points" [10], i.e. oriented such that the induced disparity is maximal : considering a line segment, for instance the projection of the translation is optimal if it induces a normal displacement of the rectilinear edge.

Moreover if can we recover the direction of the translation, as soon as a calibration is issued :

$$\mathbf{t} \equiv \mathbf{A}^{-1}\,\mathbf{s} \tag{23}$$

The parameterizations of F are very simply $F = \begin{pmatrix} 0 & 1 & s_1 \\ -1 & 0 & -s_0 \\ -s_1 & s_0 & 0 \end{pmatrix}$ for finite epipole with similar parameterization if $s_2 = 0$.

3.7 Elementary displacements of the camera: pure planar translation.

Let us now consider that we have a pure translation, but for a planar structure. Using the same equations we easily see that we are looking for a matrix $\mathbf{H}$ of the form:

$$\mathbf{H} \equiv \mathbf{I} + \frac{\mathbf{s}}{||\mathbf{s}||}\,\nu^T \tag{24}$$

Considering $\mathbf{H}^T$ we have the following eigen-value decomposition :

$$\left\{ \begin{array}{rcl} \mathbf{u} \perp \mathbf{s} & \Rightarrow & \mathbf{H}^T\,\mathbf{u} = \mathbf{u} \\ \mathbf{u} \equiv \nu & \Rightarrow & \mathbf{H}^T\,\mathbf{u} = (1 + \nu^T\,\frac{\mathbf{s}}{||\mathbf{s}||})\,\mathbf{u} \end{array} \right. \tag{25}$$

Such a collineation is only defined by five parameters (2 for the projection of the translation defined up to a scale factor and 3 for the parameter of the plane) and can thus be characterized as soon as three generic points are given, since each match generates two equations given by: $\mathbf{m}' \wedge \mathbf{H}\,\mathbf{m} = 0$, or more precisely:

$$\mathbf{m}' \wedge \mathbf{m} + (\nu^T\,\mathbf{m})\left[\mathbf{m}' \wedge \frac{\mathbf{s}}{||\mathbf{s}||}\right] = 0 \Leftrightarrow \left\{ \begin{array}{c} |\mathbf{m}, \mathbf{m}', \frac{\mathbf{s}}{||\mathbf{s}||}| = 0 \\ (\nu^T\,\mathbf{m}) = \frac{\mathbf{s}}{||\mathbf{s}||}^T\,\frac{(\mathbf{m}'^T\mathbf{m}')\mathbf{m} - (\mathbf{m}'^T\mathbf{m})\mathbf{m}'}{||\mathbf{m}' \wedge \mathbf{m}||^2} \end{array} \right. \tag{26}$$

so that given at least two non-stationary points we can calculate $\frac{\mathbf{s}}{||\mathbf{s}||}$ and then, given at least three non collinear points we can calculate ν knowing $\frac{\mathbf{s}}{||\mathbf{s}||}$.

Reciprocally, *the corresponding H-matrix is characterized by the fact that it has two eigen-values which are equal.*

In order to clarify this last point consider any matrix $\mathbf{H}$ with $\mathbf{u}_{1a}$ and $\mathbf{u}_{1b}$ the eigen-vectors corresponding to the same eigen-value λ_1 of $\mathbf{H}^T$ and $\mathbf{u}_0$ the eigen-vector corresponding to another eigen-value λ_0 of $\mathbf{H}^T$. Inverting equation (25) we easily obtain :

$$\mathbf{s} \equiv \mathbf{u}_{1a} \wedge \mathbf{u}_{1b} \quad ; \quad \nu = \frac{\lambda_0/\lambda_1 - 1}{\frac{\mathbf{s}}{||\mathbf{s}||}^T \mathbf{u}_0} \mathbf{u}_0 \tag{27}$$

and verify that $\mathbf{H}$ is of the form of equation (24). Therefore $\mathbf{H}$ *is of the form of equation (24) if and only if it has two eigen-values which are equal.*

It is thus possible to identify if a given H-matrix corresponds to a pure translation (testing if two eigen-values are equal) and if true, to recover the elements of the collineation (see [1] for a review of the calibrated case).

3.8 Elementary displacements of the camera: retinal displacement.

Let us consider the particular type of displacement for which the retinal plane is invariant. Clearly, this corresponds to rigid motion with : $t^2 = u^0 = u^1 = 0$, i.e. we have a rotation around the z axis only and of angle θ and a translation with $t^2 = 0$, while the calibration parameters might be constant or not.

In that case, we have a F-matrix of the form:

$$\mathbf{F} = \begin{pmatrix} 0 & 0 & \lambda t^1 \\ 0 & 0 & -\lambda t^0 \\ \frac{f'}{f}(-F^{02}\cos(\theta) - F^{12}\sin(\theta)) & \frac{f'}{f}(F^{02}\sin(\theta) - F^{12}\cos(\theta)) & F^{22} \end{pmatrix} \tag{28}$$

$$with \qquad F^{22} = -(F^{02}u'_0 + F^{20}u_0 + F^{12}v'_0 + F^{21}v_0)$$

This F-matrix is constrained by:

$$F^{00} = F^{01} = F^{10} = F^{11} = 0 \tag{29}$$

and we observed, from equation (28), that:

$$\frac{f'}{f} = \sqrt{\frac{(F^{20})^2 + (F^{21})^2}{(F^{02})^2 + (F^{12})^2}} \tag{30}$$

so that we can estimate the variation of the focal length, while we recover, from equation (28), the angle of the rotation as:

$$\theta = \arctan(\frac{F^{21}}{F^{20}}) - \arctan(\frac{F^{12}}{F^{02}}) \tag{31}$$

We also recover the translation up to a scale factor $\mathbf{t} \equiv \mathbf{s} \equiv (F^{02}, F^{12}, 0)^T$, and obtain a linear equation for the location of the principal point, using F^{22}.

On the reverse, considering any general matrix F, while the calibration parameters are either constant or varying, the displacement is a displacement for which the retinal plane is invariant, if and only if $F^{00} = F^{01} = F^{10} = F^{11} = 0$ since if we define $\mathbf{v} = 2 \tan\left(\frac{\theta}{2}\right) \frac{\mathbf{u}}{\|\mathbf{u}\|}$ we have:

$$\begin{pmatrix} F^{00} \\ F^{01} \\ F^{10} \\ F^{11} \end{pmatrix} \equiv \begin{pmatrix} (2\,v^0\,v^2 - 4\,v^1)\,t^1 - (4\,v^2 + 2\,v^0\,v^1)\,t^2 \\ (2\,v^1\,v^2 + 4\,v^0)\,t^1 + (v^0\,v^0 + v^2\,v^2 - v^1\,v^1 - 4)\,t^2 \\ (-2\,v^0\,v^2 + 4\,v^1)\,t^0 + (v^0\,v^0 - v^2\,v^2 - v^1\,v^1 + 4)\,t^2 \\ (-2\,v^1\,v^2 - 4\,v^0)\,t^0 + (-4\,v^2 + 2\,v^0\,v^1)\,t^2 \end{pmatrix} \tag{32}$$

which real solutions are only either $t^0 = t^1 = t^2 = 0$ or $v^0 = v^1 = t^2 = 0$, as easily verified with a symbolic calculator.

Therefore a necessary and sufficient condition to be in this situation is, for the F-matrix, to be of the form $F = \begin{pmatrix} 0 & 0 & F^{02} \\ 0 & 0 & F^{12} \\ F^{20} & F^{21} & F^{22} \end{pmatrix}$.

If now, we assume that we have a knowledge of the plane at infinity, the related matrix $\mathbf{H}_\infty$ is given by:

$$\mathbf{H}_\infty \equiv \begin{pmatrix} \frac{f'}{f}\cos(\theta) & -\frac{f'}{f}\sin(\theta) & \frac{f'}{f}\left(-\cos(\theta)\,u_0 + \sin(\theta)\,v_0\right) + u'_0 \\ \frac{f'}{f}\sin(\theta) & \frac{f'}{f}\cos(\theta) & \frac{f'}{f}\left(-\cos(\theta)\,v_0 - \sin(\theta)\,v_0\right) + v'_0 \\ 0 & 0 & 1 \end{pmatrix} \tag{33}$$

In this case, we also directly obtain the rotation angle θ, the variation of the focal length f'/f, and can estimate the new location of the optical center (u'_0, v'_0) in function of the previous value (u_0, v_0), as visible in equation (33). If the calibration parameters are constant, we directly obtain the location of the principal point. In any case, no information on the absolute value of the focal length is obtained.

This corresponds to a *a retinal planar rotation.* The particular H-matrix to be estimated is of the form: $H = \begin{pmatrix} a & b & c \\ -b & a & d \\ 0 & 0 & 1 \end{pmatrix}$.

As far as reconstruction is concerned, it has been demonstrated that we can reconstruct the scene as if it was provided by an orthographic camera [3].

Considering active vision, this yields the following mechanism : *induce a translational motion in a direction orthogonal to the average target location in the image, and in the parallel to the retinal plane, and compensate for this disparity using rotational stabilization.*

As discussed elsewhere [10], if the translation is performed such that its component along the Z-axis is zero, the retinal disparity is not dependent upon the location of the principal point [1]. As a consequence, even if this quantity is not known with a high precision (approximate calibration), the result will not be affected. Moreover, the retinal field has no singularity in the image which simplifies its estimation. It is thus a good strategy to try to perform translation parallel to the retinal plane in this case.

3.9 Elementary displacements of the camera: retinal planar displacement.

Let us now consider the fact that we are observing a planar structure, under an retinal displacement, as discussed previously.

In this case we do not have to estimate the eight parameters of the collineation, but only six of them since the matrix $\mathbf{H}$ is of the form:

$$\mathbf{H} = \begin{pmatrix} a & b & d_u \\ c & d & d_v \\ 0 & 0 & 1 \end{pmatrix} \tag{34}$$

i.e. we can assume that $\mathbf{h}^2 \equiv (0,0,1)^T$ which leads to a retinal displacement.

This fact can be easily verified with the model of equation (1) since we have:

$$\begin{array}{rcl} H^{20} & \equiv & u^0 u^2 (1-\cos(\theta)) - u^1 \sin(\theta) + t^2 n^0 \\ H^{21} & \equiv & u^1 u^2 (1-\cos(\theta)) + u^0 \sin(\theta) + t^2 n^1 \end{array} \tag{35}$$

whatever the calibration parameters are.

Solving these equations we easily demonstrate that these two values vanish if and only if *there is no rotation except around the optical axis (i.e.* $\mathbf{u} \equiv (0,0,1)^T$*) and either the translation is parallel to the retinal plane (i.e* $s^2 \equiv t^2 = 0$*) or the plane is a fronto-parallel plane (i.e. the normal of the plane* $\mathbf{n} \equiv (0,0,1)^T$*).*

In the general case of variable calibration parameters, it is not possible to recover the Euclidean parameters, since we still have 11 parameters to estimate. However, if the calibration parameters are constant, as given in [3], we can recover the direction of the translation, the angle of the rotation and the plane normal up to an indetermination. In [3] the equations have been given for small displacements and they can be easily generalized as given now:

$$\left\{ \begin{array}{rcl} \theta & = & 2\, arctan(Z) \quad with \\ & & (1+H^{00}+H^{11}+H^{11}H^{00}-H^{01}H^{10})Z^2 + 2(H^{01}-H^{10})Z + (1-H^{00}-H^{11}+H^{11}H^{00}-H^{01}H^{10}) = 0 \\ t^0 & = & \lambda\ \cos(\alpha) \\ t^1 & = & \lambda\ \sin(\alpha) \\ & & \alpha = \frac{1}{2} arctan\left(2\ \frac{(H^{11}H^{01}+H^{00}H^{10})-\cos(\theta)(H^{01}+H^{10})+\sin(\theta)(H^{11}-H^{00})}{((H^{00})^2+(H^{01})^2-(H^{10})^2-(H^{11})^2)-2(\cos(\theta)(H^{11}-H^{00})+\sin(\theta)(H^{10}+H^{01}))}\right) \\ n^0 & = & \frac{t^0(H^{00}-H^{11})+t^1(H^{01}-H^{10})}{(t^0)^2+(t^1)^2} \\ n^1 & = & \frac{t^1(H^{11}-H^{00})+t^0(H^{01}-H^{10})}{(t^0)^2+(t^1)^2} \end{array} \right. \tag{36}$$

We have two solutions for the displacement, as in the continuous case. The interpretation is discussed in [3]. Furthermore, the translation is also given up to an indeterminate λ, as expected.

Furthermore, we have one additional linear equation for the intrinsic parameters (u_0, v_0):

$$\sin(\alpha)\ \left(H^{02} - u_0\,(1-H^{00}) + v_0\,H^{01}\right) = \cos(\alpha)\ \left(H^{12} + u_0\,H^{10} - v_0\,(1-H^{11})\right) \tag{37}$$

as in the continuous case again. We do not have any constraint on the focal length f, since this quantity always appears in the equation as a factor of the indeterminate λ.

An even more degenerated form of the retinal planar displacement is to a consider a *constant retinal motion*, i.e. $a = d = 1$ and $c = b = 0$ in equation (34). This extreme situation is sometimes used for panoramic displacements. It can be easily demonstrated that this situation corresponds to *a pure retinal translation of a fronto-parallel plane, with constant calibration parameters*, with $\mathbf{s} \equiv (d_u, d_v, 0)$.

3.10 Elementary displacements of the camera: fixed axis rotation.

Let us now consider that the elementary displacement of the camera is a fixed axis rotation, while intrinsic parameters are constant. It is known that this situation is characterized by the fact that the translation is perpendicular to the axis of rotation. We have:

$$det(\mathbf{F} + \mathbf{F}^T) = 0 \tag{38}$$

It has been demonstrated that this condition is necessary and sufficient for the F-matrix to correspond to such a displacement (see [3]).

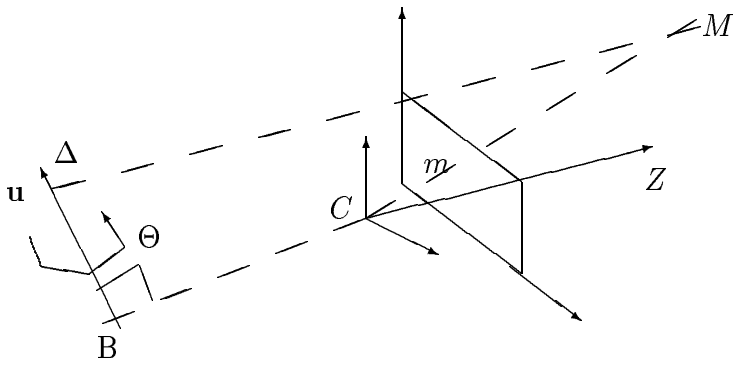


Figure 3: Notations for a fixed axis rotation

We use the notations of figure 3. We consider a vector $\mathbf{u}$ aligned with the rotation axis Δ and the point B on the axis Δ, such that $\mathbf{C} = \vec{CB} \perp \mathbf{u}$. The rotation angle is θ.

From equation (1), (4) and (5) in this case, the F-matrix is of the rather heavy form:

$$\begin{array}{c} \mathbf{F} \equiv ||\mathbf{C}|| \left[\sin(\theta)\ \tilde{\mathbf{f}}_0 + (1 - \cos(\theta))\ \left[\mathbf{f}_1\ \mathbf{f}_2^T + \mathbf{f}_2\ \mathbf{f}_1^T\right]\right] \\ \mathbf{f}_0 = \lambda^2\ \frac{2}{det(A)}\ \mathbf{A}\ \frac{\mathbf{C}\wedge\mathbf{u}}{||\mathbf{C}||} \quad and \quad \mathbf{f}_1 = \lambda\ \mathbf{A}^{-1\,T}\ \mathbf{u} \quad and \quad \mathbf{f}_2 = \lambda\ \mathbf{A}^{-1\,T}\ \frac{\mathbf{C}\wedge\mathbf{u}}{||\mathbf{C}||} \end{array} \tag{39}$$

while we have:

$$\mathbf{f}_0^T\ \mathbf{f}_1 = 0 \tag{40}$$

which is the only constraint between these three vectors[7].

[7] Consider three generic vectors $\mathbf{f}_0$, $\mathbf{f}_1$ and $\mathbf{f}_2$ such that $\mathbf{f}_0 \perp \mathbf{f}_1$ and introduce a vector $\mathbf{b} = \gamma \frac{\mathbf{f}_1 \wedge \mathbf{f}_0}{\mathbf{f}_1^T\ \mathbf{f}_1} - (\frac{1}{2} + \gamma)\ \frac{\mathbf{f}_2 \wedge \mathbf{f}_0}{\mathbf{f}_1^T\ \mathbf{f}_2}$ so that $\mathbf{b}^T\ \mathbf{f}_0 = 0$ and $\mathbf{b}^T\ (\mathbf{f}_1 \wedge \mathbf{f}_2) = \frac{(\mathbf{f}_0^T\ \mathbf{f}_2)}{2}$ for any γ. There is always a solution of equation (39)

Knowing the rotation angle θ and the matrix $\mathbf{F}$ up to a scale factor, it is then straightforward to recover $\mathbf{f}_0$ which is the vector associated to the skew-symmetric part of $\mathbf{F}$.

Similarly $\mathbf{u}_1 = \left(\frac{\mathbf{f}_1}{||\mathbf{f}_1||} + \frac{\mathbf{f}_2}{||\mathbf{f}_2||}\right)$ and $\mathbf{u}_2 = \left(\frac{\mathbf{f}_1}{||\mathbf{f}_1||} - \frac{\mathbf{f}_2}{||\mathbf{f}_2||}\right)$ are the eigen-vectors associated respectively with the positive eigen-value and negative eigen-value of the symmetric part of $\mathbf{F}$, the third eigen-value being zero as induced by equation (38). We then obtain $\mathbf{f}_1 \equiv \mathbf{u}_1 + \mathbf{u}_2$ and $\mathbf{f}_2 \equiv \mathbf{u}_1 - \mathbf{u}_2$, but up to a scale factor only.

We can recover the intrinsic calibration parameters:

$$\left\{ \begin{array}{rcl} u_0 & = & f_0^0 - \gamma\ f_2^0/f_0^2 \\ v_0 & = & f_0^1 - \gamma\ f_2^1/f_0^2 \\ f & = & \left[\sqrt{\gamma\ (f_0^0\ f_2^0 + f_0^1\ f_2^1 + f_0^2\ f_2^2) - \gamma^2\ ((f_2^0)^2 + (f_2^1)^2)}\right]/f_0^2 \end{array} \right. \tag{41}$$

up to a 1D-indetermination related to γ as found with another formalism.

A step further, we obtain two equations for the intrinsic parameters:

$$\left\{ \begin{array}{rcl} f_2^1\ (f_0^2\ u_0 - f_0^0) - f_2^0\ (f_0^2\ u_0 - f_0^1) & = & 0 \\ f^2 + u_0^2 + v_0^2 - (f_0^0\ u_0 + f_0^1\ v_0 + \gamma\ f_2^2)/f_0^2 & = & 0 \end{array} \right. \tag{42}$$

with $\gamma = [f_2^0\ (f_0^2\ u_0 - f_0^0) + f_2^1\ (f_0^2\ u_0 - f_0^1)]/[(f_2^0)^2 + (f_2^1)^2]$, so that we have one linear equation in u_0 and v_0 and direct estimation of f^2 as soon as u_0 and v_0 are estimated.

In addition, we recover the extrinsic calibration parameters:

$$\mathbf{u} \equiv \mathbf{A}^T\ \mathbf{f}_1 \quad and \quad \mathbf{C} \equiv \mathbf{u} \wedge \mathbf{A}^{-1}\ \mathbf{f}_0 \tag{43}$$

except $||\mathbf{C}|| = ||\mathbf{t}||/\sqrt{2\,(1 - \cos(\theta))}$ which cannot be recovered since we use a monocular system with a scale factor indetermination.

All these equations are singular if and only if:

$$f_0^2 = \frac{2}{f^2}\left[C^0 u^1 - C^1 u^0\right] = 0 \tag{44}$$

This singular configuration corresponds to the fact that (i) the intersection of the optical axis with retinal plane, (ii) the projection of $\mathbf{C}$ and (iii) the principal point are collinear, which can be detected in practice.

Fixed axis rotation does not induce a particular form for any H-matrix, as the reader can easily verify.

3.11 Elementary displacements of the camera: zoom displacement.

A real zoom does not correspond to a simple variation of the intrinsic parameters since the position of the optical center varies, and a translation occurs[8]. Let us now model this effect.

---

for:

$$\lambda = 1 \quad \mathbf{u} = \begin{pmatrix} 0 & 1 & 0 \end{pmatrix} \quad \frac{\mathbf{C} \wedge \mathbf{u}}{||\mathbf{C}||} = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix} \quad A^{-1\,T} = (\mathbf{b}, \mathbf{f}_1, \mathbf{f}_2)$$

as the reader can easily verify.

[8] We consider that a zoom is the combination of *a variation of the intrinsic parameters, plus a general translation, but without any rotation.* We must assume that this translation is general, and in particular not

From equation (1), (2) and (5) we have a F-matrix of the form :

$$\begin{array}{rcl} \mathbf{F} & = & \begin{pmatrix} 0 & -\lambda\, t^2 & \lambda\, f\, t^1 - F^{01}\, v_0 \\ \lambda\, t^2 & 0 & -\lambda\, f\, t^0 + F^{01}\, u_0 \\ -\lambda\, f'\, t^1 + F^{01}\, v'_0 & \lambda\, f'\, t^0 - F^{01}\, u'_0 & F^{22} \end{pmatrix} \\ F^{22} & = & (u_0\, v'_0 - v_0\, u'_0)\, F^{01} - u_0\, F^{20} - u'_0\, F^{02} - v_0\, F^{21} - v'_0\, F^{12} \end{array} \tag{45}$$

which is thus constrained by:

$$F^{00} = 0, F^{11} = 0, F^{01} = -F^{10} \tag{46}$$

Reciprocally, using equation (32) we can easily verify that this is a sufficient condition.

Here, we have 4 independent parameters (since a general F-matrix has 7 parameters and here 3 constraints occur). More precisely, knowing (u_0, v_0, f) we can recover, if $F^{01} = -\lambda\, t^2 \neq 0$, which is expected for a zoom:

$$\mathbf{t} \equiv \begin{pmatrix} F^{12} - F^{01}\, u_0 \\ -F^{02} - F^{01}\, v_0 \\ f\, F^{01} \end{pmatrix} \quad and \quad \begin{cases} u'_0 & = & \frac{f'}{f} u_0 - \frac{\frac{f'}{f} F^{12} + F^{21}}{F^{01}} \\ v'_0 & = & \frac{f'}{f} v_0 + \frac{\frac{f'}{f} F^{02} + F^{20}}{F^{01}} \end{cases} \tag{47}$$

i.e. $\mathbf{t}$ up to a scale factor and the new location of the principal point (u'_0, v'_0) while if $t^2 = 0$ we are in the case of an retinal displacement, as detailed before.

Here, we cannot recover f'/f because this expansion factor is an indetermination of the motion parameterization as discussed in [11] while equation (45) with $det(\mathbf{F}) = 0$ is equivalent to equation (47). This is somehow a very negative result, since the "zoom" effect is precisely a variation of f, non-measurable ! This is due to the fact that this is an affine quantity, as discussed in [11].

If we have a "perfect" zoom, i.e. $t^0 = t^1 = 0$, so that there is a pure translation in Z we obtain very simply: $u_0 = F^{12}/F^{01}, v_0 = F^{02}/F^{01}, u'_0 = F^{21}/F^{01}, v'_0 = F^{20}/F^{01}$ but no information on f and f'. The F-matrix has the same form as for a general zoom, unless $u_0 = u'_0$ and $v_0 = v'_0$ and in such a case the form is similar to the case of a pure translation.

If now, we consider that the affine calibration is given, we have:

$$\mathbf{H}_\infty \equiv \begin{pmatrix} \frac{f'}{f} & 0 & u'_0 - u_0 \frac{f'}{f} \\ 0 & \frac{f'}{f} & v'_0 - v_0 \frac{f'}{f} \\ 0 & 0 & 1 \end{pmatrix} \tag{48}$$

We are here in a situation where we recover, the new intrinsic calibration parameters from the old one, and the "zoom" effect since we have a knowledge of the affine calibration.

For a general planar structure, we have a H-matrix of the form:

$$\mathbf{H} \equiv \begin{pmatrix} \frac{f'}{f} + s^0\, n^0 & s^0\, n^1 & u'_0 - u_0 \frac{f'}{f} + s^0\, n^2 \\ s^1\, n^0 & \frac{f'}{f} + s^1\, n^1 & v'_0 - v_0 \frac{f'}{f} + s^1\, n^2 \\ s^2\, n^0 & s^2\, n^y & 1 + s^2\, n^2 \end{pmatrix} \tag{49}$$

related to the displacement of the principal point because, we are in the case of a thick lens and the focus mechanism induces variations of the latter without variations of the former

subject to the following constraint: $H^{01}H^{20}H^{20} - H^{00}H^{21}H^{20} - H^{10}H^{21}H^{21} + H^{11}H^{21}H^{20} = 0$.

It thus yields 7 independent equations which allow to compute: $\mathrm{s} \equiv (H^{01}H^{20}, H^{10}H^{21}, H^{20}H^{21})$, a constraint on the plane parameter: $n^0 H^{21} = n^1 H^{20}$ and extract two equations for the intrinsic parameters:

$$\begin{cases} u'_0 &= \frac{f'}{f}u_0 + \frac{H^{01}}{H^{21}} + \frac{f'}{f}\frac{H^{01}H^{22} - H^{02}H^{21}}{H^{01}H^{20} - H^{00}H^{21}} \\ v'_0 &= \frac{f'}{f}v_0 + \frac{H^{10}}{H^{20}} + \frac{f'}{f}\frac{H^{10}H^{22} - H^{12}H^{20}}{H^{10}H^{21} - H^{11}H^{20}} \end{cases}$$

as for a general rigid object, but again no information on f'/f.

This analysis allow us to conclude that -with our model- a zoom is not a very attractive displacement, when considering the self-calibration and reconstruction problem, because we loose the knowledge on the calibration parameters in this case. Therefore, as in [13] or in a simpler case as in [13], where a model of the variation of the intrinsic parameters is defined and calibrated is thus mandatory. Another alternative is to consider a thick lens model as in [13].

Let us finally analyze the zoom displacement qualitatively :

- To detect unexpected objects, the best conservative configuration for the zoom is to be minimum (the focal length being smallest, the field of view is wider). This extremal configuration corresponds also to a situation where object size and projected displacements are minimal.

- However, if the field of view is too wide then so will be the density of edges and an artificial disparity will be induced by matching errors. Zooming into the observed object will overcome insufficient resolution.

This leads to a general criterion for zoom control : *the focal length is to be increased if and only if this reduces the residual disparity between two frames for the observed object, and it is to be tuned to minimize this disparity.*

4 Defining a hierarchical motion module

4.1 Combining different models of displacements

Following the previous discussion, when we estimate a rigid displacement, we consider several cases, depending on the nature of the displacement. Collecting all constraints proposed in the previous section, we can describe the following set of models:

Considering a rigid structure, the following class of displacements can be identified:

Class of Displacement	Parameterization or constraint	Information Recovered	Number of Parameters
Pure rotation	$\mathbf{F}=0$	$\mathbf{H}_\infty, \mathbf{t}=0$	0
Z-axis pure translation	$\mathbf{F}=\begin{pmatrix}0&-1&0\\1&0&0\\0&0&0\end{pmatrix}$	$\mathbf{H}_\infty=\mathbf{R}=\mathbf{I}, \mathbf{t}\equiv\mathbf{z}$	0
Pure retinal translation	$\mathbf{F}=\begin{pmatrix}0&0&a\\0&0&b\\-a&-b&0\end{pmatrix}, \|\|\mathbf{F}\|\|=1$	$\mathbf{H}_\infty=\mathbf{R}=\mathbf{I}, \quad \mathbf{t}\equiv\begin{pmatrix}\cos(\theta)\\\sin(\theta)\\0\end{pmatrix}$ $\theta=atan(\frac{s_1}{s_0})$	1
Pure translation	$\mathbf{F}=\begin{pmatrix}0&c&a\\-c&0&b\\-a&-b&0\end{pmatrix}, \|\|\mathbf{F}\|\|=1$	$\mathbf{H}_\infty=\mathbf{R}=\mathbf{I}$	2
Retinal displacement	$\mathbf{F}=\begin{pmatrix}0&0&a\\0&0&b\\c&d&e\end{pmatrix}, \|\|\mathbf{F}\|\|=1$	$\mathbf{R}, \mathbf{t}/\|\|\mathbf{t}\|\|, eq(\mathbf{a})$	4
Zoom displacement	$\mathbf{F}=\begin{pmatrix}0&f&a\\-f&0&b\\c&d&e\end{pmatrix}, \quad \begin{matrix}c\,b-a\,d=0\\\|\|\mathbf{F}\|\|=1\end{matrix}$	$\mathbf{R}=\mathbf{I}, \mathbf{t}/\|\|\mathbf{t}\|\|, eq(\mathbf{a})$	4
Fixed axis rotation	$det(\mathbf{F}+\mathbf{F}^T)=0$, $det(\mathbf{F})=0$, $\|\|\mathbf{F}\|\|=1$	$eq(\mathbf{a})$	6
Retinal translation	$det(\mathbf{F})=0, s^2=0, \|\|\mathbf{F}\|\|=1$	$\mathbf{t}\equiv\begin{pmatrix}\cos(\theta)\\\sin(\theta)\\0\end{pmatrix}$ $\theta=atan(\frac{s_1}{s_0})$ $, eq(\mathbf{a})(Kruppa)$	6
General rigid displacement	$det(\mathbf{F})=0, \|\|\mathbf{F}\|\|=1$	$eq(\mathbf{a})(Kruppa)$	7

where $eq(\mathbf{a})$ means that we obtain equations about the intrinsic calibration parameters, these equations being either linear equations or the quartic Kruppa equations, as specified. In these cases, it is not possible to maintain an estimation of all calibration parameters.

In the planar case, we have:

Class of Displacement	Parameterization or constraint	Information Recovered	Number of Parameters
Stationary structure	$\mathbf{H}=\mathbf{I}$	$\mathbf{R}=\mathbf{I}, \mathbf{t}=0, \mathbf{a}=\mathbf{a}'$	0
Constant retinal displacement	$\mathbf{H}=\begin{pmatrix}0&0&a\\0&0&b\\0&0&1\end{pmatrix}$	$\mathbf{R}=I, \mathbf{t}\equiv(a,b,0), \mathbf{a}=\mathbf{a}', \mathbf{n}\equiv\mathbf{z}$	2
Retinal planar zoom	$\mathbf{H}=\begin{pmatrix}c&0&a\\0&c&b\\0&0&1\end{pmatrix}$	eq($\mathbf{a}$)	4
Retinal planar rotation	$\mathbf{H}=\begin{pmatrix}c&d&a\\-d&c&b\\0&0&1\end{pmatrix}$	$\mathbf{R}$, eq($\mathbf{a}$)	4
Pure planar retinal translation	$\mathbf{H}=\mathbf{I}+\mathbf{s}\nu^T, s^2=1$	$\mathbf{R}=\mathbf{I}, \mathbf{s}/\|\|\mathbf{s}\|\|, \nu$	5
Pure planar translation	$\mathbf{H}=\mathbf{I}+\mathbf{s}\nu^T, s^2=0$	$\mathbf{R}=\mathbf{I}, \mathbf{s}/\|\|\mathbf{s}\|\|, \nu$	5
Retinal planar displacement	$\mathbf{H}=\begin{pmatrix}a&b&c\\d&e&f\\0&0&1\end{pmatrix}$	$\mathbf{R}$, $\mathbf{t}/\|\|\mathbf{t}\|\|$, $\mathbf{n}$, eq($\mathbf{a}$)	6
General planar displacement	$\mathbf{H}=\begin{pmatrix}a&b&c\\d&e&f\\g&h&1\end{pmatrix}$		8

In fact some other variants have also been introduced in order to have alternative models with very few parameters. For instance a model with zero parameters, corresponding to a collineation equal to the identity, i.e. a stationary structure is introduced. This allows to have a simple model assuming that points are not moving.

Furthermore, this set of model has a very interesting structure, i.e. some models are generalizations of others. This allows to take as best model the first model, starting from the bottom, which statistical significance is smaller that every models immediately higher in the hierarchy.

This is illustrated in figure 4.

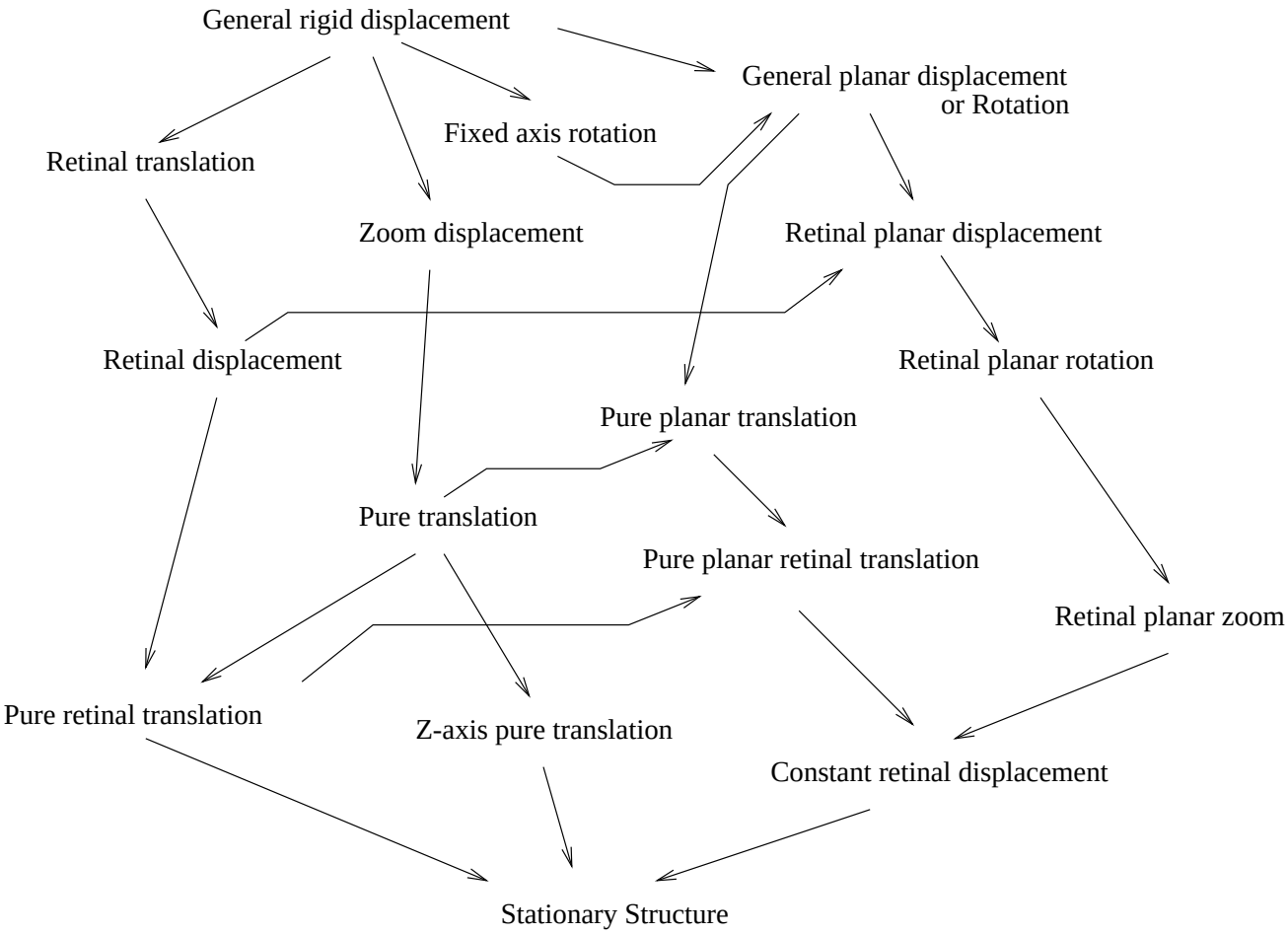


Figure 4: A hierarchy of models of rigid and planar displacements

4.2 Using a statistical framework.

4.2.1 The Early-vision module.

Feature points corresponding to high curvature points are extracted from each image. In our application, we use the "Harris" corner detector [2], and as reported in [15], we perform a correlation operation and select those locations for which the correlation score is high. We may find both bad locations and false matches, or matches which do not correspond to the quantity we want to estimate, because they belongs to moving objects.

When considering an image sequence, each point will be tracked from the first to the last view and then the trajectory will be interpolated using a polynomial model in order to obtain a subpixel mechanism of localization of the points.

4.2.2 Eliminating outliers.

In order to eliminate these outliers, and following previous authors in this field [7, 15] we use a variant of the least-median square algorithm: *the trimmed-least-median-square*. This method estimates the parameters by finding the smallest value for the trimmed-median[9] of the squared residuals computed for the entire data set. Please refer to [7, 15] for details.

It turns out that this method is very robust to false matches as well as outliers due to bad locations. The obtained estimate is refined, at last, solving a least-squares problem, as discussed now.

In our case, it is very easy to estimate the $\mathbf{x}$ parameter from a set of matches.

- When estimating H-matrices which minimize the retinal disparity we use the five proposed parameterizations and solve equation (7) for a minimal set of points.
- When estimating F-matrices which minimize the retinal disparity we use the seven proposed parameterizations and solve equation (5) for a minimal set of points.

4.2.3 Refining the estimation.

As noted in [6], the least-median-square *efficiency* is poor in the presence of Gaussian noise. The efficiency of a method is defined as the ratio between the lowest achievable variance for the estimated parameters and the actual variance provided by the given method. To compensate for this deficiency, we further carry out a weighted least-squares procedure minimizing a criterion of the form:

$$\mathbf{x}_\bullet = argmin_{\mathbf{x}} \underbrace{\left[\sum_{\{\mathbf{m}\}} w_{\mathbf{m}} f_{\mathbf{m}}(\mathbf{x})^2\right] / \left[\sum_{\{\mathbf{m}\}} w_{\mathbf{m}}\right]}_{\left[\epsilon_{\mathbf{f},\{\mathbf{m}\}}\right]^2} \tag{50}$$

The matches having $w_{\mathbf{m}} = 0$ are outliers and should not be further taken into account. The parameter $\mathbf{x}$ is finally estimated by solving the weighted least-squares problem.

In our case, the weighted least-squares criterion have been defined in equations (6) and (8).

4.2.4 Computing covariances and comparing different models.

In order to average properly these values, at each step of the computation, a covariance is estimated. It is very easy to estimate these covariances in our case, since we minimize a least-square criterion and obtain:

$$\mathbf{x}_\bullet = argmin_{\mathbf{x}} \underbrace{\left[\sum_{\{\mathbf{m}\}} w_{\mathbf{m}} f_{\mathbf{m}}(\mathbf{x})^2\right] / \left[\sum_{\{\mathbf{m}\}} w_{\mathbf{m}}\right]}_{\left[\epsilon_{\mathbf{f},\{\mathbf{m}\}}\right]^2} \quad \Rightarrow \Lambda_{\mathbf{x}_\bullet}^{-1} \simeq \sum_{\{\mathbf{m}\}} w_{\mathbf{m}} \frac{\partial f_{\mathbf{m}}(\mathbf{x})}{\partial \mathbf{x}} \frac{\partial f_{\mathbf{m}}(\mathbf{x})}{\partial \mathbf{x}}^T \tag{51}$$

[9] The ν-trimmed-median is defined as follows, for $\nu \in [0,1]$: let us sort the n residual in increasing order. The ν-trimmed-median is the i-th value in this series, with $i = \nu\, n$. The median corresponds to the middle value of this series, but we can select any other value between 0 and 1.

This computation allows not only to estimate $\mathbf{x}_\bullet$ but also its *information* (defined as the inverse of a covariance) $\Lambda_{\mathbf{x}_\bullet}^{-1}$. Moreover, assuming that the uncertainty on $f_{\mathbf{m}}(\mathbf{x})$ is corrupted by an additive Gaussian noise, the quantity :

$$\chi^2 = \frac{\left[\epsilon_{\mathbf{f},\{\mathbf{m}\}}\right]^2}{card(\{\mathbf{m}\}) - dim(\mathbf{x})} \tag{52}$$

is related to a χ-*square* distribution, for which the lower χ^2 the more significant the estimation.

This test will be used to verify if a model with less parameters, although its residual error is expected to be higher, is not more robust in terms of data representation.

In order to avoid testing all models using the hierarchy of figure 4, we can start the estimation from the root of this hierarchy, the *stationary structure assumption*, and program the system to take as best model the first model which weighted residual error is smaller that every model higher in the hierarchy. This should allow us to get a model with a minimal number of parameter, and a maximum significance. It is implicitly assumed that the significance is monotonic in the hierarchy, i.e. that if a model is less significant than a more specific model, more general models are even less significant.

4.2.5 Implementing a motion module

The data representation for such a composite rigid displacement is:

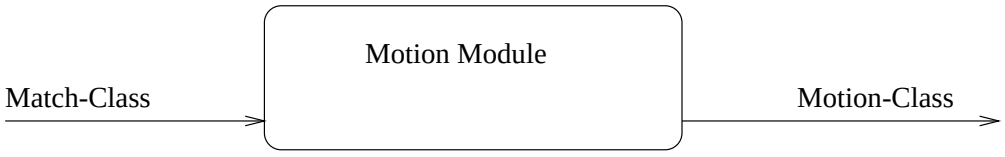
```
{
 /* Type of model of displacement */
 motion_class, /* Motion class which corresponds to the best estimation */
 {pa, pb, ...}, /* Probability of error for each motion class */
 /* Parameters defining the planar or rigid displacement */
Either
 F, C_F, /* F-matrix and inverse of its covariance */
Or
 H, C_H, /* H-matrix and inverse of its covariance */
 /* List of matches coherent with this displacements */
 {m1, m2, ...},
 {p1, p2, ...},  /* Probability to belong to this motion class */

} MotionClass;
```

The data representation for a match used as input of the module is is the following:

```
{
 m        = {  u,  v },        /* 2D Location in the first view */
 m'       = { u', v' },        /* 2D Location in the second view */
 C_m      = { Cuu, Cuv, Cvv }, /* Inverse of the covariance of the match,
                                  we use wm=sqrt(Cuu*Cvv) in the criterion */
} MatchClass;
```

Collecting all this information, we easily obtain a motion module which takes an estimation of matches as input and outputs information on the displacement parameters and of the class of displacement, as shown here :



4.3 Experimental results

We use a Least Median Square minimization to detect oultiers (0 outliers means that there is more than 50 % outliers or ν % outliers). Then, we refine the computation of the F-matrix or H-matrix on the points which are not outliers, and determine thus the residual error, in pixels.

The method, which allows us to determine which kind of displacement we have in the scene, is to search in the previous graph from the leaf to the top of the tree ("general rigid displacement") the displacement which have the least residual error (when a parent has an error bigger than a son, we stop the progression of the computation in this branch).

4.3.1 Using synthetic data

Let us now experiment the estimation of different models as discussed in this paper.

We first consider a synthetic scene corresponds to a set of two planes as for the real scene taken in consideration in the sequel (see figure 4.3.2). We analyze some particular case of displacements, and add a noise on each match, with a standard deviation, with $\sigma = 0.3$ pixel, which corresponds to usual precision of sub-pixelic corner detectors and matchers.

In each case, the residual error corresponds to the following square root of the quadratic error defined by the criteria of equation (6) or (8).

Simulation of a pure translation. The displacement is : t = (21, 13, 30), w = (0, 0, 0).

We write F_{th} the theoretical matrix and F_{exp} the matrix obtained :

$$F_{th} = \begin{bmatrix} -0 & -0.00069705 & 0.420089 \\ 0.00069705 & -0 & -0.568793 \\ -0.420089 & 0.568793 & 0 \end{bmatrix}, F_{exp} = \begin{bmatrix} 0 & 0.000715133 & -0.417255 \\ -0.000715133 & 0 & 0.570874 \\ 0.417255 & -0.570874 & 0 \end{bmatrix} \quad (53)$$

The results for different hypotheses are collected in table (1).

displacement	number of outliers	residual error
stationary structure	0	18.1247
pure retinal translation	64	2.43827
planar retinal displacement	45	1.06696
retinal displacement (F22 = 0)	54	0.440651
retinal displacement (F22 = 1)	55	0.433421
pure translation	63	0.393062
zoom displacement (s2 = 0; s'2 = 0; s0.s'0 != 0)	63	7.94626
zoom displacement linear (s2 = 1; s'2 = 1)	48	0.480495
general rigid displacement	55	0.395976

Table 1: Table of residues for a pure translation ($\sigma = 0.3$ pixel).

In this case, it is clear that the system will consider that the displacement is a pure translation, as expected. This is still the case with the addition of noise.

Simulation of a pure retinal translation. The displacement is : t = (21, 13, 0), w = (0, 0, 0)

We write F_{th} the theoretical matrix and F_{exp} the matrix obtained :

$$F_{th} = \begin{bmatrix} -0 & 0 & 0.372189 \\ -0 & -0 & -0.601228 \\ -0.372189 & 0.601228 & 0 \end{bmatrix}, F_{exp} = \begin{bmatrix} 0 & 0 & 0.372188 \\ 0 & 0 & -0.601229 \\ -0.372188 & 0.601229 & 0 \end{bmatrix} \quad (54)$$

The results for different hypotheses are collected in table (2).

displacement	number of outliers	residual error
stationary structure	0	17.8517
pure retinal translation	44	0.405941
planar retinal displacement	97	0.748497
retinal displacement (F22 = 0)	59	0.396306
retinal displacement (F22 = 1)	59	0.474059
pure translation	47	0.409366
zoom displacement (s2 = 0; s'2 = 0; s0.s'0 != 0)	0	0.580022
zoom displacement linear (s2 = 1; s'2 = 1)	57	0.422898
general rigid displacement	72	0.400766

Table 2: Table of residues for a pure retinal displacement ($\sigma = 0.3$ pixel).

In this case, the residual error is minimal not for the pure retinal translation but for a more general retinal displacement. However, because the system chooses the displacement for which the residual error is minimal with respect to displacements which are direct generalization, the pure retinal translation is still selected, because it corresponds to the first local minimum in the tree of models. Our heuristic is thus quite important here.

Simulation of a retinal displacement. The displacement is : t = (21, 13, 0), w = (0, 0, 0.4)

We write F_{th} the theoretical matrix and F_{exp} the matrix obtained :

$$F_{th} = \begin{bmatrix} -0 & 0 & 0.00407479 \\ -0 & -0 & -0.00658236 \\ -0.00122967 & 0.00764325 & -0.99994 \end{bmatrix}, F_{exp} = \begin{bmatrix} 0 & 0 & -0.00358023 \\ 0 & 0 & 0.00730432 \\ 0.000448378 & -0.00813084 & 0.999934 \end{bmatrix} \quad (55)$$

The results for different hypotheses are collected in table (3).

Again, we still have a minimal residual error for the right displacement, as expected. Note that a planar retinal displacement is also a plausible model, and might be detected as the most plausible displacement, if the level of noise increases. This is due to the fact that the variation in depth in the synthetic scene is not important, so that for high level of noise, the perception of the relief can be deleted because of the noise.

Simulation of a planar retinal displacement. The displacement is : t = (21, 13, 0), w = (0, 0, 0.4)

We write H_{th} the theoretical matrix and H_{exp} the matrix obtained :

$$H_{th} = \begin{bmatrix} 0.00845897 & -0.00297253 & 0.573987 \\ 0.00379271 & 0.00713407 & -0.818739 \\ 0 & 0 & 0.00772858 \end{bmatrix}, H_{exp} = \begin{bmatrix} 0.00828338 & -0.00295254 & 0.583445 \\ 0.00376835 & 0.00699834 & -0.81203 \\ 0 & 0 & 0.00759828 \end{bmatrix} \quad (56)$$

displacement	number of outliers	residual error
stationary structure	5	69.9445
pure retinal translation	46	46.112
planar retinal displacement	100	0.807336
retinal displacement (F22 = 0)	106	0.697904
retinal displacement (F22 = 1)	59	0.41378
pure translation	46	37.9302
zoom displacement (s2 = 0; s'2 = 0; s0.s'0 != 0)	59	37.0459
zoom displacement linear (s2 = 1; s'2 = 1)	0	46.279
general rigid displacement	72	0.421729

Table 3: Table of residues for a retinal displacement ($\sigma = 0.3$ pixel).

The results for different hypotheses are collected in table (4).

displacement	number of outliers	residual error
stationary structure	4	92.2992
pure retinal translation	26	33.504
planar retinal displacement	2	0.602899
retinal displacement (F22 = 0)	27	0.426322
retinal displacement (F22 = 1)	33	0.367844
pure translation	38	17.0126
zoom displacement (s2 = 0; s'2 = 0; s0.s'0 != 0)	15	8.03349
zoom displacement linear (s2 = 1; s'2 = 1)	0	10.7797
general rigid displacement	30	0.381183

Table 4: Table of residues for a planar retinal displacement ($\sigma = 0.3$ pixel).

In this case, the planar retinal displacement is well detected but, retinal displacement would be chosen as the optimal model. This is due to the fact that a retinal displacement also corresponds to the displacement of a fronto-parallel plane, while we have chosen a plane which is not far from this configuration.

This shows that the "decision" is not unique and that the output of this module must not be considered as a unique answer but as a set of hypotheses with different levels of probability.

Simulation of general displacement. The displacement is : t = (21, 13, 30), w = (0.2, 0.01, 0.4)

We write F_{th} the theoretical matrix and F_{exp} the matrix obtained :

$$F_{th} = \begin{bmatrix} -1.31391e-06 & -2.92072e-06 & 0.0028296 \\ 3.21358e-06 & -1.83e-06 & -0.00217196 \\ -0.000864567 & 0.00348618 & -0.999987 \end{bmatrix}, F_{exp} = \begin{bmatrix} 1.61428e-06 & 3.48703e-06 & -0.00291376 \\ -3.79558e-06 & 2.00566e-06 & 0.00189159 \\ 0.000829183 & -0.00334547 & 0.999988 \end{bmatrix} \quad (57)$$

The results for different hypotheses are collected in table (5).

Here, very clearly, the best model is the general displacement. This is also an expected result, since statistical tests always easily *reject* erroneous hypotheses. In other words, we

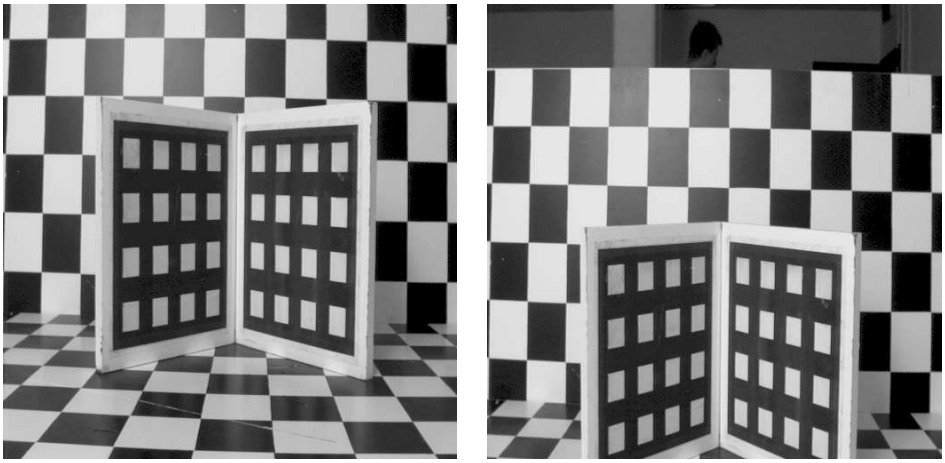
displacement	number of outliers	residual error
stationary structure	0	169.397
pure retinal translation	44	37.8552
planar retinal displacement	75	2.10893
retinal displacement (F22 = 0)	84	1.97447
retinal displacement (F22 = 1)	68	1.2561
pure translation	59	35.3121
zoom displacement (s2 = 0; s'2 = 0; s0.s'0 != 0)	44	55.8793
zoom displacement linear (s2 = 1; s'2 = 1)	0	56.8783
general rigid displacement	59	0.444719

Table 5: Table of residues for a general displacement without zoom ($\sigma = 0.3$ pixel).

have no risk to accept a biased model with a smaller number of enough parameters, but only to use a model with too many parameters. The explanation of this effect is simple: the estimator not only estimates the true displacement but also might try to parameterize the particular configuration of the noise.

4.3.2 An example with real data : grid scene.

We consider a sequence of 16 images as illustrated in figure (4.3.2), for which the displacement is an approximative retinal translation, with some erroneous rotation because of the actual set-up. A retinal displacement is thus expected.



The early-vision module has provided matches between 44 points and the errors are given in table (6).

The model is correctly estimated also in this real case, which thus allow us to conclude on the validity of the proposed mechanism.

4.3.3 An example with real data : external scene.

We consider a sequence of 9 images as illustrated in figure (4.3.3), for which the displacement is an approximative translation, with some erroneous. A pure translation is thus expected.

displacement	number of outliers	residual error
stationary structure	0	17.9633
pure retinal translation	0	17.4948
planar retinal displacement	11	6.52437
retinal displacement (F22 = 0)	8	1.95843
retinal displacement (F22 = 1)	8	0.735489
pure translation	0	14.7264
zoom displacement (s2 = 0; s'2 = 0; s0.s'0 != 0)	0	11.677
zoom displacement linear (s2 = 1; s'2 = 1)	4	0.887803
general rigid displacement	11	0.84458

Table 6: Table of residues for the real scene.

The early-vision module has provided matches between 141 points and the errors are given in table (7).

displacement	number of outliers	residual error
stationnary structure	34	5.8149
pure retinal translation	45	3.24404
planar retinal displacement	31	4.73934
retinal displacement (F22 = 0)	42	3.23533
retinal displacement (F22 = 1)	50	1.78939
pure translation	44	1.67575
zoom displacement (s2 = 0; s'2 = 0; s0.s'0 != 0)	44	3.19459
zoom displacement linear (s2 = 1; s'2 = 1)	44	1.12893
general rigid displacement	27	1.55442

Table 7: Table of residues for an external scene.

The values of errors confirm us the validity of our model.

4.3.4 An example with real data : general displacement.

We expect, in the case illustrated in figure (4.3.4), that all particular cases are not valid and only the general displacement case will be valid.


The early-vision module has provided matches between 74 points and the errors are given in table (8).

displacement	number of outliers	residual error
stationnary structure	13	48.3722
pure retinal translation	17	11.9222
planar retinal displacement	22	15.1314
retinal displacement (F22 = 0)	27	1.58154
retinal displacement (F22 = 1)	22	1.49724
pure translation	18	10.6164
zoom displacement (s2 = 0; s'2 = 0; s0.s'0 != 0)	0	6.05225
zoom displacement linear (s2 = 1; s'2 = 1)	9	4.08582
general rigid displacement	25	0.606527

Table 8: Table of residues for a real general displacement.

5 Conclusion

In the present paper we have reviewed and completed the description of a general framework which allows not only to estimate a minimal parameterization of the rigid displacement between two frames, but also to determine several particular cases which occur in practice and have important advantages with respect to the calibration problem. This is true for several standard displacement, except a zoom displacement which seems to be a singular case, for the proposed model.

The statistical framework to implement these equations has been already described in [11] and has been applied here to the estimation of collineations from a minimal parameterization. This paper however generalizes the set of models to general rigid displacements, and proposes a complete analysis of the underlying rigid displacement in each case.

Similar attempts to use degenerated models of parameterization of motion have been already issued in the past [11, 8, 12]. However, we collect here new results about the Euclidean representation associated to each parameterization. Furthermore, the implementation also

integrates two new aspects: (i) clustering data and (ii) testing different models to represent the data.

Finally, this work tries to develop -with a certain degree of completeness- all the different singularities which occur for a rigid displacement and which can be detected without calibration. A practical motion module has been developed and successfully experimented.

A step further, we will use this hierarchical approach to not only parameterize the retinal displacement but also analyze the calibration of the visual system and recover the scene structure. A preliminary study has been issued [3] for retinal displacements.

References

[1] O. Faugeras. *Three-dimensional Computer Vision: a geometric viewpoint.* MIT Press, Boston, 1993.

[2] C. Harris and M. Stephens. A combined corner and edge detector. In *Proceedings Alvey Conference*, pages 189–192, 1988.

[3] D. Lingrand and T. Viéville. Dynamic foveal 3d sensing using affine models. Technical Report RR-2687, INRIA, 1995.

[4] T. Luong. *Matrice Fondamentale et Calibration Visuelle sur l'Environnement.* PhD thesis, Université de Paris-Sud, Orsay, 1992.

[5] W. Press, B. Flannery, S. Teukolsky, and W. Vetterling. *Numerical recipes, the art of scientific computing.* Cambridge University Press, Cambridge, U.S.A., 1988.

[6] P. Rousseeuw and A. Leroy. *Robust Regression and Outlier Detection.* John Wiley & Sons, New York, 1987.

[7] L. Shapiro and M. Brady. Rejecting outliers and estimating errors in an orthogonal regression framework. Tech.Report OUEL 1974/93, Dept. Engineering Science, University of Oxford, Feb. 1993.

[8] P. Torr, A. Zisserman, and S. Maybank. Robust detection of degenerated configurations for the fundamental matrix. In *5th International Conference on Computer Vision*, pages 1037–1042, 1995.

[9] T. Viéville. Autocalibration of visual sensor parameters on a robotic head. *Image and Vision Computing*, 12, 1994.

[10] T. Viéville, E. Clergue, R. Enciso, and H. Mathieu. Experimentating with 3D vision on a robotic head. *Robotics and Autonomous Systems*, 14(1), 1995.

[11] T. Viéville, C. Zeller, and L. Robert. Using collineations to compute motion and structure in an uncalibrated image sequence. *International Journal of Computer Vision*, 1995. To appear.

[12] C. Wiles and M. Brady. Closing the loop on multiple motion. In *5th International Conference on Computer Vision*, pages 308–313, 1995.

[13] R. Willson. *Modeling and Calibration of Automated Zoom Lenses.* PhD thesis, Department of Electical and Computer Engineering, Carnegie Mellon University, 1994.

[14] C. Zeller and O. Faugeras. Applications of non-metric vision to some visual guided tasks. In *The 12th Int. Conf. on Pattern Recognition*, pages 132–136, 1994.

[15] Z. Zhang, R. Deriche, Q.-T. Luong, and O. Faugeras. A robust approach to image matching: Recovery of the epipolar geometry. In *Proc. International Symposium of Young Investigators on Information\Computer\Control*, pages 7–28, Beijing, China, Feb. 1994.

Acknowledgments We are especially thankful to **O.D. Faugeras** for some powerful ideas which are at the origin of this work. This work is partially achieved under **Esprit Project P8878/REALISE**.

Unité de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unité de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unité de recherche INRIA Rhône-Alpes, 46 avenue Félix Viallet, 38031 GRENOBLE Cedex 1
Unité de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unité de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
ISSN 0249-6399