INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

A Projective Framework for Structure and Motion Recovery from Two Views of a Piecewise Planar Scene

Adrien Bartoli, Peter Sturm and Radu Horaud

N° 4070

October 2000

THÈME 3

Rapport de recherche

ISSN 0249-6399 ISRN INRIA/RR--4070--FR+ENG

A Projective Framework for Structure and Motion Recovery from Two Views of a Piecewise Planar Scene

Adrien Bartoli, Peter Sturm and Radu Horaud

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

Rapport de recherche n° 4070 — October 2000 — 30 pages

Abstract: In this paper, we consider the problem of finding an optimal reconstruction from two views of a piecewise planar scene. We consider the general case of uncalibrated cameras, hence place us in a projective framework. In this case, there is no meaningful metric information about the object space that could be used to define optimization criteria. Taking into account that the images are then the only spaces where an optimization process makes sense, there is a need at each step of the reconstruction process, from the detection of planar structures to motion estimation and actual 3D reconstruction, of a consistent image level representation of geometric 3D structures. In our case, we need to represent camera motion and 3D points that are subject to coplanarity constraints. It is well known that camera motion between two views can be represented on the image level via the epipolar geometry (fundamental matrix). Coplanarity constraints can be expressed via a collection of 2D homographies. Unfortunately, these algebraic entities are over-parameterized in the sense that the 2D homographies must in addition obey constraints imposed by the epipolar geometry. We are thus looking for a minimal and consistent representation of motion (epipolar geometry) and structure (points+homographies) that in addition should be easy to use for minimizing reprojection error in a bundle adjustment manner. In this paper, we propose such a representation and use it to devise fast and accurate estimation methods for each step of the reconstruction process, including image point matching, plane detection and optimal triangulation of planes and points on planes. We make extensive use of the quasi-linear optimization principle. A great number of experimental results show that the new methods give superior results compared to approaches that do not estimate motion and multi-planar structure simultaneously and consistently, even in cases when the observed scene is not perfectly coplanar.

Key-words: epipolar geometry, plane homography, consistent image level representation, constrained estimation, quasi-linear optimization, maximum likelihood estimation, plane detection, points on plane triangulation.

Unité de recherche INRIA Rhône-Alpes
655, avenue de l'Europe, 38330 Montbonnot-St-Martin (France)
Téléphone : +33 4 76 61 52 00 — Télécopie +33 4 76 61 52 52

Un système projectif de reconstruction d'une scène plane par morceau à partir de deux vues

Résumé : Dans cet article, nous considérons le problème de trouver une reconstruction optimale à partir de deux vues d'une scène plane par morceau. Nous considérons le cas de caméras non calibrées, qui nous place dans un cadre projectif. Dans ce cas, il n'y a pas d'information métrique significative pouvant servir à définir un critère. Les images étant alors les seuls espaces où un processus d'optimization à un sens, il y a besoin, à chaque étape du processus de reconstruction, de la détection des structures planes à l'estimation de la reconstruction et du mouvement entre les caméras, d'une représentation des structures géométriques 3D au niveau image. Dans notre cas, nous avons besoin de représenter le mouvement entre les deux caméras et des points 3D sujets à des contraintes de coplanarité. Il est connu que le mouvement entre les deux caméras peut être représenté par la géométrie épipolaire (ou de manière équivalente la matrice fondamentale). Les contraintes de coplanarité peuvent être exprimées par un ensemble d'homographies 2D. Malheureusement, ces entités algébriques sont sur-paramétrées au sens où les homographies 2D doivent satisfaire certaines contraintes imposées par la géométrie épipolaire. Nous cherchons donc une représentation minimale et consistante de la géométrie épipolaire et de la structure 3D qui devra être facile à utiliser pour minimiser une erreur de reprojection dans un processus d'alignement de faisceaux. Nous proposons une telle représentation et l'utilisons for construire des méthodes rapides et précises pour chaque étape du processus de reconstruction, incluant l'appariement des points image, la détection de plans et la triangulation optimale des plans et de points sur des plans. Nous utilisons en particulier le principe d'optimisation quasi-linéaire. Un grand nombre d'expérimentations montre que la nouvelle méthode donne des résultats de qualité supérieure aux approches non consistantes ou basées uniquement sur des points. Ceci est valable même dans certains cas où la scène observée n'est pas parfaitement plane par morceau.

Mots-clés : géométrie épipolaire, homographie plane, représentation consistante au niveau image, estimation contrainte, optimisation quasi-linéaire, estimation au maximum de vraisemblance, détection de plans, triangulation de points sur des plans.

1 Introduction

The recovery of structure and motion from images is one of the key goals in photogrammetry and computer vision. The special case of piecewise planar reconstruction is particularly important due to the large number of such scenes in man-made environments (e.g. buildings, floors, ceilings, roads, walls, furniture, etc.).

A complete system that would be able to achieve automatically such reconstruction tasks from any images (and so in a projective framework) and in a quick and optimal way would incorporate both the step of plane detection and an optimal structure and motion recovery. Consequently, and if we qualify an algebraic representation to be *consistent* when its number of dof (degrees of freedom) strictly corresponds to the essential dof of what it represents, there is a need for:

- a consistent image-level representation of plane structures;
- a consistent, fast and accurate plane homography estimator;
- a method for optimally reconstruct planes (and points on planes) while estimating motion.

The image based instead of *3D*-based representation of geometric constraints is necessary to allow matching constraints and is essential in a projective framework. In such a framework, it is well known that the geometry of two views of a piecewise planar scene can be represented by the *epipolar geometry* for the motion and a set of *plane homographies* (plus image points related by them) for the structure. These are very strong constraints from which we can expect better reconstruction results than from the traditional point-only based methods [9]. Using this plane representation, the above problems are difficult for the following reasons:

- image level algebraic structures that represent motion and piecewise planar structure for two views are an over-parameterization (e.g. for an m planes scene, the essential dof are $7 + 3m$, i.e. 7 for the epipolar geometry [12] and 3 for each plane in *3D* space, whereas the algebraic ones are $9(m + 1)$, i.e. 9 for each 3×3 matrix that represent the epipolar geometry [12] and each plane homography [10]) and the majority of constraints are non-linear;
- a consistent, fast and accurate homography estimator will incorporate such constraints, preferably not need a non-linear optimization and minimize a geometrically meaningful residual in both images.

In the following two paragraphs we review existing work and describe our contributions to the problems posed above.

1.1 Background and related work

There is a lot of work on interest point detection and image matching. Here, we consider that basic image point correspondences are available. By basic we mean that a certain amount of these correspondences may be wrong. These matches do not have to be clustered in coplanar groups in advance.

Maybe the first attempt to give a minimal parameterization of the epipolar geometry in a particular case (i.e. when both epipoles are finite) is described in [11, 12]. This work has been extended in [19] to the general case where the author gives an algorithm to compute the epipolar geometry and the best "map" (in other words parameterization) among 36 different choices. However, the optimization procedure is costly due to a step by step determination of the appropriate map. Moreover, the link with plane homographies has not been made.

In [3, 15], methods to compute motion and structure from motion in a piecewise planar environment are given. These methods work in the case of calibrated images and do not yield an *MLE* (Maximum Likelihood Estimator). The main problem with these methods is that there is not any image based global framework. The constraint of coplanarity has been studied in [14] but without an explicit image level parameterization of geometric structures. The results obtained in this work show that the accuracy of reconstruction is not better but that it appears visually more acceptable.

1.2 Our contributions

Our contributions are the following. We propose a plane homography estimator constrained by a given epipolar geometry and respecting the constraints given above, in other words, given a motion which sets a *3D* framework, we are able to estimate planes on the image level. We show that the optimization process can be conducted using the quasi-linear principle instead of a non-linear one such as Levenberg-Marquardt [13], which is of significant interest from an implementation and complexity point of view. Next, we define a complete and minimal structure and motion algebraic representation and use it to optimally reconstruct planes and points on planes, where we reduce the 36 maps [19] for the epipolar geometry to 3 using an additional constraint and an image rotation and we incorporate plane homographies in the parameterization. In other words, this step optimally refines the previously estimated structure and motion. Finally, we present a framework for automatic scene segmentation into coplanar clusters and reconstruction of the recovered structures. This step relies on the previously defined image level algebraic representation of plane homographies and devised estimators. It supplies an initialization to the process of optimal structure and motion estimation. A great number of experimental results (for both simulated and real images) validate each part of the work.

1.3 Paper organization

The next sections present respectively some preliminaries needed to read this paper, the study of plane homography estimation from point correspondences, the definition of a consistent image level structure for two views of a piecewise planar scene and a framework for an automatic reconstruction system. We then conclude and give perspectives for future work. Finally, an appendix gives an implementation oriented formulation of the main results obtained in this paper.

2 Preliminaries

In this section, we give some preliminaries required for an easy reading of this paper, starting with our notations and the description of the pinhole camera model and the two view relations. We then examine the transformation induced by two projections of a plane and how to reconstruct points on planes. Finally, we give possibilities to upgrade a projective reconstruction to a metric framework.

2.1 Notations

Throughout this paper, projective (resp. Euclidean) spaces of dimension n are typeset $\mathbb{P}^n$ (resp. $\mathbb{E}^n$). We use homogeneous coordinates which implies that points, lines, planes and transformations are defined up to a non-null scalar. Equality of homogeneous coordinates is designated by $\sim$.

Vectors are typeset in boldface ($\mathbf{x}$, $\mathbf{X}$) and matrices using sans-serif fonts (F, H). Points, lines and planes are represented by vectors and designated by upper case letters for *3D* points ($\mathbf{X}$), lower case letters for *2D* points ($\mathbf{x}$) and greek letters for planes (π). The *2D* points in the second image are followed by a prime ($\mathbf{x}'$). There is not any formal difference between a physical entity and its algebraic representations.

Point and set-of-point correspondences are respectively designated by ($\mathbf{x} \leftrightarrow \mathbf{x}'$) and $\{\mathbf{x} \leftrightarrow \mathbf{x}'\}$. The cross product between two 3-vectors is designated by $\times$, the associated matrix by $[.]_\times$ and the scalar product between two n-vectors $\mathbf{u}$ and $\mathbf{v}$ by $\mathbf{u}^\mathsf{T}\mathbf{v}$ where $^\mathsf{T}$ is the transposition. Identity matrix of dimension $m \times m$ and null n-vector are respectively typeset I_m and $\mathbf{0}_n$. The abbreviation "dof" will often be used for "degrees of freedom".

2.2 Camera model

A camera is described by the widely used pinhole model. It is modeled by a 3×4 matrix P mapping points $\mathbf{X}$ from the projective space $\mathbb{P}^3$ to points $\mathbf{x}$ on the projective plane $\mathbb{P}^2$:

$$\mathbf{x} \sim \mathsf{P}\mathbf{X}.$$

2.3 Two view geometry

There is a strong link between the two projections $\mathbf{x}$ and $\mathbf{x}'$ of the same *3D* point on the cameras. This link is modeled by the well known epipolar geometry [12, 19], expressed as:

$$\mathbf{x}'^\mathsf{T}\mathsf{F}\mathbf{x} = 0, \tag{1}$$

where F is a 3×3 matrix called the fundamental matrix, encapsulating all the epipolar geometry. The fundamental matrix has 7 dof and its rank is 2. The left and right epipoles are respectively designated by $\mathbf{e}$ and $\mathbf{e}'$ and defined such that $\mathsf{F}\mathbf{e} = \mathsf{F}^\mathsf{T}\mathbf{e}' = \mathbf{0}$.

It is usefull to recall the expressions of the uncalibrated projection matrices P and P' with respect to a given *reference plane homography* H_r:

$$\begin{aligned} \mathsf{P} &\sim (\mathrm{I}_3|\mathbf{0}_3) \\ \mathsf{P}' &\sim (\mathsf{H}_r|\mathbf{e}'), \end{aligned} \tag{2}$$

where matrix $\underline{\mathsf{H}}$ fixes the projective base and can be, as $\mathbf{e}'$, extracted from F up to a 3-parameters ambiguity (see below).

2.4 Plane homographies

The projections $\mathbf{x}$ and $\mathbf{x}'$ of the same *3D* point lying on a plane are related by a plane homography described by a full-rank 3×3 matrix H:

$$\mathbf{x}' \sim \mathsf{H}\mathbf{x}. \tag{3}$$

This matrix is linked to the epipolar geometry [10] by the following relation:

$$\mathsf{F} \sim [\mathbf{e}']_\times \mathsf{H}. \tag{4}$$

Furthermore, if we fix a reference plane homography H_r, any other plane homography H_i can be characterized by a 3-vector $\mathbf{a}_i$ using:

$$\mathsf{H}_i \sim \mathsf{H}_r + \mathbf{e}'\mathbf{a}_i^\mathsf{T}. \tag{5}$$

The reference homography can be any plane homography. A singular one can be extracted from the fundamental matrix in the following way:

$$\mathsf{H}_r \sim [\mathbf{e}']_\times \mathsf{F}. \tag{6}$$

2.5 Projective reconstruction of points on planes

For a given pair $(\mathbf{x} \leftrightarrow \mathbf{x}')$ of corresponding points satisfying a plane homography H (equation (3)), the corresponding *3D* point $\mathbf{X}$ lies on a plane related to H and is given by:

$$\mathbf{X} \sim \left(\mathbf{x}^\mathsf{T}, \mathbf{x}^\mathsf{T}\mathbf{a}\right)^\mathsf{T},$$

where $\mathbf{a}$ gives the parameters of H relative to the reference homography H_r, i.e. in the projective base defined by equations (2).

2.6 From projective to metric reconstruction

It is well known that a projective reconstruction can be updated to metric using a special *3D* homography H_u which depends on the intrinsic parameters of both cameras. This homography can be estimated in two ways:

- an *automatic* way: using an autocalibration method (for a list of references, refer to [10]). In the case of two views, one has to assume some intrinsic parameters to be a priori known [6];
- a *manual* way: by defining the transformation of a projective base into a Euclidean equivalent. In practice, one has just to designate five points and give their metric coordinates in an Euclidean frame.

3 Plane homography estimation from point matches

In this section, we first describe two kinds of distance in the Euclidean plane $\mathbb{E}^2$, an algebraic and a geometric one. We then establish the relationship between them to emphasize the bias of the algebraic distance, leading to the principle of quasi-linear estimation. Secondly, we show how to estimate a plane homography from coplanar point correspondences with a basic quasi-linear estimator. Next, we show how a new estimator can be devised, based on quasi-linearity and minimizing a symmetric (in both images) residual. Moreover, the result obtained using this estimator is a fully epipolar geometry compatible plane homography.

3.1 Algebraic and geometric distance in $\mathbb{E}^2$

Let us consider two points of the Euclidean plane $\mathbb{E}^2$ represented by their homogeneous 3-vector coordinates $\mathbf{x}$ and $\mathbf{y}$.

The *Euclidean distance* d in $\mathbb{E}^2$ is given by:

$$d^2(\mathbf{x}, \mathbf{y}) = \left(\frac{x_1}{x_3} - \frac{y_1}{y_3}\right)^2 + \left(\frac{x_2}{x_3} - \frac{y_2}{y_3}\right)^2. \tag{7}$$

Now let us construct the *algebraic distance* d_a. The two points are the same if and only if $\mathbf{x}$ and $\mathbf{y}$ are equal up to a non-null multiplicative scalar or equivalently if their cross-product is the null 3-vector:

$$\mathbf{x} \sim \mathbf{y} \Leftrightarrow \mathbf{v} = \mathbf{0} \text{ where } \mathbf{v} = \mathbf{x} \times \mathbf{y}.$$

As a consequence, the squared $\mathcal{L}_2$-norm of the cross-product, $\mathbf{v}^\mathsf{T}\mathbf{v}$, can be used as a measure of distance between these two points. The development of its algebraic expression leads to the fact that only two elements of $\mathbf{v}$ are linearly independent. Consequently, we may define the *algebraic distance* d_a between two points by:

$$\begin{aligned} d_a^2(\mathbf{x}, \mathbf{y}) &= v_1^2 + v_2^2 = ||\mathsf{S}(\mathbf{x} \times \mathbf{y})||^2 \text{ where } \mathsf{S} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \\ &= (x_1 y_3 - x_3 y_1)^2 + (x_2 y_3 - x_3 y_2)^2, \end{aligned} \tag{8}$$

which is a bilinear expression in $\mathbf{x}$ and $\mathbf{y}$. The choice for the elements v_1 and v_2 is motivated by two reasons:

- it is essential for the important equation (9) of similarity between algebraic and Euclidean distances to be established;
- the points considered belong to the Euclidean and not the projective plane (our definition of d_a is no longer reasonable for $\mathbb{P}^2$ due to possible infinite points i.e. if a point, for example $\mathbf{x}$, lies on $\mathbf{l}_\infty$, while the other, $\mathbf{y}$ does not, $d_a^2(\mathbf{x}, \mathbf{y}) = ||\mathbf{x}||^2$ does not depend on $\mathbf{y}$ and if both points $\mathbf{x}$ and $\mathbf{y}$ lie on $\mathbf{l}_\infty$, $d_a^2(\mathbf{x}, \mathbf{y}) = 0$, even if the points are distinct).

It is important to note that d_a does not have a physical meaning because it is not linked to actual measurement error and is not invariant to gauge freedom in the measuments.

The two distances d and d_a are related as follows:

$$d^2(\mathbf{x}, \mathbf{y}) = w_{\mathbf{x},\mathbf{y}} d_a^2(\mathbf{x}, \mathbf{y}) \text{ where } w_{\mathbf{x},\mathbf{y}} = \frac{1}{(x_3 y_3)^2}. \tag{9}$$

This equation relates the bilinear distance d_a and the non-linear distance d. The factors $w_{\mathbf{x},\mathbf{y}}$ can be thought of as weights or bias [8] between the two distances. The next sections illustrate how this can be used to devise quasi-linear estimators in various cases.

3.2 Unconstrained estimation

Quasi-linear estimators have been studied for various purposes [1]. It has been observed that this kind of estimator is well-suited when all constraints can be linearly expressed [1] but may fail in other cases (e.g. for the estimation of the fundamental matrix [19]).

Here, we give a simple completely linear example in order to explain the principle of quasi-linear estimators and to evaluate possible performances of such processes. Our example is the non symmetric estimation of a plane homography from image correspondences. Unconstrained estimation means that we do not have any a priori knowledge about the two views' geometry and the observed scene, apart from the fact that the observed points are coplanar in space. Non symmetric means that the error is being minimized is only one image.

Let $\{\overline{\mathbf{x}} \leftrightarrow \overline{\mathbf{x}}'\}$ be a set of n corresponding image points arising from two projections of a plane and $\{\mathbf{x} \leftrightarrow \mathbf{x}'\}$ the noisy observations of these points. These matches are related by an 8 dof plane homography H according to $\overline{\mathbf{x}}' \sim \mathsf{H}\overline{\mathbf{x}}, \forall(\overline{\mathbf{x}} \leftrightarrow \overline{\mathbf{x}}')$. To estimate the 9 algebraic parameters of this homography, a common approximation [10] is to assume that only the second set of image points are noisy (i.e. $\{\mathbf{x}\} = \{\overline{\mathbf{x}}\}$), which is generally wrong but yields a linear estimator for the elements of H when d_a is used. This implies that the residual is to be minimized only in the second image:

$$\mathsf{H} \sim \underset{\mathsf{H}}{argmin} \sum_{\{\overline{\mathbf{x}} \leftrightarrow \mathbf{x}'\}} d_a^2(\mathbf{x}', \mathsf{H}\overline{\mathbf{x}}). \tag{10}$$

This criterion optimizes the non symmetric transfer algebraic distance d_a and can be expressed using a linear equation system. The ambiguity arising from the free scale factor of H is removed using the additional constraint $||\mathsf{H}||^2 = 1$. However, despite its advantages, this criterion is bad because of the previously given reasons about the bias of d_a. To overcome this problem, one can replace d_a by d, but the criterion obtained is no longer linear:

$$\mathsf{H} \sim \underset{\mathsf{H}}{argmin} \sum_{\{\overline{\mathbf{x}} \leftrightarrow \mathbf{x}'\}} d^2(\mathbf{x}', \mathsf{H}\overline{\mathbf{x}}). \tag{11}$$

Indeed, it has to be optimized using a complicated and costly non-linear optimization method such as Levenberg-Marquardt or Newton [13]. Fortunately this drawback can be overcome using a quasi-linear method in the following way. Using equation (9), (11) becomes:

$$\mathsf{H} \sim \underset{\mathsf{H}}{argmin} \sum_{\{\overline{\mathbf{x}} \leftrightarrow \mathbf{x}'\}} w^2_{\mathbf{x}',\mathsf{H}\overline{\mathbf{x}}}\, d_a^2(\mathbf{x}', \mathsf{H}\overline{\mathbf{x}}). \tag{12}$$

The problem of non-linearity now comes from the weight factors $w_{\mathbf{x}',\mathsf{H}\overline{\mathbf{x}}}$. If they are known, the criterion becomes linear. This leads to the following iterative algorithm. Weights $w_{\mathbf{x}',\mathsf{H}\overline{\mathbf{x}}}$ are initialized to 1 and iteratively updated. The loop is ended when the weights or the geometric transfer residual converge. This algorithm is summarized in table 1. It is a quasi-linear optimization and can be viewed as a convergence from the minimization of an algebraic to a Euclidean residual. This estimator can be used with a minimum of four point correspondences.

1. *initialization:* $w_{\mathbf{x}',\mathsf{H}\overline{\mathbf{x}}} = 1$;
2. *estimation:* estimate H from equation (12) using standard weighted least squares;
3. *weighting:* use H to update weights $w_{\mathbf{x}',\mathsf{H}\overline{\mathbf{x}}}$ according to equation (9);
4. *iteration:* iterate steps 2. and 3. until convergence (see text).

Table 1: Quasi-linear non symmetric estimation of a plane homography from point correspondences.

The efficiency of the quasi-linearity principle described above is validated using simulated data to compare four estimators: *Lin**, *Lin*, *NLin* and *QLin*.

- *Lin** is based on criterion (10) and using raw image coordinates (this is a *DLT* which stands for Direct Linear Transform [7]);
- *Lin* is based on the same criterion but uses a prior data normalization (this is a *DLT*);
- *NLin* is based on an iterative optimization of criterion (11) using Levenberg-Marquardt [13];
- *QLin* is based on the algorithm of table 1.

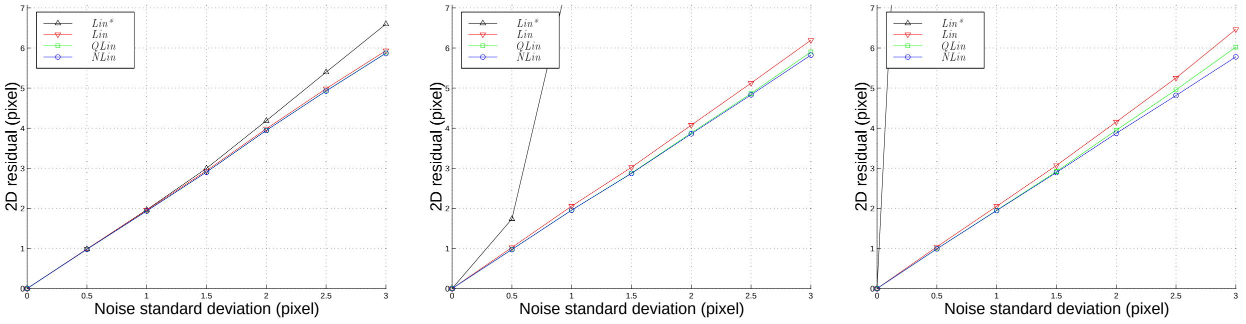
The residual measured is the *2D* non symmetric geometric transfer *RMS* (Root Mean Square) given by:

$$E_1(\mathsf{H}) = \sqrt{\frac{1}{n} \sum_{\{\mathbf{x} \leftrightarrow \mathbf{x}'\}} d^2(\mathbf{x}', \mathsf{H}\mathbf{x})}.$$

The test bench is constituted by n points lying on a one meter squared plane observed by two cameras at a distance ρ. Gaussian centered noise is added to the image points.

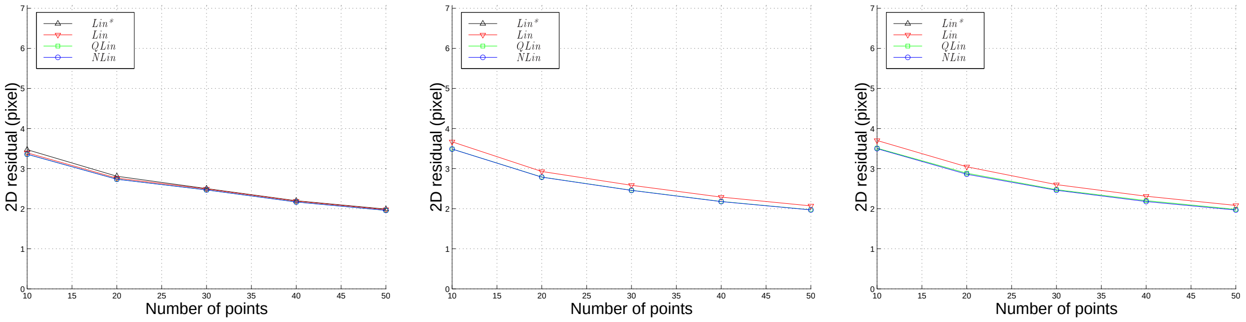
Figure 1 shows results obtained when varying noise standard deviation between 0 and 3 pixels (for $\rho=3$, 11 and 20 meters and using n=50 points) and figure 2 when varying n from 10 to 50 points (for $\rho=3$, 11 and 20 meters and using a 1 pixel standard deviation noise).

Figure 1 shows that *Lin** quickly degrades with distance and is only usable for small distances (i.e. for stable configurations) and that the last three algorithms perform better, *NLin* being the best and *Lin* the worst, in an acceptable range of difference, for all distances. Figure 2 shows that results are enhanced for all algorithms and for all distances when using more point correspondences.



(a) (b) (c)

Figure 1: Comparison of *2D* residual E_1 vs noise standard deviation for the estimators *Lin**, *Lin*, *NLin* and *QLin* for n=50 points. The distance plane/cameras ρ is (a): 3, (b): 11 and (c): 20 meters.



(a) (b) (c)

Figure 2: Comparison of *2D* residual E_1 vs number of points n used for the estimators *Lin**, *Lin*, *NLin* and *QLin* for a 1 pixel standard deviation noise. The distance plane/cameras ρ is (a): 3, (b): 11 and (c): 20 meters. Results of *Lin** are out of range for (b) and (c).

These conclusions are not related to true measurements. They just demonstrate how the algorithms can "stick" to the data provided but not how far from the truth the results provided by these algorithms are. Indeed, we expect that, even if the resulting homography sticks equally well to the data for all distances (graphs 1(a), 1(b) and 1(c) are nearly identical for the better methods), the homography might be worse if the distance increases.

Let us consider another measure: a distance between the estimated homography H and the true one $\overline{\mathsf{H}}$. We use the Frobenius norm of the difference of these two "normalized" matrices:

$$E_{\mathcal{F}}(\overline{\mathsf{H}}, \mathsf{H}) = ||\frac{\overline{\mathsf{H}}}{\overline{\mathsf{H}}_{kl}} - \frac{\mathsf{H}}{\mathsf{H}_{kl}}||^2 \text{ where } kl = \underset{ij}{argmax}\, \overline{\mathsf{H}}_{ij}.$$

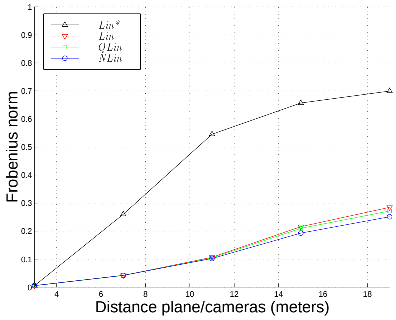


Figure 3: Comparison of the distance $E_{\mathcal{F}}$ between the estimated homography H and the true one $\overline{\mathsf{H}}$ for the estimators *Lin**, *Lin*, *NLin* and *QLin* vs the distance plane/cameras ρ for n=50 points and a 1 pixel standard deviation noise.

Figure 3 shows results obtained using this measure vs the distance plane/cameras ρ for n=50 and a 1 pixel standard deviation noise. We clearly see that, even if the residual E_1 is roughly the same for all distances, the quality of the estimated homography actually degrades with distance (i.e. less stable configurations). Consequently, in these cases, several important things, such as the orientation of the plane can not be well determined. Moreover, the estimated homography can not be used to automatically detect and add point correspondences that belong to the observed plane.

The cost of these algorithms can be compared using the number of linear systems solved during the process of optimization:

- *Lin** and *Lin*: 1;
- *NLin*: between 2 and 20;
- *QLin*: between 2 and 8.

Our conclusions are that the quality and cost augment from *Lin* to *NLin* and that *QLin* is a good compromise. The other advantages of *QLin* vs *NLin* are its easy implementation (it is just a loop over the linear method *Lin*) and a more stable convergence to a global optimum [1] (it is less sensitive to local minima than *NLin*).

3.3 Epipolar geometry constrained estimation

According to the previous section, quasi-linear estimators perform well under certain constraints and from various points of view. The question is now, how can we devise a "good" plane homography estimator based on the quasi-linear principle. Such an estimator follows the reduced approach [17] (i.e. minimizes a geometric residual in both images) and incorporates the external global constraint of compatibility with a given epipolar geometry, reducing the number of unknowns from 8 to 3 for each homography. Our constraints to build a "good" estimator are then:

- minimize a symmetric geometrical residual;
- the optimization process can be conducted using the principle of quasi-linearity;
- the result is fully compatible with a given epipolar geometry.

Now, let us devise such an estimator, starting from an algebraic expression of our constraints:

$$\mathsf{H} \sim \underset{\mathsf{H}}{argmin} \sum_{\{\mathbf{x} \leftrightarrow \mathbf{x}'\}} \left(d^2(\mathbf{x}', \mathsf{H}\mathbf{x}) + d^2(\mathbf{x}, \mathsf{H}^{-1}\mathbf{x}') \right). \tag{13}$$

The scale ambiguity and the compatibility with a given epipolar geometry are respectively fixed by the constraints $||\mathsf{H}||^2 = 1$ and equation (5). The epipolar geometry is supposed to be given in terms of the two epipoles and a reference plane homography (i.e. $\mathbf{e}, \mathbf{e}' \sim \mathsf{H}_r\mathbf{e}, \mathsf{H}_r$). Equation (5) reduces the number of unknowns and allows to rewrite the problem under the form:

$$\mathsf{H} \sim \mathsf{H}_r + \mathbf{e}'\mathbf{a}^\mathsf{T} \text{ where } \mathbf{a} = \underset{\mathbf{a}}{argmin} \sum_{\{\mathbf{x} \leftrightarrow \mathbf{x}'\}} \left(d^2(\mathbf{x}', \mathsf{H}\mathbf{x}) + d^2(\mathbf{x}, \mathsf{H}^{-1}\mathbf{x}') \right), \tag{14}$$

where $\mathbf{a}$ is respectively linked to H and H^{-1} by equations (5) and (16). Applying the principle of quasi-linearity to optimize equation (14) requires to express the non-linear geometric distance d in terms of the bilinear algebraic distance d_a. Using equation (9), we obtain for both distance terms of the sum in equation (14):

$$d^2(\mathbf{x}', \mathsf{H}\mathbf{x}) = w^2_{\mathbf{x}', \mathsf{H}\mathbf{x}} d_a^2(\mathbf{x}', \mathsf{H}\mathbf{x}) \text{ and } d^2(\mathbf{x}, \mathsf{H}^{-1}\mathbf{x}') = w^2_{\mathbf{x}, \mathsf{H}^{-1}\mathbf{x}'} d_a^2(\mathbf{x}, \mathsf{H}^{-1}\mathbf{x}').$$

The problem is now reduced to express linearly the two algebraic distances in terms of the variable $\mathbf{a}$. For the first term, it is rather direct by expanding it according to equation (8), introducing equation (5) and after some minor algebraic manipulations it becomes:

$$d_a^2(\mathbf{x}', \mathsf{H}\mathbf{x}) = ||\mathsf{S}[\mathbf{x}']_\times \mathbf{e}'\mathbf{x}^\mathsf{T}\mathbf{a} + \mathsf{S}[\mathbf{x}']_\times \mathsf{H}_r\mathbf{x}||^2, \tag{15}$$

which is linear in terms of $\mathbf{a}$ when the squared-norm is ignored. For the second term, it is less direct since the linear expression of H^{-1} in terms of $\mathbf{a}$ is not so trivial. This parameterization is the following:

$$\mathsf{H}^{-1} \sim \mathbf{e}\mathbf{a}^\mathsf{T}\mathsf{H}_r^{-1} - (1 + \mathbf{a}^\mathsf{T}\mathbf{e})\mathsf{H}_r^{-1}. \tag{16}$$

The reader can easily check its validity by doing the simple matrix product $(\mathsf{H}_r + \mathbf{e}'\mathbf{a}^\mathsf{T})(\mathbf{e}\mathbf{a}^\mathsf{T}\mathsf{H}_r^{-1} - (1 + \mathbf{a}^\mathsf{T}\mathbf{e})\mathsf{H}_r^{-1}) = \alpha \mathrm{I}_3 \sim \mathrm{I}_3, \alpha \in \mathbb{R}$. Note that $\mathbf{e}' = \mathsf{H}\mathbf{e}$ and not just $\mathbf{e}' \sim \mathsf{H}\mathbf{e}$ is required for this parameterization to be valid. As above, expanding the expression of distance using equation (8) and introducing the parameterization (16) finally yields:

$$d_a^2(\mathbf{x}, \mathsf{H}^{-1}\mathbf{x}') = ||\mathsf{S}[\mathbf{x}]_\times [[\mathsf{H}_r^{-1}\mathbf{x}']_\times \mathbf{e}]_\times \mathbf{a} - \mathsf{S}[\mathbf{x}]_\times \mathsf{H}_r^{-1}\mathbf{x}'||^2 \tag{17}$$

Introducing these two linear developments of distances into equation (14) yields the expression:

$$\begin{aligned} \mathsf{H} \sim \mathsf{H}_r + \mathbf{e}' \Bigg(\underset{\mathbf{a}}{argmin} \sum_{\{\mathbf{x} \leftrightarrow \mathbf{x}'\}} \Big(& \\ & w^2_{\mathbf{x}', \mathsf{H}\mathbf{x}} ||\mathsf{S}[\mathbf{x}']_\times \mathbf{e}'\mathbf{x}^\mathsf{T}\mathbf{a} + \mathsf{S}[\mathbf{x}']_\times \mathsf{H}_r\mathbf{x}||^2 + \\ & w^2_{\mathbf{x}, \mathsf{H}^{-1}\mathbf{x}'} ||\mathsf{S}[\mathbf{x}]_\times [[\mathsf{H}_r^{-1}\mathbf{x}']_\times \mathbf{e}]_\times \mathbf{a} - \mathsf{S}[\mathbf{x}]_\times \mathsf{H}_r^{-1}\mathbf{x}'||^2 \Big) \Bigg)^\mathsf{T}. \end{aligned} \tag{18}$$

This equation can be rewritten in a linear system (one equation per point correspondence) which allows to solve for $\mathbf{a}$ and then deduce H. The corresponding quasi-linear algorithm is summarized in table 2. The stopping criterion is, as for the algorithm *QLin* of table 1, the convergence of weights or residual. This estimator can be used with a minimum of three point correspondences.

Figures 4 and 5 show results obtained using simulated data to compare three estimators: *LinCSym*, *NLinCSym* and *QLinCSym*.

- *LinCSym* is based on a linear optimization of criterion (13) using the algebraic distance d_a instead of d (this is a *DLT*);
- *NLinCSym* is based on an iterative optimization of criterion (13) using Levenberg-Marquardt [13];
- *QLinCSym* is based on the algorithm of table 2.

Let $(\mathbf{e}, \mathbf{e}' = \mathsf{H}_r\mathbf{e}, \mathsf{H}_r)$ be a given epipolar geometry,

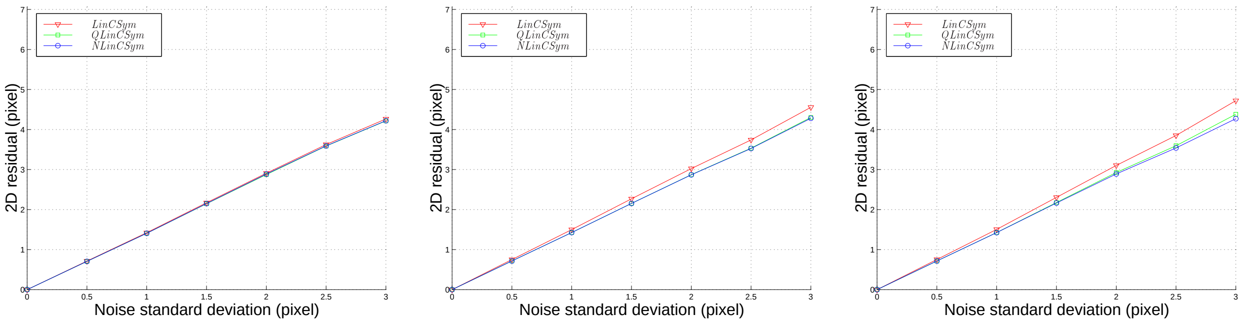
1. *initialization:* $w_{\mathbf{x}',\mathsf{H}\mathbf{x}} = w_{\mathbf{x},\mathsf{H}^{-1}\mathbf{x}'} = 1$;
2. *estimation:* estimate $\mathbf{a}$ then H from equation (18) using standard weighted least squares;
3. *weighting:* use H to update weights $w_{\mathbf{x}',\mathsf{H}\mathbf{x}}$ and $w_{\mathbf{x},\mathsf{H}^{-1}\mathbf{x}'}$ using equation (9);
4. *iteration:* iterate steps 2. and 3. until convergence (see text).

Table 2: Quasi-linear constrained symmetric estimation for plane homography from point correspondences.

The residual measured is the *2D* symmetric geometric transfer *RMS* given by:

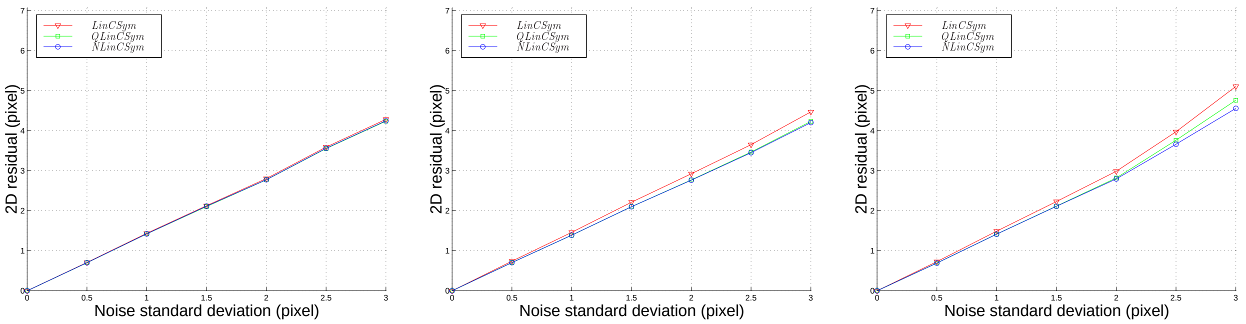
$$E_2(\mathsf{H}) = \sqrt{\frac{1}{n} \sum_{\{\mathbf{x} \leftrightarrow \mathbf{x}'\}} \left(d^2(\mathbf{x}', \mathsf{H}\mathbf{x}) + d^2(\mathbf{x}, \mathsf{H}^{-1}\mathbf{x}') \right)}$$

We have used n=50 points lying on three faces of a *3D* one meter sided cube at a distance ρ of 3, 11 and 20 meters from the cameras. Gaussian centered noise standard deviation varying between 0 and 3 pixels has been added to the image points. We either use the true epipolar geometry or a maximum likehood estimate from all simulated points. The homography H is estimated from the points of one face of the cube.



(a) (b) (c)

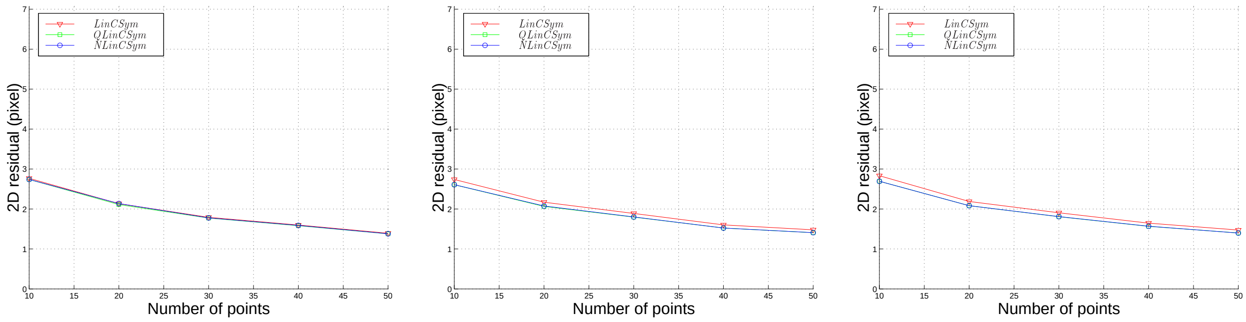
Figure 4: Comparison of the *2D* residual E_2 vs noise standard deviation for the estimators *LinCSym*, *NLinCSym* and *QLinCSym* for n=50 points per plane and 3 planes. The distance plane/cameras ρ is (a): 3, (b): 11 and (c): 20 meters. The true fundamental matrix is used.



(a) (b) (c)

Figure 5: Comparison of the *2D* residual E_2 vs noise standard deviation for the estimators *LinCSym*, *NLinCSym* and *QLinCSym* for n=50 points per plane and 3 planes. The distance plane/cameras ρ is (a): 3, (b): 11 and (c): 20 meters. The maximum likelihood estimate of the fundamental matrix is used.

Figures 4 and 5 show that for all distances, the algorithms tested can stick equally well to the data, both in the case when the true epipolar geometry is used and in the case when an estimated one is used. Figure 6 shows that results are improved for all distances and methods when the number of points increases.



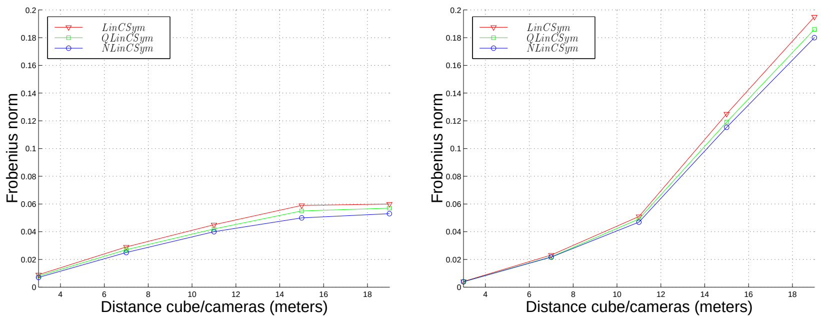
(a) (b) (c)

Figure 6: Comparison of the *2D* residual E_2 vs the number n of points per plane for the estimators *LinCSym*, *NLinCSym* and *QLinCSym* for a 1 pixel standard deviation noise. The distance plane/cameras ρ is (a): 3, (b): 11 and (c): 20 meters. The maximum likelihood estimate of the fundamental matrix is used.

The problem is that these measures only reflect how algorithms stick to data, which is why we use now the previously defined measure of distance $E_{\mathcal{F}}$ between two plane homographies defined above. Figure 7 clearly shows that:

- results are better when the true fundamental matrix is used;
- the quality degrades when the cube moves away from the cameras.

The consequences are the same as described in the case of the unconstrained homography estimation.



(a) (b)

Figure 7: Comparison of the distance $E_{\mathcal{F}}$ between the estimated homography H and the true one $\overline{\mathsf{H}}$ for the estimators *LinCSym*, *NLinCSym* and *QLinCSym* vs the distance plane/cameras ρ for a 1 pixel standard deviation noise, n=50 points per plane and 3 planes. (a): the true fundamental matrix is used and (b): the maximum likelihood estimate is used.

4 Projective structure and calibration from two images

In this section, we aim at extending the minimal parameterization of the epipolar geometry presented in [12, 19] to the case of the complete structure and motion for two views of a piecewise planar scene. This is done via a constrained and minimal (so consistent) parameterization of the fundamental matrix and plane homographies. The principal result is an *MLE* for the parameters of the scene.

First, we investigate the strong link between epipolar geometry and plane homographies from which we deduce a minimal parameterization based on a *reference plane homography*. Next, we retrieve the results of [19]

Point in left image	$\mathbf{x}$
Associated epipolar line in left image	$[\mathbf{e}]_\times \mathbf{x}$
Intersection with $\mathbf{d}$	$[\mathbf{d}]_\times [\mathbf{e}]_\times \mathbf{x}$
Corresponding point on $\mathbf{d}'$ and on the associated epipolar line in the right image	$\widetilde{\mathsf{H}}[\mathbf{d}]_\times [\mathbf{e}]_\times \mathbf{x}$
Parameterization of plane homographies	$(\widetilde{\mathsf{H}}[\mathbf{d}]_\times [\mathbf{e}]_\times + \mathbf{e}'\mathbf{a}^\mathsf{T})\mathbf{x}$

Table 3: Construction of constrained plane homographies. The reference homography H_r can be set by fixing the free inhomogeneous 3-vector $\mathbf{a}$.

in a different way and show how to use them to devise the *MLE* for the epipolar geometry. This estimator is then extended to use the piecewise planar structure of the scene and eventually points that do not lie on any plane. Finally, we show experimental results that validate our approach in various cases.

4.1 Epipolar geometry and plane homographies

Plane homographies and epipolar geometry are strongly linked. This section aims at defining what the minimal algebraic image components of a piecewise planar scene are and what is a possible consistent parameterization of these components.

In [19], the components of the epipolar geometry are defined as the two epipoles $\mathbf{e}$ and $\mathbf{e}'$ and the epipolar transformation which we call $\widetilde{\mathsf{h}}$ (explained below). Each epipole has 2 dof and the epipolar transformation has 3, which yields the well known 7 dof for the epipolar geometry. In the remainder of this paper, we use this parameterization under the notation $\nu = \{\mathbf{e}, \mathbf{e}', \widetilde{\mathsf{h}}\}$.

The question left is now how do we parameterize plane homographies given the epipolar geometry. If, as in [19], we do not parameterize plane homographies directly from ν, this is only possible using a reference homography, let us say H_r, deduced from F using equation (6) and then equation (5) for each plane homography H_i. The drawback of this approach is that the parameterization of plane homographies is not direct, which results in a more complicated expression than below.

On the other hand, if we manage to parameterize this reference homography H_r directly from ν in a simpler way, one can easily compute F using equation (4) ($\mathsf{F} \sim [\mathsf{e}']_\times \mathsf{H}$) and any plane homography as above. This solution, which has not been explored until now, is described in the following. We introduce, for an easy reading, two functions f and h which allow to deduce respectively the fundamental matrix and any plane homography given its 3-vector parameters from ν ($\mathsf{F} \sim f(\nu)$ and $\mathsf{H}_i \sim h(\nu, \mathbf{a}_i)$ where $\mathbf{a}_i$ characterizes H_i).

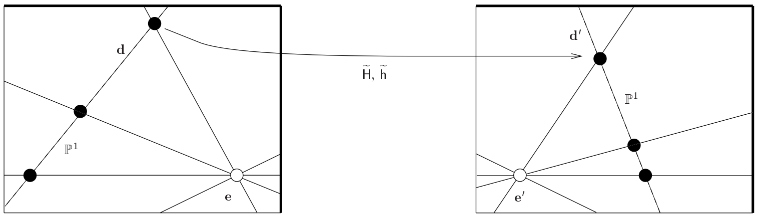


Figure 8: Epipolar pencils ($\mathbb{P}^1$ spaces) formed in each image by epipolar lines are linked by a 2×2 homography $\widetilde{\mathsf{h}}$. Its extension to $\mathbb{P}^2$ is $\widetilde{\mathsf{H}}$.

Let us construct the parameterization. Epipolar lines form a pencil with the epipole as node (see figure 8) which is isomorphic to a projective line. Consequently, epipolar lines are linked between the two images by a 2×2 homography, *the epipolar transformation* which we denote $\widetilde{\mathsf{h}}$. Based on this, we are now going to characterize an homography H_r by considering a point $\mathbf{x}$ in the left image and associating to it a point $\mathbf{x}'$ on the right epipolar line associated to $\mathbf{x}$. The method consists in intersecting the epipolar pencil with a line $\mathbf{d}$ in the left image and $\mathbf{d}'$ in the right one and expressing $\widetilde{\mathsf{h}}$ between $\mathbf{d}$ and $\mathbf{d}'$. This line homography is then extended to a plane homography $\widetilde{\mathsf{H}}$ (not unique, see below for a possible solution), from which we can deduce H_r. This construction is algebraically given in table 3 and illustrated on figure 9. The 3-vector $\mathbf{a}$ can be freely chosen to simplify the expression of H_r in the different cases. Lines $\mathbf{d}$ and $\mathbf{d}'$ are chosen not to contain respectively epipoles $\mathbf{e}$ and $\mathbf{e}'$.

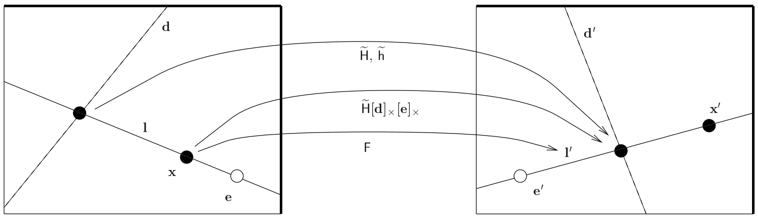


Figure 9: Principle of the parameterization of the epipolar geometry.

The crucial point of the method is the extension of $\widetilde{\mathsf{h}}$ to $\widetilde{\mathsf{H}}$ because its structure depends on $\mathbf{d}$ and $\mathbf{d}'$. Moreover, the finite or infinite nature of the epipoles changes the number of dof of the epipolar geometry. Actually we have to choose a special element of $\mathbf{d}$ and $\mathbf{d}'$ which is non null which yields $3 \times 3 = 9$ possibilities of parameterization or maps and either impose a constraint on the elements of $\widetilde{\mathsf{h}}$ such as $||\widetilde{\mathsf{h}}|| = 1$ or divide by the largest of its four elements, which finally yields $9 \times 4 = 36$ maps. This number has already been mentioned in [19]. Here, we choose to use the constraint $||\widetilde{\mathsf{h}}||^2 = 1$ with 9 maps rather than 36 maps.

For example, we derive here in detail the important case when both epipoles are finite ($\mathbf{e} \sim (e_1, e_1, 1)^\mathsf{T}$ and $\mathbf{e}' \sim (e'_1, e'_2, 1)^\mathsf{T}$). The epipolar geometry has then its full 7 dof. The only lines which never contain $\mathbf{e}$ and $\mathbf{e}'$ are respectively $\mathbf{l}_\infty \sim \mathbf{l}'_\infty \sim (0,0,1)^\mathsf{T}$ which motivates the choice $\mathbf{d} \sim \mathbf{d}' \sim (0,0,1)^\mathsf{T}$. The epipolar transformation $\widetilde{\mathsf{h}}$ is then expressed between the lines at infinity of the images. Let $\mathbf{p}$ and $\mathbf{p}'$ be respectively the intersection of $\mathbf{l}$ and $\mathbf{d}$ and of $\mathbf{l}'$ and $\mathbf{d}'$ where $\mathbf{l}$ and $\mathbf{l}'$ are the left and right epipolar lines associated to the point $\mathbf{x}$. Let us see the impact of these choices on the form of $\widetilde{\mathsf{H}}$. Indeed, $\mathbf{p} \in \mathbf{l}_\infty \Rightarrow \mathbf{p} \sim (p_1, p_2, 0)^\mathsf{T}$ and $\mathbf{p}' \in \mathbf{l}'_\infty \Rightarrow \mathbf{p} \sim (p'_1, p'_2, 0)^\mathsf{T}$ which implies $\begin{pmatrix} p_1 \\ p_2 \end{pmatrix} \sim \widetilde{\mathsf{h}} \begin{pmatrix} p'_1 \\ p'_2 \end{pmatrix}$ and $\widetilde{\mathsf{H}} \sim \begin{pmatrix} \widetilde{\mathsf{h}} & \mathbf{q} \\ \mathbf{0}_2^\mathsf{T} & \end{pmatrix}$ where $\mathbf{q}$ is a 3-vector which can be freely chosen (to simplify expressions). Consequently, $\widetilde{\mathsf{H}}$ is not unique. Here we choose $\mathbf{q} = \mathbf{0}_3$. In this case $\widetilde{\mathsf{H}}$ does not correspond to any physical plane. The reference homography is then deduced starting from table 3:

$$\mathsf{H}_r \sim \widetilde{\mathsf{H}}[\mathbf{l}_\infty]_\times[\mathbf{e}]_\times + \mathbf{e}'\mathbf{a}^\mathsf{T} \sim \begin{pmatrix} -\widetilde{\mathsf{h}} & \widetilde{\mathsf{h}}(e_1, e_2)^\mathsf{T} \\ \mathbf{0}_2^\mathsf{T} & 0 \end{pmatrix} + \mathbf{e}'\mathbf{a}^\mathsf{T}.$$

Remembering that the 3-vector $\mathbf{a}$ can be freely chosen for the reference homography, we set it to $\mathbf{a} = (0,0,1)^\mathsf{T}$ which yields a final expression for H_r that is non singular ($det(\mathsf{H}_r) = -det(\widetilde{\mathsf{h}})$) and affine, thus handled easily (in particular for the estimation of H_r^{-1}):

$$\mathsf{H}_r \sim \begin{pmatrix} -\widetilde{\mathsf{h}} & \widetilde{\mathsf{h}}(e_1, e_2)^\mathsf{T} + (e'_1, e'_2)^\mathsf{T} \\ \mathbf{0}_2^\mathsf{T} & 1 \end{pmatrix}.$$

We have also studied the cases when both epipoles are infinite (which corresponds to the case when the optical center of each camera is in the focal plane of the other) and when one of them is finite (for example, $\mathbf{e}$ is at infinity and $\mathbf{e}'$ is finite). When an epipole is at infinity, it has only one dof. Consequently, to complete the parameterization, one has to assume that it is not on a certain axis of the image frame coordinate system. Here, we assume that the epipole at infinity, $\mathbf{e}$ for example, is not on the vertical axis and so we choose $\mathbf{d} \sim (1,0,0)^\mathsf{T}$ which is feasible because it corresponds to the fact that cameras are not above each other. In practice, if $\mathbf{e}$ lies on or is close to the vertical axis, it is possible to rotate the image in an appropriate way to make the parameterization valid. This choice for $\mathbf{d}$ implies that $\mathbf{e}$ has the form $\mathbf{e} \sim (1, e_2, 0)^\mathsf{T}$. The expressions of H_r are not established in detail here but are summarized in table 4. When both epipoles are at infinity, the epipolar geometry has 1+1+3=5 dof and when one epipole is finite, it has 1+2+3=6 dof. The 36 original maps are then reduced to 9 using the constraint $||\widetilde{\mathsf{h}}||^2=1$. These maps account for the following cases:

1. both epipoles are finite: 1 map;

2. one epipole is at infinity and the other is finite: 4 maps;

3. both epipoles are at infinity: 4 maps.

Now, let us see in more detail for these cases if the number of maps can be reduced:

1. the number of map can not be reduced;

2. the number of maps can be reduced using the fact that the role of the two images can be inversed, which leaves 2 maps. Among those, the map corresponding to the infinite epipole lying on the vertical axis can be removed using an appropriate image rotation (which leaves the residual invariant) yielding 1 map for this case;

3. the number of maps can be reduced to 1 using the same reasoning as above.

Consequently, we have reduced the 36 original maps to 3 by a simple constraint and optional image transformations. The expressions of these 3 maps are given in detail both for H_r and F in tables 4 and 5.

$\mathbf{e}$	$\mathbf{e}'$	dof	ν	H_r
$\not\infty$	$\not\infty$	7	$\widetilde{\mathsf{h}}$, e_1, e_2, e'_1, e'_2	$\begin{pmatrix} -\widetilde{\mathsf{h}} & \widetilde{\mathsf{h}}(e_1,e_2)^\top + (e'_1,e'_2)^\top \\ \mathbf{0}_2^\top & 1 \end{pmatrix}$
$\not\infty$	∞	6	$\widetilde{\mathsf{h}}$, e_1, e_2, e'_2	$\begin{pmatrix} \mathbf{0}_2^\top & 1 \\ -\widetilde{\mathsf{h}} & \widetilde{\mathsf{h}}(e_1,e_2)^\top + (e'_2,0)^\top \end{pmatrix}$
∞	∞	5	$\widetilde{\mathsf{h}}$, e_2, e'_2	$\begin{pmatrix} 1 & \mathbf{0}_2^\top \\ \widetilde{\mathsf{h}}(e_2,0)^\top + (e'_2,0)^\top & -\widetilde{\mathsf{h}} \end{pmatrix}$

Table 4: Parameterization of a reference plane homography H_r in terms of the finite or infinite nature of the epipoles.

There also exist particular two view configurations when the epipolar geometry takes a simplified form. An important part of these cases have been studied in [18]. The general principle of parameterization developed here can easily be applied to these configurations.

4.2 Optimal estimation of epipolar geometry

Papers dealing with the estimation of the epipolar geometry [12, 19] conclude that the optimal way needs a parameterization of the fundamental matrix that enforces the geometric constraint of null determinant and which express the algebraic constraint of scale ambiguity in a convenient way. Such a parameterization is proposed in [19]. We obtain here the same results but in a different way, starting from the parameterization of the reference homography proposed above. The fundamental matrix is retrieved using equation (4) with $\mathsf{H}_r \sim \widetilde{\mathsf{H}}$. Table 5 shows forms of the fundamental matrix in the three cases developed previously for the reference homography. The case when the two epipoles are finite strictly corresponds to the results obtain in [12].

$\mathbf{e}$	$\mathbf{e}'$	dof	ν	F
$\not\infty$	$\not\infty$	7	$\widetilde{\mathsf{h}}$, e_1, e_2, e'_1, e'_2	$\begin{pmatrix} c & d & -ce_1 - de_2 \\ -a & -b & ae_1 + be_2 \\ ae'_2 - ce'_1 & be'_2 - de'_1 & k \end{pmatrix}$ with $k = (ce_1 + de_2)e'_1 - (ae_1 + be_2)e'_2$
$\not\infty$	∞	6	$\widetilde{\mathsf{h}}$, e_1, e_2, e'_2	$\begin{pmatrix} -ce'_2 & -de'_2 & (ce_1 + de_2)e'_2 \\ c & d & -ce_1 - de_2 \\ -a & -b & ae_1 + be_2 \end{pmatrix}$
∞	∞	5	$\widetilde{\mathsf{h}}$, e_2, e'_2	$\begin{pmatrix} ce_2e'_2 & -ce'_2 & -de'_2 \\ -ce_2 & c & d \\ ae_2 & -a & -b \end{pmatrix}$

Table 5: Fundamental matrix parameterization according to the finite or infinite nature of the epipoles $\mathbf{e}$ and $\mathbf{e}'$. a, b, c and d are the coefficients of the epipolar transformation $\widetilde{\mathsf{h}}$.

Now that the parameterization of the fundamental matrix with respect to the finite of infinite nature of the epipoles is set, a step further consists to choose which one of these parameterizations to use in a given situation and when this choice has to be made. In [19], as all models (called maps) have the same number of dof, the author proposes to choose the one that is locally the least singular by maximizing determinants of minors from a jacobian matrix. As this criterion is local, it has to be included in the optimization process which is complicated, from a coding and cost point of view.

A way to enhance such a process might be to choose the appropriate map using a model selection criterion such as *MDL* or *AIC* [16] and not at every step of the optimization process. Indeed, experimental results obtained in [7] show that the *DLT* algorithm to compute the epipolar geometry, also called 8 point algorithm performs well when using a data normalization: the algorithm does not yield gross errors, it is slightly less accurate than a non-linear estimation and so is appropriate to choose a model for the fundamental matrix.

Once the parameterization has been chosen, one can use two kinds of criteria: the reduced and the direct approach [17]. Let us investigate first the reduced approach criterion. We are only looking for ν, the parameters of the epipolar geometry which optimize:

$$\mathsf{F} = \underset{\nu}{argmin} \sum_{\{\mathbf{x} \leftrightarrow \mathbf{x}'\}} \left(d^2(\mathbf{x}', \mathsf{F}\mathbf{x}) + d^2(\mathbf{x}, \mathsf{F}^\mathsf{T}\mathbf{x}') \right), \tag{19}$$

where the previously chosen parameterization is enforced via the constraint $\mathsf{F} \sim f(\nu)$ and where $d(\mathbf{x}, \mathbf{l})$ is the Euclidean distance between the point $\mathbf{x}$ and the line $\mathbf{l}$. On the other hand, the direct approach is optimal in the sense of the maximum likelihood (under the assumption of i.i.d. centered Gaussian noise) and optimizes over ν and point coordinates. The criterion to optimize is:

$$\{\mathsf{F}, \{\widehat{\mathbf{x}} \leftrightarrow \widehat{\mathbf{x}}'\}\} = \underset{\nu, \{\widehat{\mathbf{x}} \leftrightarrow \widehat{\mathbf{x}}'\}}{argmin} \sum_{\{\mathbf{x} \leftrightarrow \mathbf{x}'\}} \left(d^2(\mathbf{x}, \widehat{\mathbf{x}}) + d^2(\mathbf{x}', \widehat{\mathbf{x}}') \right), \tag{20}$$

under the constraints $\mathsf{F} \sim f(\nu)$ and $\widehat{\mathbf{x}}'^\mathsf{T} \mathsf{F} \widehat{\mathbf{x}} = 0$. This last constraint forces corrected points $\{\widehat{\mathbf{x}} \leftrightarrow \widehat{\mathbf{x}}'\}$ to satisfy perfectly the epipolar geometry. This criterion yields the *MLE* for the epipolar geometry and a general scene structure. The final algorithm is summarized in table 6.

1. *initialisation:* estimate F with the normalized 8 points algorithm [7];
2. *parameterization choice:* choose the appropriate parameterization (table 5) according to the nature of the epipoles $\mathbf{e}$ et $\mathbf{e}'$ (see text);
3. *first refinement (optional):* optimize the criterion (19) using a non-linear optimization method such as Levenberg-Marquardt [13];
4. *maximum likelihood:* optimize using the same technique the cost function (20) to obtain the maximum likelihood estimate (see text).

Table 6: Epipolar geometry parameterization and parameter estimation algorithm.

4.3 Optimal and consistent estimation of epipolar geometry and plane homographies

In this section, we investigate deeper how to optimally estimate the complete projective structure and motion for a piecewise planar scene. In other words, we extend the approach presented in the previous section to make use of the a priori known structure of the scene. We assume the following to be known:

- a suitable parameterization of the epipolar geometry and its parameters;
- a partition of point correspondences into coplanar groups (and individual points not belonging to any such group).

The structure of a planar scene is represented here by the equations of each plane of the scene. The set of parameters of these equations is denoted $\mathbf{A} = \{\mathbf{a}_1 \ldots \mathbf{a}_m\}$ where m is the number of modeled scene planes. These parameters are related to the corresponding plane homographies via equation (5). As above, we use the reduced and direct approach but this time we optimize over the set of variables $\nu \cup \mathbf{A}$, i.e. the parameters of the epipolar geometry (motion) and of the different planes of the scene (structure). The criterion for the reduced approach is then:

$$\{\mathsf{F}, \mathsf{H}_1, \ldots, \mathsf{H}_m\} = \underset{\nu \cup \mathbf{A}}{argmin} \sum_{j=1}^{m} \sum_{\{(\mathbf{x} \leftrightarrow \mathbf{x}') \in \pi_j\}} \left(d^2(\mathbf{x}', \mathsf{H}_j \mathbf{x}) + d^2(\mathbf{x}, \mathsf{H}_j^{-1} \mathbf{x}') \right), \tag{21}$$

under the constraints $\mathsf{F} \sim f(\nu)$ and $\mathsf{H}_j \sim h(\nu, \mathbf{a}_j)$ which enforce respectively the previously chosen parameterization and the relationship of global consistence, i.e. compatibility of each homography with the epipolar geometry. The notation π_j denotes the plane corresponding to equation $\mathbf{a}_j$. It is important to remark that the epipolar geometry is implicitly estimated via plane homographies and not directly using equation (1). Optimization of equation (21) requires the estimation of H_j^{-1} at each step of the process. This is a costly operation which is improved using the fact that each H_j is parameterized in a particular way and using equation (16). This equation requires the knowledge of H_r^{-1} which can be easily estimated taking into account its particular form established previously (e.g. its affine form in the case of finite epipoles).

The direct approach consists in optimizing the following criterion when point coordinates are involved and constrained to lie on the plane they correspond to:

$$\{\mathsf{F}, \mathsf{H}_1, \ldots, \mathsf{H}_m, \{\widehat{\mathbf{x}}\}\} = \underset{\nu \cup \mathbf{A} \cup \{\widehat{\mathbf{x}}\}}{argmin} \sum_{j=1}^{m} \sum_{\{(\mathbf{x} \leftrightarrow \mathbf{x}') \in \pi_j\}} \left(d^2(\mathbf{x}, \widehat{\mathbf{x}}) + d^2(\mathbf{x}', \mathsf{H}_j \widehat{\mathbf{x}}) \right), \tag{22}$$

under the constraints $\mathsf{F} \sim f(\nu)$ and $\mathsf{H}_j \sim h(\nu, \mathbf{a}_j)$. This criterion yields the *MLE* for the epipolar geometry, the plane equations and the points of these planes. Note that this is true only in the case when each point belongs to only one plane (e.g. if $\mathbf{x} \in \pi_1$ and $\mathbf{x} \in \pi_2$, we can not guarantee that $\mathsf{H}_1 \widehat{\mathbf{x}} \sim \mathsf{H}_2 \widehat{\mathbf{x}}$ which is in contradiction with optimality). It is also possible to make points $\{(\mathbf{y} \leftrightarrow \mathbf{y}') \notin \pi\}$ that do not belong to any plane contribute to the estimation by including them in the following way:

$$\begin{aligned} \{\mathsf{F}, \mathsf{H}_1, \ldots, \mathsf{H}_m, \{\widehat{\mathbf{x}}\}, \{\widehat{\mathbf{y}} \leftrightarrow \widehat{\mathbf{y}}'\}\} = \underset{\nu, \mathbf{a}_1, \ldots, \mathbf{a}_m, \{\widehat{\mathbf{x}}\}, \{(\widehat{\mathbf{y}} \leftrightarrow \widehat{\mathbf{y}}') \notin \pi\}}{argmin} \Bigg(\\ \sum_{j=1}^{m} \sum_{\{(\mathbf{x} \leftrightarrow \mathbf{x}') \in \pi_j\}} \left(d^2(\mathbf{x}, \widehat{\mathbf{x}}) + d^2(\mathbf{x}', \mathsf{H}_j \widehat{\mathbf{x}}) \right) + \\ \sum_{\{(\mathbf{y} \leftrightarrow \mathbf{y}') \notin \pi\}} \left(d^2(\mathbf{y}, \widehat{\mathbf{y}}) + d^2(\mathbf{y}', \widehat{\mathbf{y}}') \right) \Bigg), \end{aligned} \tag{23}$$

using the same constraints as for the optimization of criterion (22) plus $\widehat{\mathbf{y}}'^{\mathsf{T}} \mathsf{F} \widehat{\mathbf{y}} = 0$ to ensure that points that do not lie on any plane fully satisfy the epipolar geometry. The main two terms of equation (23) can be weigthed depending on the situation (e.g. the number of planes and points). The global algorithm is summarized in table 7.

1. *initialization* F*:* use algorithm 6 to obtain a suitable parameterization of the epipolar geometry and its parameters ν;
2. *initialization* $\mathsf{H}_1 \ldots \mathsf{H}_m$*:* estimate parameters $\mathbf{a}_1, \ldots, \mathbf{a}_m$ of plane homographies $\mathsf{H}_1, \ldots, \mathsf{H}_m$ using a constrained linear method (see previous section);
3. *first refinement (optional):* optimize the criterion (21) using a non-linear optimization method such as Levenberg-Marquardt [13];
4. *maximum likelihood:* optimize using the same technique the cost function (22) or (23) to obtain the maximum likelihood estimate (see text).

Table 7: Optimal coherent structure and motion estimation algorithm.

4.4 Experimental results

We have used simulated data to validate our approach. The test bench is the same as in the previous section. The residual measured is the *3D* non symmetric *RMS* between the estimated projective reconstruction $\{\mathbf{X}\}$ and the true Euclidean one $\{\overline{\mathbf{X}}\}$:

$$E_3 = \sqrt{\frac{1}{n} \sum_{\{\mathbf{X} \leftrightarrow \overline{\mathbf{X}}\}} d^2(\mathsf{H}_3 \mathbf{X}, \overline{\mathbf{X}})},$$

where H_3 is a *3D* homography estimated via a non-linear minimization of E_3 using the Levenberg-Marquardt method [13].

The estimators compared can be divided in two sets. First, the epipolar geometry based estimators (referred to as *Methods F*):

- *FLin+BA*, based on a normalized 8 point algorithm [7] to estimate the fundamental matrix and an optimal bundle adjustement [9] that correct points so that they satisfy exactly this epipolar geometry;
- *FML*, based on the algorithm of table 6;
- *trueF+BA*, based on an optimal bundle adjustement [7] that correct points so that they satisfy exactly the true epipolar geometry.

Then, the plane homography based estimators (referred to as *Methods H*):

- *HiML+FML*, based on a maximum likelihood estimation of plane homographies [10] to ensure a first correction of point correspondences and then a maximum likelihood estimation of the epipolar geometry (algorithm of table 6) to complete the reconstruction;
- *consHiML*, based on the algorithm of table 7;
- *trueHi+BA*, based on an optimal bundle adjustement of points [10] using the true plane homographies.

We have carried out two sets of experiments, where *3D* points lie in:

1. perfectly coplanar groups;
2. near coplanar groups.

Let us comment on the first set of experiments. Figure 10 shows results obtained when varying noise standard deviation between 0 and 3 pixels (for ρ=3, 11 and 20 meters and using n=50 points per plane) and figure 11 when varying n from 10 to 50 points (for ρ=3, 11 and 20 meters and using a 1 pixel standard deviation noise).

The results can be divided in two sets: the set *Methods F* and the *Methods H* one. Figure 10 and 11 show that for all distances, levels of noise and numbers of points, *Methods H* perform better (the residual is at least two times lower) than *Methods F*. If we investigate the first set of methods *Methods F* in more detail, we can conclude that:

- *trueF+BA* performs better than *FML* which itself performs better than *FLin+BA*;
- these methods diverge when distance and noise standard deviation increase too much.

For the second set *Methods H*, we can say that:

- *trueHi+BA* performs better than *consHiML* which itself performs better than *HiML+FML*;
- these methods do not degrade too much when distance and noise standard deviation increase.

Consequently, we can conclude that in the case of perfectly coplanar points, and taking into account that in real cases, methods *trueF+BA* and *trueHi+BA* are not always available (they require a priori knowledge about the camera configuration or the scene structure), *Methods F* fail when a certain amount of noise and a certain distance ρ are reached while *Methods H* do not. Moreover, when the former do not fail, they perform worse than the latter.

Now, let us investigate the second set of experiments, when *3D* points are only nearly coplanar. Figure 12 shows results obtained when varying planes unflatness, i.e. when *3D* points do not lie perfectly on planes: coplanar points are offset vertically from their planes by a random distance (Gaussian noise with standard

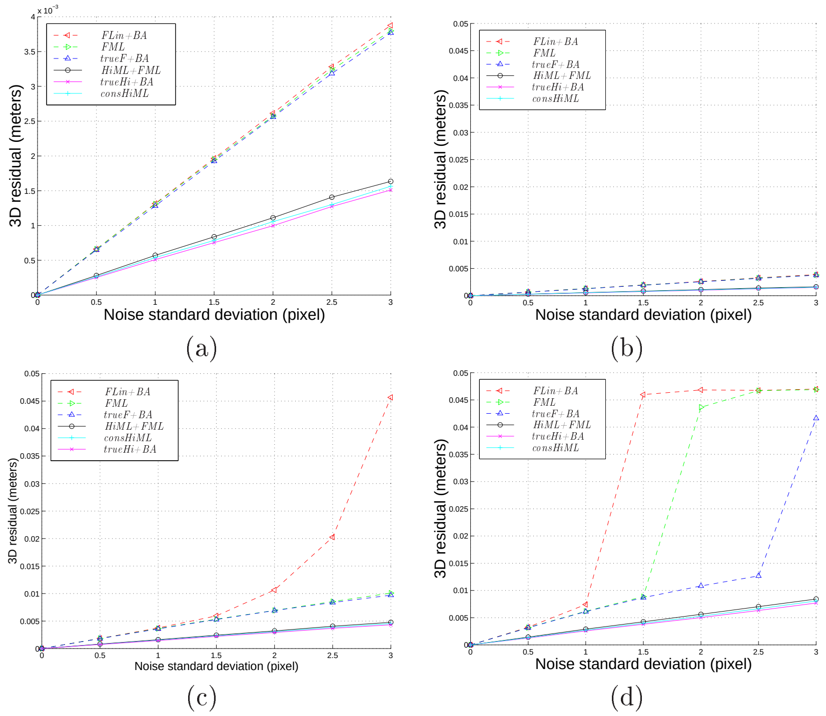


Figure 10: Comparison of the *3D* residual E_3 for the estimators *FLin+BA*, *FML*, *trueF+BA*, *HiML+FML*, *consHiML* and *trueHi+BA* vs noise standard deviation for n=50 points per plane and 3 planes. The distance cube/cameras ρ is (a) and (b): 3, (c): 11 and (d): 20 meters. Simulated *3D* points are perfectly coplanar. Note that (a) and (b) are identical with different scales.

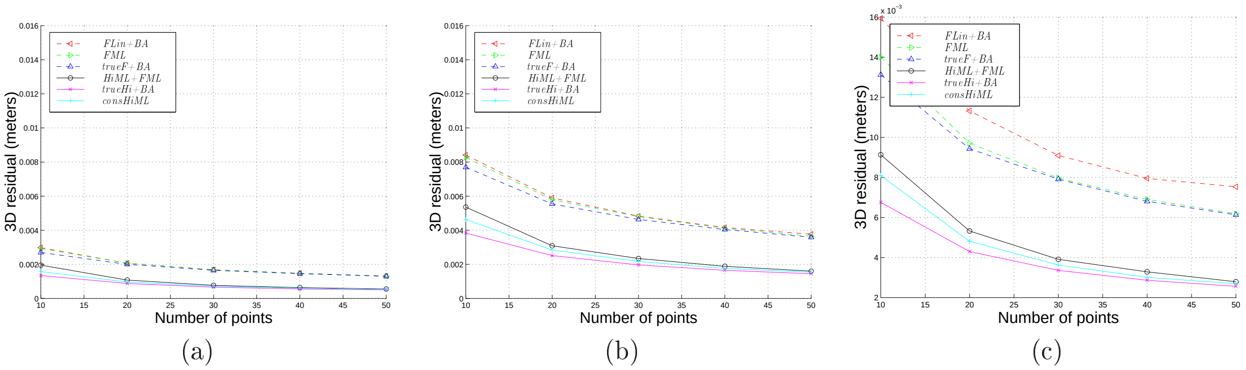
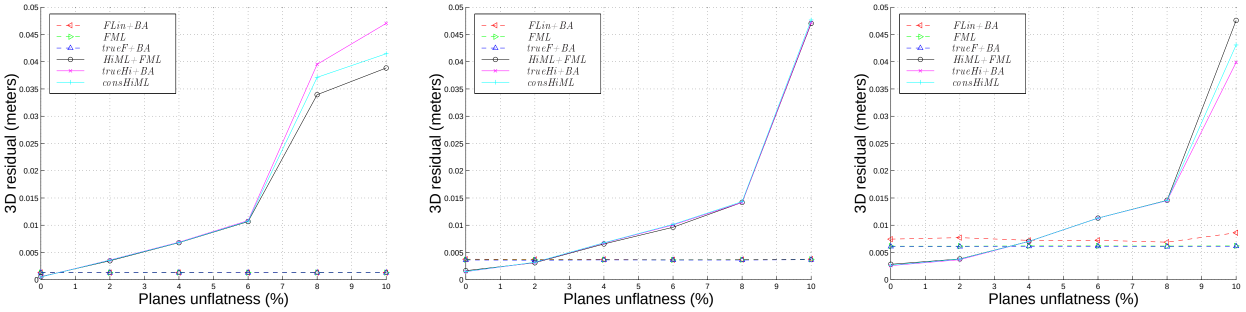
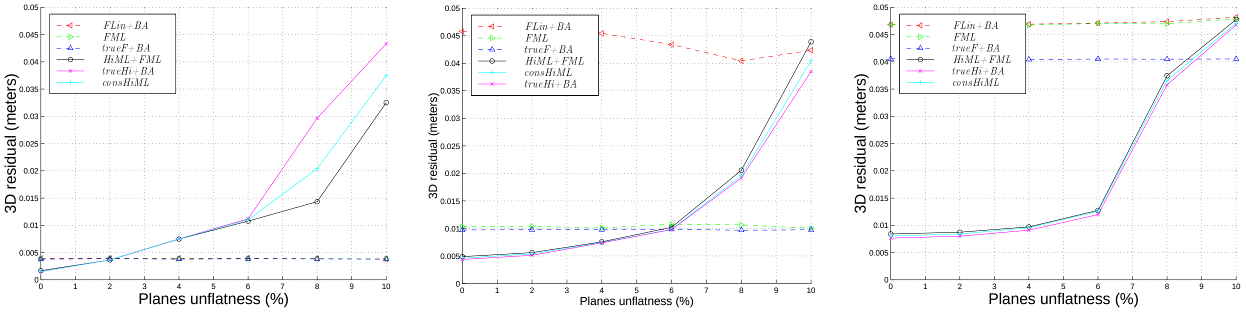


Figure 11: Comparison of the *3D* residual E_3 for the estimators *FLin+BA*, *FML*, *trueF+BA*, *HiML+FML*, *consHiML* and *trueHi+BA* vs the number n of point correspondences per plane for a 1 pixel standard deviation noise. The distance cube/cameras ρ is (a): 3, (b): 11 and (c): 20 meters. Simulated *3D* points are perfectly coplanar.



(a) (b) (c)

Figure 12: Comparison of the *3D* residual E_3 for the estimators *FLin+BA*, *FML*, *trueF+BA*, *HiML+FML*, *consHiML* and *trueHi+BA* and *trueHi+BA* vs the unflatness of *3D* planes for n=50 points per plane, 3 planes and a 1 pixel standard deviation noise. The distance cube/cameras ρ is (a) and (b): 3, (c): 11 and (d): 20 meters.



(a) (b) (c)

Figure 13: Comparison of the *3D* residual E_3 for the estimators *FLin+BA*, *FML*, *trueF+BA*, *HiML+FML*, *consHiML* and *trueHi+BA* vs the unflatness of *3D* planes for n=50 points per plane, 3 planes and a 3 pixel standard deviation noise. The distance cube/cameras ρ is (a) and (b): 3, (c): 11 and (d): 20 meters.

	3 m.	11 m.	20 m.
1 pixel	0.5%	2%	4%
3 pixels	2%	6%	9%

Table 8: Ratio ε of the plane unflatness and the total object size vs the distance cube/cameras ρ and noise standard deviation.

deviation between 0 and 0.1 meters, for ρ=3, 11 and 20 meters and using n=50 points per plane). A 1 pixel standard deviation noise is added to image points. Figure 13 shows the results using the same situation but a 3 pixel standard deviation noise.

Once again the behaviour of these methods can be divided in the two same sets as above. Let us denote ε the ratio between the planes unflatness and the size of the simulated cube where *Methods H* begin to perform worse than *Methods F*, e.g. for figure 12 (a), ε=0.5%.

Figures 12 and 13 allow to establish table 8 which clearly shows that the less stable the configuration is (a high standard deviation noise and/or a high distance cube/cameras ρ), the higher ε is. This means that there are a lot of cases when *3D* points are not perfectly coplanar and when the plane based method will perform better than any method based only on points. The values of ε given in table 8 represent relatively large variations in practice. A great majority of real planes are below such variations.

5 Plane detection and scene reconstruction

In this section, we present a complete reconstruction system for two views of a piecewise planar scene. The different steps of this system are:

1. detect and match interest points;
2. detect the projections of planes;
3. estimate structure and motion parameters using the algorithms presented in the previous section to projectively reconstruct the scene.

The first step is done via the estimation of the epipolar geometry using the method presented in [19]. The second step, detection of the planar structures, is achieved using two kinds of criterion:

- an algebraic one based on plane homographies;
- a topological one based on proximity of coplanar points.

None of these criteria really reflects physical planes, that's why we have to merge them to obtain a reasonable result. The algebraic detection of planes is based on a recursive robust estimation of plane homographies using the algorithm *RANSAC* [4] and the constrained linear estimator for plane homographies presented in the previous section. We estimate the dominant plane and then remove points that lie on it to recursively estimate the next dominant plane. The topological detection is based on an iterative grouping of each coplanar point set according to the proximity of the points (as in [2] for a dominant *3D* homography estimation). Once this has been done, the third and last step can be started. It consists in applying the algorithm from the previous section to the data obtained with the two first steps. The result is a projective *3D* reconstruction of the scene. The user has just to give the borders of each plane in one of the two images to complete the texture mapping of the reconstruction. An update to a metric reference frame can finally be done using autocalibration or by asking the user to give the Euclidean coordinates points.

We present here two reconstructions of piecewise planar scenes. The first one, designated by *Cubic Scene* deals with a well arranged interior scene (figure 14) and the second one, designated by *Game System* is a real building snapshot (figure 15).

More than 100 interest points per image have been detected and automatically matched while robustly estimating the epipolar geometry [19] for *Cubic Scene*. Using plane homographies and proximity of coplanar points, 5 planar structures have been detected and manually completed to 7 (see figure 16 (a) and (b) respectively).

The automatic plane detection is achieved via a *RANSAC*-like [5] scheme. This robustly find the parameters of the successive dominant plane homographies. The loop is ended when less than 3 matches are compatible with the dominant plane.

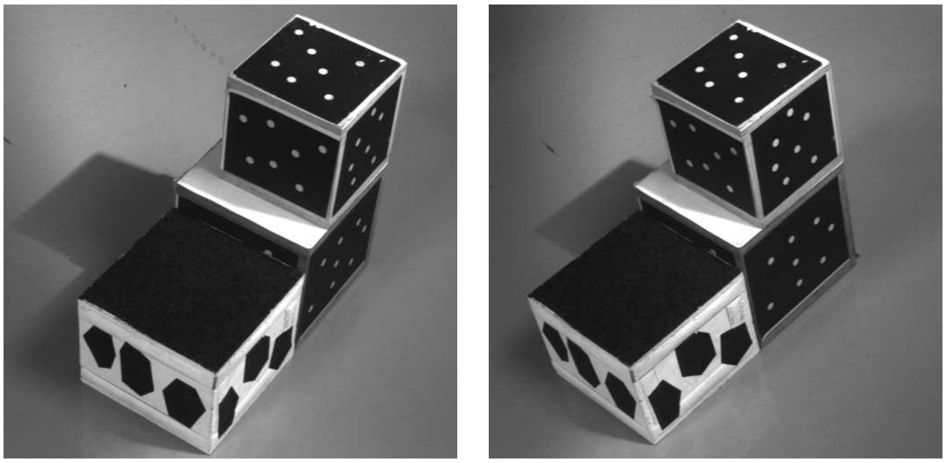


Figure 14: The *Cubic Scene* stereo image pair.

Figure 15: The *Game System* stereo image pair.

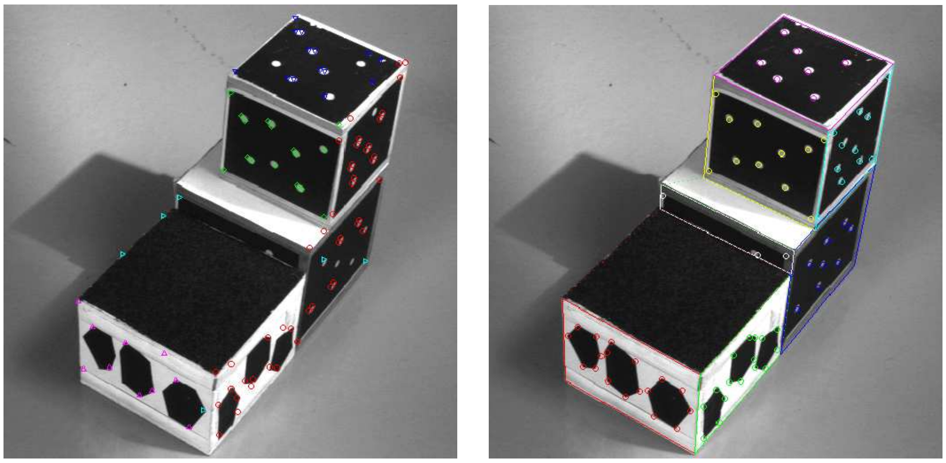


Figure 16: Automatically detected (a) and manually completed (b) planar structures.

An optimal projective reconstruction, see figure 17, has then been obtained using previously presented algorithms.

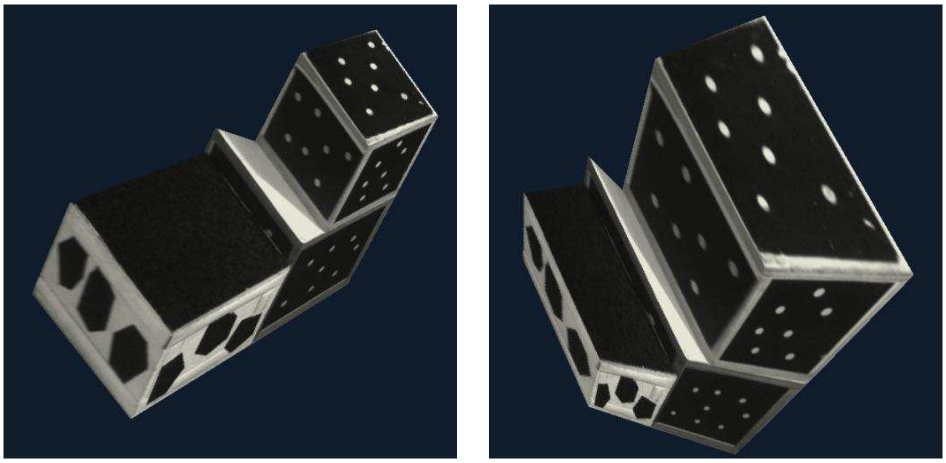


Figure 17: Textured rendering of the recovered projective model for the *Cubic Scene* stereo image pair from different points of view.

This reconstruction has been upgraded to a metric equivalent, see figure 18, from which we have measured some significant metric values such as equal sides length ratios and angles. Figure 19 clearly shows that these values are very close to the expected ones. Finally, figure 20 shows an example of augmented reality by adding the Euclidean model recovered to a virtual world.

For the second image pair *Game System*, the point correspondences and epipolar geometry have been estimated using the technique proposed in [19]. The automatic clustering into coplanar groups detects then only the 2 main planes of the scene (i.e. the front and side of the building) and the rest of the modeled planes have been established manually. We have then performed a projective and a metric equivalent reconstruction, for which images from different points of view are visible in figure 21 and 22 respectively.

6 Discussion

In this paper, we describe a method for estimating the complete optimal structure and motion from two uncalibrated views of a piecewise planar scene. The resulting geometric structures are represented on the image level by a fundamental matrix and a collection of plane homographies (equivalent to *3D* planes) and points on planes. All these structures have been made algebraically consistents without using a final bundle adjustement of the reconstruction. Such a process has been used only to compute the maximum likelihood estimate of structure and motion in a final stage.

We summarize the conclusions for the three main topics we have delt with.

First, the epipolar geometry constrained estimation of a plane homography. The propose method follows the reduced approach thanks to a novel formulation of the inverse homography in the constrained case. The optimization is quasi-linearly conducted. Experimental results show that:

- the resulting homography is better when the estimation is constrained;
- the quasi-linear optimization can achieve in this case performances that are very close to a non-linear one.

Second, the optimal triangulation process. We reduce the 36 original maps for the parameterization of the fundamental matrix to 3 and use them to devise an optimal estimator. A consistent formulation for plane homographies allows then to compute the optimal structure and motion of the scene. Experimental results show that:

- plane based methods are clearly superior to methods that only reconstruct the individual points;
- the new consistent formulation enhances the accuracy of the reconstruction compared to an inconsistent one;

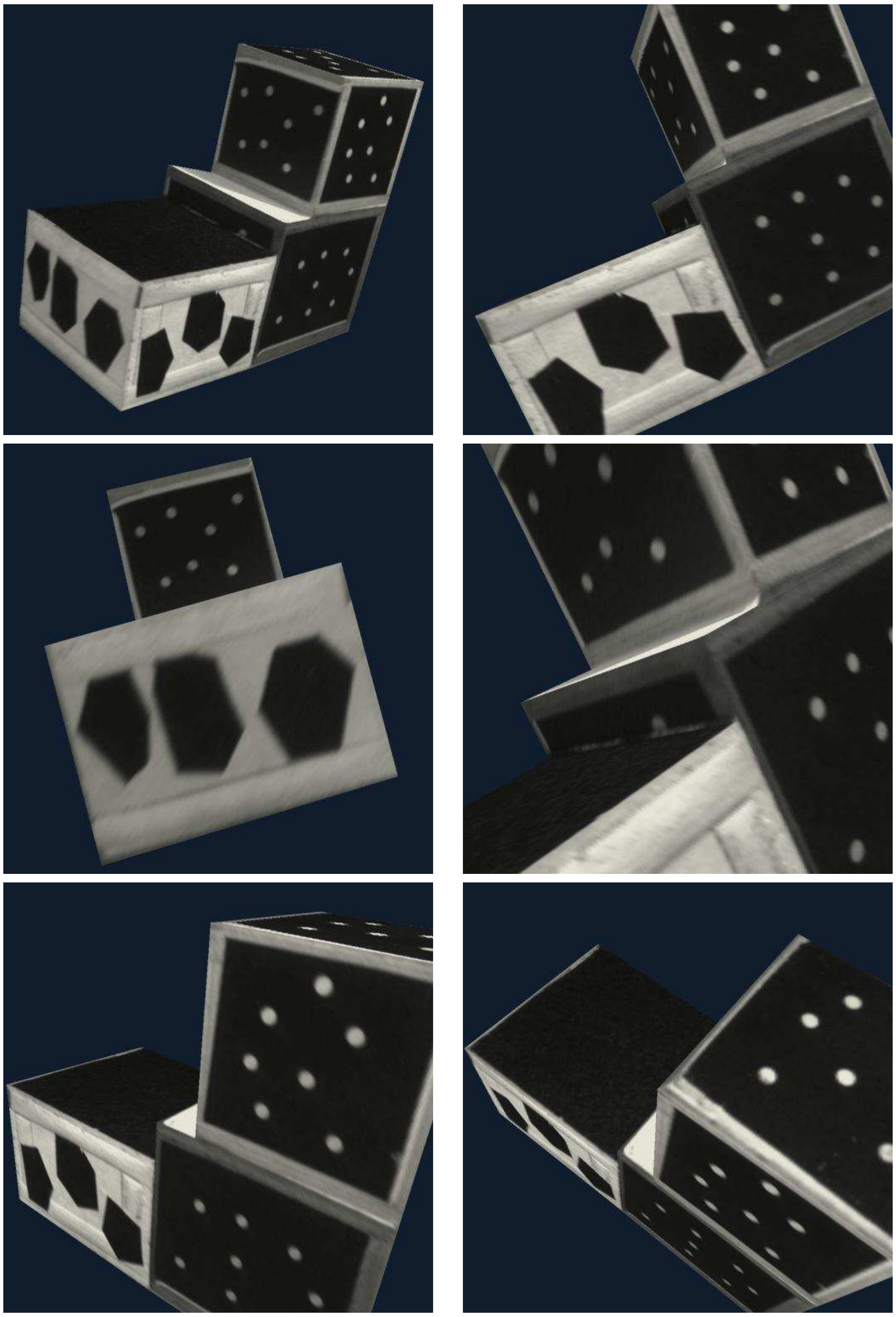
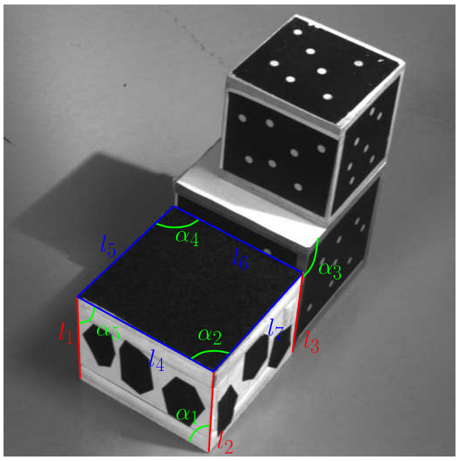


Figure 18: Textured rendering of the recovered metric model for the *Cubic Scene* stereo image pair from different points of view.



l_1/l_2	l_1/l_3	l_4/l_5	l_4/l_6	l_4/l_7
1.0280	0.9655	1.0400	1.0297	1.0232

$2\alpha_1/\pi$	$2\alpha_2/\pi$	$2\alpha_3/\pi$	$2\alpha_4/\pi$	$2\alpha_5/\pi$
0.9743	0.9816	0.9964	0.9857	0.9948

Figure 19: Metric measures on the Euclidean reconstruction from the *Cubic Scene* stereo image pair.

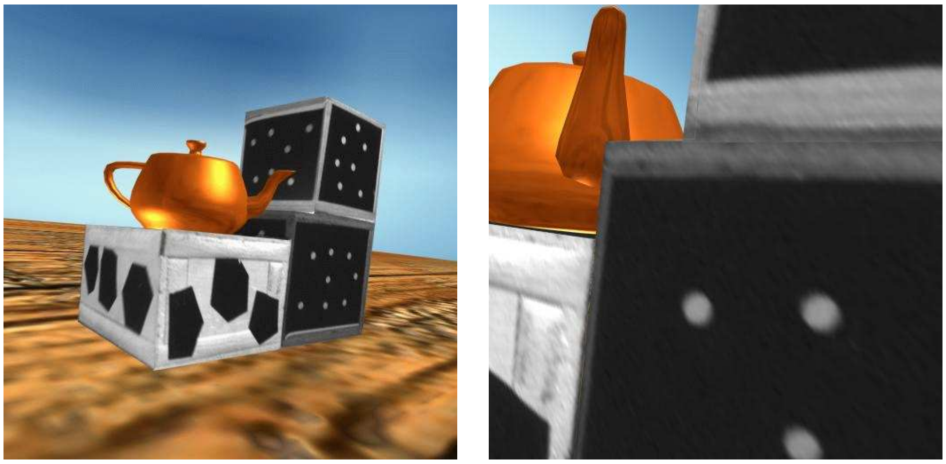


Figure 20: Textured rendering from different points of view of the recovered metric model for the *Cubic Scene* stereo image pair added to a virtual world.

Figure 21: Textured rendering of the recovered projective model for the *Game System* stereo image pair from different points of view.

Figure 22: Textured rendering of the recovered metric model for the *Game System* stereo image pair from different points of view.

- plane based methods achieve good performances in a number of situations where the observed scene is not perfectly piecewise planar.

Third, the image level plane detection. Experimental results show that it is very difficult to devise an image level criterion corresponding to a physical plane. For the moment, the system can only be semi-automatic, that is to say that an intervention of the user in the process of plane detection might be necessary in many situations.

Our work perspectives are three-fold. We plan to:

- extend the approach to more than one stereo image pair: this implies to complete the consistent representation of structure and motion;
- devise a quasi-linear process for the bundle adjustment of points on planes: indeed, this is currently achieved using a non-linear optimization and as said in this paper, there are some reasons to prefer the quasi-linearity;
- enhance the plane detection: current results of such detection systems can be greatly robustified using more than one image pair.

References

[1] D. Demirdjian and R. Horaud. Motion-egomotion discrimination and motion segmentation from image-pair streams. *Computer Vision and Image Understanding*, 78(1):53–68, April 2000.

[2] D. Demirdjian, A. Zisserman, and R. Horaud. Stereo autocalibration from one plane. In *Proceedings of the 6th European Conference on Computer Vision, Dublin, Ireland*, 2000.

[3] O. Faugeras and F. Lustman. Motion and structure from motion in a piecewise planar environment. In *International Journal of Computer Vision*, 1988.

[4] M.A. Fischler and R.C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis. Technical report, SRI International, Menlo Park, CA, 1980.

[5] M.A. Fischler and R.C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Graphics and Image Processing*, 24(6):381 – 395, June 1981.

[6] R. Hartley. Extraction of focal lengths from the fundamental matrix. Technical report, G.E. CRD, Schenectady, NY, 1993.

[7] R. Hartley. In defence of the 8-point algorithm. In *Proceedings of the 5th International Conference on Computer Vision, Cambridge, Massachusetts, USA*, pages 1064–1070, June 1995.

[8] R. Hartley. Minimizing algebraic error. *Proceedings of the 6th International Conference on Computer Vision, Bombay, India*, 1998.

[9] R. Hartley and P. Sturm. Triangulation. In *Proceedings of* ARPA *Image Understanding Workshop, Monterey, California, USA*, pages 957–966, November 1994.

[10] R.I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision.* Cambridge University Press, June 2000.

[11] Q.T. Luong. *Matrice fondamentale et autocalibration en vision par ordinateur.* Thèse de doctorat, Université de Paris-Sud, Orsay, France, December 1992.

[12] Q.T. Luong and O. Faugeras. The fundamental matrix: Theory, algorithms and stability analysis. *International Journal of Computer Vision*, 17(1):43–76, 1996.

[13] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. *Numerical Recipes in C - The Art of Scientific Computing.* Cambridge University Press, 2nd edition, 1992.

[14] R. Szeliski and P.H.S. Torr. Geometrically constrained structure from motion : Points on planes. In *European Workshop on 3D Structure from Multiple Images of Large Scale Environments*, June 1998.

[15] Jean-Philippe Tarel and Jean-Marc Vézien. A generic approach for planar patches stereo reconstruction. In *Proceedings of the Scandinavian Conference on Image Analysis*, pages 1061–1070, Uppsala, Sweden, 1995. http://www-rocq.inria.fr/syntim/textes/scia95-fra.html.

[16] P.H.S. Torr. An assessment of information criteria for motion model selection. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Puerto Rico, USA*, pages 47–52, June 1997.

[17] B. Triggs. Optimal estimation of matching constraints. In *3D Structure from Multiple Images of Large-scale Environments SMILE'98.* Springer Verlag, 1998.

[18] T. Viéville and D. Lingrand. Using singular displacements for uncalibrated monocular visual systems. In B. Buxton and R. Cipolla, editors, *Proceedings of the 4th European Conference on Computer Vision, Cambridge, England*, volume 1065 of *Lecture Notes in Computer Science*, pages 207–216. Springer-Verlag, April 1996.

[19] Z. Zhang. Determining the epipolar geometry and its uncertainty: A review. Technical Report RR 2927, INRIA, July 1996.

A Implementation oriented formulation of main results

In this appendix, we give some implementation oriented mathematical developments which can help the interested reader for an easy coding of some techniques presented above.

A.1 Epipolar geometry constrained plane homography estimation

In this section, we give a convenient way for the implementation of the estimators *LinCSym* and *QLinCSym*. These two algorithms are based on equation (14). In order to be optimized using the quasi-linear principle, this equation needs the development of two others, equations (15) and (17). Let us see how this can be done for a point correspondence ($\mathbf{x} \leftrightarrow \mathbf{x}'$). A good way is to successively compute the following:

- $g_1 = x'_2 e'_3 - e'_2$;
- $g_2 = e'_1 - x'_1 e'_3$;
- $l_i = \mathbf{h}_i^\mathsf{T} \mathbf{x}$;

where $\mathbf{h}_i^\mathsf{T}$ is the i^{th} row of H_i and $\mathbf{e}' \sim (e'_1, e'_2, e'_3)^\mathsf{T}$ is the right epipole, and form the following linear system:

$$\begin{pmatrix} g_1 x_1 & g_1 x_2 & g_1 \\ g_2 x_1 & g_2 x_2 & g_2 \end{pmatrix} \mathbf{a} = \begin{pmatrix} l_2 - x'_2 l_3 \\ x'_1 l_3 - l_1 \end{pmatrix},$$

which is equivalent to the first main term in the sum of equation (14).

A step further consists in computing:

- $l'_i = \mathbf{h}_i^{-\mathsf{T}} \mathbf{x}'$;
- $l''_1 = l'_2 e_3 - l'_3 e_2$;
- $l''_2 = l'_3 e_1 - l'_1 e_3$;
- $l''_3 = l'_1 e_2 - l'_2 e_1$.

where $\mathbf{h}_i^{-\mathsf{T}}$ is the i^{th} row of $\mathsf{H}_i^{-\mathsf{T}}$ and $\mathbf{e} \sim (e_1, e_2, e_3)^\mathsf{T}$ is the left epipole, and form the following linear system:

$$\begin{pmatrix} -l''_3 - x_2 l''_2 & x_2 l''_1 & l''_1 \\ x_1 l''_2 & -x_1 l''_1 - l''_3 & l''_2 \end{pmatrix} \mathbf{a} = \begin{pmatrix} x_2 l'_3 - l'_2 \\ l'_1 - x_1 l'_3 \end{pmatrix}.$$

A.2 Estimate the parameters of the epipolar geometry given the fundamental matrix and an appropriate parameterization

In this section, we give a simple way to estimate the parameters ν of the epipolar geometry given the fundamental matrix F in the case when both epipoles are finite studied in detail below. In this case, denoting by F_{ij} the elements of the given F, the parameters $\nu = \{\mathbf{e}, \mathbf{e}', \widetilde{\mathsf{h}}\} = \{e_1, e_2, e'_1, e'_2, a, b, c, d\}$ under the constraint $a^2 + b^2 + c^2 + d^2 = 1$ can be recovered using the following steps:

- $a = -\mathsf{F}_{21}$;
- $b = -\mathsf{F}_{22}$;
- $c = \mathsf{F}_{11}$;
- $d = \mathsf{F}_{12}$;
- $t = \frac{1}{ad - bc}$;
- $e_1 = t(b\mathsf{F}_{13} + d\mathsf{F}_{23})$;
- $e_2 = -t(a\mathsf{F}_{13} + c\mathsf{F}_{23})$;
- $e'_1 = t(b\mathsf{F}_{31} - a\mathsf{F}_{32})$;
- $e'_2 = t(d\mathsf{F}_{31} - c\mathsf{F}_{32})$.

A.3 Estimate the inverse of the reference homography

The reference homography H_r, as well as its inverse, H_r^{-1}, is needed in many processes in this work (equation (16) for the estimators *LinCSym* and *QLinCSym*, equation (14) for the estimator *NLinCSym* (see text in the corresponding section) etc.). According to the parameterization of plane homographies established above, H_r can take a particular form, e.g. an affine in the case of finite epipoles. We give here the closed-form expression of H_r^{-1} in this case, in terms of the parameters ν:

$$\mathsf{H}_r^{-1} \sim \begin{pmatrix} -\widetilde{\mathsf{h}}^{-1} & \widetilde{\mathsf{h}}^{-1}(e_1', e_2')^\mathsf{T} + (e_1, e_2)^\mathsf{T} \\ \mathbf{0}_2^\mathsf{T} & 1 \end{pmatrix}.$$

Unité de recherche INRIA Rhône-Alpes
655, avenue de l'Europe - 38330 Montbonnot-St-Martin (France)

Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)
Unité de recherche INRIA Sophia Antipolis : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399