HAL
open science

Structure and Motion From Two Uncalibrated Views Using Points on Planes

Adrien Bartoli, Peter Sturm, Radu Horaud

To cite this version:

Adrien Bartoli, Peter Sturm, Radu Horaud. Structure and Motion From Two Uncalibrated Views Using Points on Planes. Third International Conference on 3-D Digital Imaging and Modeling, May 2001, Quebec City, Canada. pp.83-90, 10.1109/IM.2001.924410 . hal-00094781

HAL Id: hal-00094781

https://hal.science/hal-00094781v1

Submitted on 26 May 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

3DIM'01 - THIRD INTERNATIONAL CONFERENCE ON 3D DIGITAL IMAGING AND MODELING, QUEBEC CITY, CANADA, PP. 83-90, JUNE 2001.

Structure and Motion from Two Uncalibrated Views Using Points on Planes

Adrien Bartoli Peter Sturm Radu Horaud

INRIA Rhône-Alpes, 655, Avenue de l'Europe
38334 Saint Ismier Cedex, France. *first.last@inria.fr*

Abstract

This paper is about the problem of structure and motion recovery from two views of a rigid scene. Especially, we deal with the case of scenes containing planes, i.e. there are sets of coplanar points. Coplanarity is a strong constraint for both structure recovery and motion estimation. Most existing works do only exploit one of the two aspects, or, if both, then in a sub-optimal manner. A typical example is to estimate motion (epipolar geometry) using raw point correspondences, to perform a 3D reconstruction and then to fit planes and maybe correct 3D point positions to make them coplanar. In this paper, we present an approach to estimate camera motion and piecewise planar structure simultaneously and optimally: the result is the estimation of camera motion and 3D structure, that minimizes reprojection error, while satisfying the piecewise planarity. The estimation problem is minimally parameterized using 2D entities – epipoles, epipolar transformation, plane homographies and image points – subsequently deriving the corresponding 3D entities is trivial. Experimental results show that the reconstruction is of clearly superior quality compared to traditional methods based only on points, even if the scene is not perfectly piecewise planar.

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.). Piecewise planar structures constitute very strong geometric constraints from which we can expect better reconstruction results than from the traditional methods based only on points.

We propose a projective framework including the *MLE* (Maximum Likelihood Estimator) for structure and motion recovery from two views of a piecewise planar scene, which provides the flexibility of working with uncalibrated or partially calibrated images.

Such an estimator needs an algebraic representation of geometric structures, either in *3D* or image-based. Both approaches have advantages and drawbacks. In the *3D* case, it is difficult to enforce geometric constraints (e.g. express a point that belongs to a plane with only two parameters), especially in a projective framework. In [11] for example, the structure is corrected to be quasi-piecewise planar during bundle adjustment via heavily weighted additional equations. So, on the one hand, the problem is overparameterized, on the other hand large terms enter the equation system, which might affect numerical stability. As for the image-based approach, *3D* geometric entities are usually represented indirectly (e.g. an homography matrix for a plane), consistent modeling can thus be non-trivial, especially if the number of images is not small.

As the most important goal in this paper is to devise the *MLE*, we choose the image-based approach that allows to represent points on planes. The major difficulty is then to express consistently the image entities that represent the *3D* scene geometry and camera motion. Consistent means that the number of dof (degrees of freedom) of the representation is the same as that of the essential dof.

The analysis of the algebraic entities on the image level reveals that the number of algebraic dof is higher than that of essential dof, which results in an overparameterization. Indeed, the motion can be represented by the epipolar geometry, i.e. 7 essential dof but 9 algebraic ones when expressed via the fundamental matrix. Each plane has 3 essential dof but induces a plane homography which has 9 algebraic ones. Finally, each point lying on a plane has 2 essential dof but its projections onto the images contain 4 algebraic ones.

The main result of this paper is a *consistent parameterization on the image level* expressing the entire scene geometry. The parameterization and the corresponding *MLE* are given in §§2 and 3 respectively. A method to obtain initial values for plane parameters is described in §4. The proposed approach is validated using simulated data in §3 where we investigate in particular the behaviour of the esti-

mator in the case of approximately piecewise planar scenes. Experimental results using real images are shown in §5 followed by our conclusions and perspectives.

In the following two paragraphs, we review existing work and give some preliminaries.

1.1. Previous Work

One of the first attempts to give a minimal parameterization of camera motion in the case of two uncalibrated views is [7] where the author addresses the particular case of finite epipoles and devises a minimal parameter set, called a *map*, for the epipolar geometry. This work has been extended to the general case in [17] where the author devises all 36 possible maps (different ways of parameterizing rank-2-ness of the fundamental matrix and of dealing with the scale uncertainty). The optimization procedure is costly because of a step by step choice of the appropriate map and the practical interest is limited due to the high number of possible maps. The link with plane homographies has not been made in these two works.

In [3, 12], two different methods for structure and motion in a piecewise planar environment are proposed in the case of calibrated images. They do not include an image level representation and do not yield an *MLE*.

The constraint of coplanarity has been studied in [11, 16]. The *3D* representation of structures does not permit to model points on planes (see above). The results obtained in [11] show that the accuracy of the reconstruction is not better compared to not using coplanarity information but that it appears visually smoother when using planar structures.

In our approach, where all geometric entities are expressed on the image level, points on planes are minimally parameterized so that they really lie on planes. The *MLE* is then obtained without using superfluous equations.

1.2. Preliminaries

We use perspective projection to model cameras. In this case, two projections $\mathbf{x}$ and $\mathbf{x}'$ of the same *3D* point $\mathbf{X}$ are related via ${\mathbf{x}'}^{\mathsf{T}}\mathsf{F}\mathbf{x} = 0$ where F is the 3×3 rank deficient fundamental matrix representing the epipolar geometry [7]. The two epipoles $\mathbf{e}$ and $\mathbf{e}'$ are defined as $\mathsf{F}\mathbf{e} = \mathsf{F}^{\mathsf{T}}\mathbf{e}' = \mathbf{0}$.

If the point $\mathbf{X}$ lies on a plane, $\mathbf{x}$ and $\mathbf{x}'$ are related via $\mathbf{x}' \sim \mathsf{H}\mathbf{x}$ where H is a full rank (in general) 3×3 matrix H representing a plane homography and $\sim$ denotes the equality up to a non-null scale factor. Any plane homography H is linked to the fundamental matrix via $\mathsf{F} \sim [\mathbf{e}']_\times \mathsf{H}$ where $[.]_\times$ denotes the matrix associated with the cross product, i.e. $[\mathbf{v}]_\times \mathbf{q} = \mathbf{v} \times \mathbf{q}$. This implies that if we fix a reference plane homography H_r, H can be written [9]:

$$\mathsf{H} \sim \mathsf{H}_r + \mathbf{e}'\mathbf{a}^{\mathsf{T}}, \tag{1}$$

where $\mathbf{a}$ is an inhomogeneous 3-vector. The equation of the plane $\boldsymbol{\pi}$ inducing H is $(\mathbf{a}^{\mathsf{T}} | -1)^{\mathsf{T}}$ in the projective basis defined by the projection matrices $\mathsf{P} \sim (\mathrm{I}_3 | \mathbf{0}_3)$ and $\mathsf{P}' \sim (\mathsf{H}_r | \mathbf{e}')$. In this basis, the reconstruction of a point lying on the plane $\boldsymbol{\pi}$, given e.g. its projection $\mathbf{x}$ in the first image, writes:

$$\mathbf{X}^{\mathsf{T}} \sim (\mathbf{x}^{\mathsf{T}} | \mathbf{x}^{\mathsf{T}}\mathbf{a}). \tag{2}$$

Throughout this paper, we use the Levenberg-Marquardt algorithm [10] to conduct non-linear optimization processes.

2 A Consistent Structure and Motion Parameterization

In this section, we define a consistent image level representation for the structure and motion of a piecewise planar scene. This parameterization is consistent in the sense that its number of dof strictly corresponds to the number of essential dof of the geometry, namely 7 for the epipolar geometry, 3 for each modeled plane and 2 for each point on a modeled plane.

Let us first consider the case of the epipolar geometry. Its components are defined in [7, 17] as the set $\nu = \{\mathbf{e}, \mathbf{e}', \widetilde{\mathsf{h}}\}$ where $\mathbf{e}$ and $\mathbf{e}'$ are the two epipoles and $\widetilde{\mathsf{h}}$ the epipolar transformation (a $\mathbb{P}^1$ homography relating the two epipolar pencils, see below). In [17], it is shown that 36 maps are necessary to cover all two view configurations.

A plane is parameterized by a 3-vector $\mathbf{a}$. Equation (1) gives then the corresponding plane homography H, necessary to obtain image-based constraints. The difficulty is then to choose the reference homography H_r. A possibility is $\mathsf{H}_r \sim [\mathbf{e}']_\times \mathsf{F}$ [9], but this yields a singular H_r of a complicated closed-form expression.

Our method relies on the introduction of a singular *2D* homography called the *extended epipolar transformation* that we denote $\widetilde{\mathsf{H}}$ and that is formed uniquely from ν. It is the basis for a consistent expression of the epipolar geometry and a regular reference plane homography of a simplified form. Each point lying on a modeled plane will be represented by a single image point.

In the next section, we present the extended epipolar transformation and a way to reduce the 36 original maps to 3 basic ones. We then construct a parameterization of a reference homography and of the fundamental matrix. Finally, we consider the problem of selecting the appropriate map.

2.1. The Extended Epipolar Transformation

We clarify the role of the epipolar transformation $\widetilde{\mathsf{h}}$ and show how the construction of the extended one $\widetilde{\mathsf{H}}$, leads to 36 possible cases. We then give a detailed construction in

the case where both epipoles are finite. The notations correspond to those shown in figure 1.

The epipolar transformation $\widetilde{\mathsf{h}}$ is the *1D* homography between the two pencils of epipolar lines, induced by the epipolar geometry. In order to materialize this relation, one needs to express the epipolar lines as elements of $\mathbb{P}^1$. This can be done by intersecting the epipolar pencils with a line $\mathbf{d}$ in the left image and a line $\mathbf{d}'$ in the right one. The intersections $\mathbf{p}$ and $\mathbf{p}'$ of two corresponding epipolar lines with $\mathbf{d}$ and $\mathbf{d}'$ induce two $\mathbb{P}^1$ points $\bar{\mathbf{p}}$ and $\bar{\mathbf{p}}'$ related by $\widetilde{\mathsf{h}}$. The lines $\mathbf{d}$ and $\mathbf{d}'$ are chosen to not contain the epipoles.

The extended epipolar transformation $\widetilde{\mathsf{H}}$ is a $\mathbb{P}^2$ homography relating the points $\mathbf{p}$ and $\mathbf{p}'$ in the same way $\widetilde{\mathsf{h}}$ relates $\bar{\mathbf{p}}$ and $\bar{\mathbf{p}}'$. Consequently, it is not unique and may be singular. Note that $\widetilde{\mathsf{H}}$ does not necessarily correspond to a world plane.

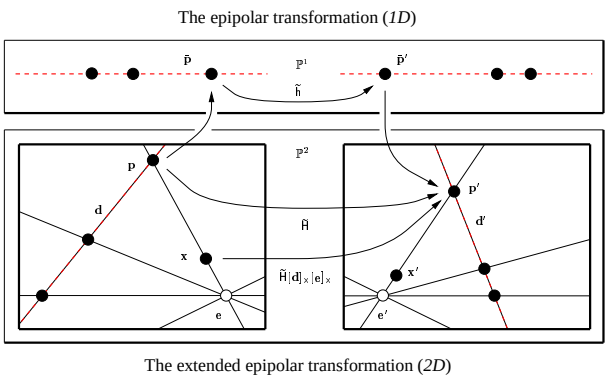

Figure 1. The role of the extended epipolar transformation.

The construction of $\widetilde{\mathsf{H}}$ from $\widetilde{\mathsf{h}}$ is a crucial point as it depends on the choice of $\mathbf{d}$ and $\mathbf{d}'$ which itself depends on the two view configuration at hand. Actually, the problem is to express $\bar{\mathbf{p}}$ and $\bar{\mathbf{p}}'$ as functions of $\mathbf{p}$ and $\mathbf{p}'$ respectively. The idea seems straigthforward: as $\mathbf{p}$ lies on the line $\mathbf{d}$, we can form $\bar{\mathbf{p}}$ with two significant elements of $\mathbf{p}$. Which ones are significant depends on the value of $\mathbf{d}$. To make this relationship explicit, we express the constraint that $\mathbf{p} \in \mathbf{d}$ by the equation $\mathbf{p}^\mathsf{T}\mathbf{d} = 0$. If we develop this bilinear expression, it immediately follows that expressing one element p_i of $\mathbf{p}$ in terms of $\mathbf{d}$ and of the other elements of $\mathbf{p}$ requires that the corresponding element of $\mathbf{d}$ satisfies $d_i \neq 0$. As $i \in 1\ldots3$, there are 3 possibilities. This reasoning has also to be done for $\bar{\mathbf{p}}'$, which yields 3×3=9 maps.

In a similar way, expressing the 2×2 matrix $\widetilde{\mathsf{h}}$ with only 3 parameters requires to fix an element to a constant value. The 4 possible choices for this element finally lead to the 9×4=36 maps of [17].

Let us derive in detail the case when both epipoles are finite. The only lines that never contain $\mathbf{e}$ and $\mathbf{e}'$ are the lines at infinity in each image, hence $\mathbf{d} \sim \mathbf{d}' \sim (0,0,1)^\mathsf{T}$. The intersection points take the form $\mathbf{p} \sim (\bar{\mathbf{p}}, 0)^\mathsf{T}$ and $\mathbf{p}' \sim (\bar{\mathbf{p}}', 0)^\mathsf{T}$ and are related by the epipolar transformation via $\bar{\mathbf{p}}' \sim \widetilde{\mathsf{h}}\bar{\mathbf{p}}$. Consequently, the extension of $\widetilde{\mathsf{h}}$ takes the form $\widetilde{\mathsf{H}} \sim \begin{pmatrix} \widetilde{\mathsf{h}} & \mathbf{q} \\ \mathbf{0}_2^\mathsf{T} & \end{pmatrix}$ where $\mathbf{q}$ is any 3-vector, chosen here as $\mathbf{q} = \mathbf{0}$.

There also exist particular two view configurations [15] where the extended epipolar transformation might take a simplified form. The principle of parameterization developed here can easily be applied to them.

2.2. Reducing the Number of Maps

The 36 maps are not all necessary to model all situations. We state in table 1 a set of 3 basic maps to which we can bring back every case. This is important for an easy implementation of the parameterization: all the cases that we remove for the construction of $\widetilde{\mathsf{H}}$ will not appear for that of H_r and F.

This reduced set of maps is obtained by fixing first the scale ambiguity of $\widetilde{\mathsf{h}}$ using an additional constraint such as $||\widetilde{\mathsf{h}}||^2 = 1$. The number of maps reduces to 9. An appropriate image rotation (which leaves the residual invariant) ensure that an infinite epipole, e.g. $\mathbf{e}$ does not lie on the chosen line $\mathbf{d}$ (here $\mathbf{d} \sim (1,0,0)^\mathsf{T}$). Using then the fact that the role of the images can be exchanged, the number of maps reduces to 3.

Note that the number of essential dof is not the same for each case because a finite epipole has 2 but an infinite one has only 1.

case	$\mathbf{e}$	$\mathbf{e}'$	#dof	ν
1	$\not\infty$	$\not\infty$	7	$\widetilde{\mathsf{h}}, e_1, e_2, e_1', e_2'$
2	$\not\infty$	∞	6	$\widetilde{\mathsf{h}}, e_1, e_2, e_2'$
3	∞	∞	5	$\widetilde{\mathsf{h}}, e_2, e_2'$

Table 1. The 3 basic cases that have to be parameterized.

2.3. Construction of the Reference Homography

This section aims at giving the expression of a reference homography H_r in terms of the parameters ν of the epipolar geometry. As any regular plane homography can be used, we choose a practical form in the 3 parameters space.

Let us start by giving a general formula to express plane homographies from $\widetilde{\mathsf{H}}$. According to the interpretation of $\widetilde{\mathsf{h}}$ and $\widetilde{\mathsf{H}}$ given in §2.1, we use the following construction. For a given point $\mathbf{x}$ in the left image, we find the corresponding

epipolar line and intersect it with the line $\mathbf{d}$ to get the point $\mathbf{p}$. Using $\widetilde{\mathsf{H}}$, we retrieve the point $\mathbf{p}'$ of the right image. Plane homographies are then obtained by adding the term $\mathbf{e}'\mathbf{a}^\mathsf{T}$ where the inhomogeneous 3-vector $\mathbf{a}$ parameterizes the family [9]. This yields the formula:

$$\mathsf{H}_\mathbf{a} \sim \widetilde{\mathsf{H}}[\mathbf{d}]_\times[\mathbf{e}]_\times + \mathbf{e}'\mathbf{a}^\mathsf{T}. \tag{3}$$

When fixing $\mathbf{a}$, the above expression gives a reference homography. Table 2 shows the derivation for the 3 basic cases. These expressions are practical in the sense that H_r is regular ($det(\mathsf{H}_r) = -det(\widetilde{\mathsf{h}})$ in all cases) and that their forms are affine-like (a row with two zeros and a one) and so handled easily, e.g. for the computation of H_r^{-1}.

We derive in detail the case 1. In this case, we have $\mathbf{d} \sim \mathbf{l}_\infty \sim (0,0,1)^\mathsf{T}$. Using the formulation established previously for $\widetilde{\mathsf{H}}$, we deduced from equation (3):

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

Choosing $\mathbf{a} = (0,0,1)^\mathsf{T}$ yields the affine expression given in table 2.

case	$\mathbf{a}$	H_r
1	$(0,0,1)^\mathsf{T}$	$\begin{pmatrix} -\widetilde{\mathsf{h}} & \widetilde{\mathsf{h}}(e_1,e_2)^\mathsf{T} + (e'_1,e'_2)^\mathsf{T} \\ \mathbf{0}_2^\mathsf{T} & 1 \end{pmatrix}$
2	$(0,0,1)^\mathsf{T}$	$\begin{pmatrix} \mathbf{0}_2^\mathsf{T} & 1 \\ -\widetilde{\mathsf{h}} & \widetilde{\mathsf{h}}(e_1,e_2)^\mathsf{T} + (e'_2,0)^\mathsf{T} \end{pmatrix}$
3	$(1,0,0)^\mathsf{T}$	$\begin{pmatrix} 1 & \mathbf{0}_2^\mathsf{T} \\ \widetilde{\mathsf{h}}(e_2,0)^\mathsf{T} + (e'_2,0)^\mathsf{T} & -\widetilde{\mathsf{h}} \end{pmatrix}$

Table 2. Parameterization of a reference plane homography H_r for the 3 basic cases defined in table 1.

2.4. Construction of the Fundamental Matrix

Let us give the general formulation of the fundamental matrix using the same reasoning as above. Given a point $\mathbf{x}$ in the left image, we know how to associate to it a point $\mathbf{p}'$ lying on its associated epipolar line in the right image. The last step to obtain this epipolar line is then to link $\mathbf{p}'$ and $\mathbf{e}'$. The fundamental matrix is then expressed by $\mathsf{F} \sim [\mathbf{e}']_\times\widetilde{\mathsf{H}}[\mathbf{d}]_\times[\mathbf{e}]_\times$. The resulting form of F for the 3 basic cases is given in table 3. The expression for case 1 (obtained with $\mathbf{d} \sim \mathbf{l}_\infty \sim (0,0,1)^\mathsf{T}$) coincides with that given in [7, 17].

2.5. Choosing the Best Map

A method to choose the best map for the fundamental matrix is given in [17]. Provided an initial guess, it consists in selecting the map that locally is the least singular. The major drawback of this criterion is that, as it is local, it has to be included in the optimization loop. Consequently, all 36 maps are tried at each step of the optimization process, which represents a non negligible cost. Moreover, this does not take into account that the number of dof is not the same for all maps.

A way to enhance such a process is to devise a criterion which takes into account the particular number of dof of each map. Consequently, it might be very appropriate to use a model selection criterion such as *MDL* or *AIC* [13, 6] and not at every step of the optimization process.

case	F
1	$\begin{pmatrix} c & d & -ce_1 - de_2 \\ -a & -b & ae_1 + be_2 \\ ae'_2 - ce'_1 & be'_2 - de'_1 & k \end{pmatrix}$ with $k = (ce_1 + de_2)e'_1 - (ae_1 + be_2)e'_2$
2	$\begin{pmatrix} -ce'_2 & -de'_2 & (ce_1 + de_2)e'_2 \\ c & d & -ce_1 - de_2 \\ -a & -b & ae_1 + be_2 \end{pmatrix}$
3	$\begin{pmatrix} ce_2e'_2 & -ce'_2 & -de'_2 \\ -ce_2 & c & d \\ ae_2 & -a & -b \end{pmatrix}$

Table 3. Fundamental matrix parameterization for the 3 basic cases defined in table 1. The scalars a, b, c and d are the coefficients of the epipolar transformation $\widetilde{\mathsf{h}}$.

2.6. Summary

In this section, we have developed minimal parameterizations of the fundamental matrix and a reference plane homography, for all cases of finite/infinite epipoles. These parameterizations (see tables 2 and 3) lead to very simple closed-form expressions for plane homographies. This enables a very efficient simultaneous optimization of structure and motion, as described in §3.

3. Optimal Estimation

In this section, we derive a criterion to compute the optimal structure and motion of a piecewise planar scene including non coplanar points. Optimal is taken in the sense of the maximum likelihood under the assumption of i.i.d. centered Gaussian noise in measured image point coordinates.

Two functions give respectively the fundamental matrix $\mathsf{F} \sim f(\nu)$ and any plane homography $\mathsf{H}_j \sim h(\nu, \mathbf{a}_j)$ from

the parameters ν of the epipolar geometry and a plane equation $\mathbf{a}_j$ (according to the expressions in tables 2 and 3 and equation (1)).

3.1. A Criterion for the *MLE*

We apply the parameterization defined in §2 to make use of a completely piecewise planar scene structure. We then complete the approach to address the case of a partially piecewise planar and partially general scene structure, i.e. containing both points belonging and not belonging to coplanar group.

Each point lying on a modeled plane is parameterized in the left image. Its corresponding point in the right image is given by applying the adequate plane homography. For a given map, the residual to minimize is then given by $\mathcal{R}_{\mathsf{H}_j} = \sum_{\{(\mathbf{x}\leftrightarrow\mathbf{x}')\in\boldsymbol{\pi}_j\}} \left(d^2(\mathbf{x},\widehat{\mathbf{x}}) + d^2(\mathbf{x}',\mathsf{H}_j\widehat{\mathbf{x}})\right)$, where $\boldsymbol{\pi}_j$ denotes the plane of equation $\mathbf{a}_j$ corresponding to the plane homography H_j and $d(.,.)$ the Euclidean distance.

The optimal parameters are obtained as [14]:

$$\{\mathsf{F},\mathbf{a}_1,\ldots,\mathbf{a}_m,\{\widehat{\mathbf{x}}\}\} = \underset{\nu\cup\{\mathbf{a}_1,\ldots,\mathbf{a}_m\}\cup\{\widehat{\mathbf{x}}\}}{argmin} \sum_{j=1}^{m} \mathcal{R}_{\mathsf{H}_j}, \quad (4)$$

under the constraints $\mathsf{F} \sim f(\nu)$ and $\mathsf{H}_j \sim h(\nu,\mathbf{a}_j)$. Note that maximum likelihood estimates are achieved only in the case when each point belongs to only one plane (e.g. if $\mathbf{x}\in\boldsymbol{\pi}_1$ and $\mathbf{x}\in\boldsymbol{\pi}_2$, we can not guarantee that $\mathsf{H}_1\widehat{\mathbf{x}} \sim \mathsf{H}_2\widehat{\mathbf{x}}$). The epipolar geometry is implicitly estimated via plane homographies[1].

It is also possible to make points $\{\mathbf{y}\leftrightarrow\mathbf{y}'\}$ that do not belong to any plane, contribute to the estimation:

$$\{\mathsf{F},\mathbf{a}_1,\ldots,\mathbf{a}_m,\{\widehat{\mathbf{x}}\},\{\widehat{\mathbf{y}}\leftrightarrow\widehat{\mathbf{y}}'\}\} = \underset{\nu\cup\{\mathbf{a}_1,\ldots,\mathbf{a}_m\}\cup\{\widehat{\mathbf{x}}\}\cup\{\widehat{\mathbf{y}}\leftrightarrow\widehat{\mathbf{y}}'\}}{argmin} \left(\left(\sum_{j=1}^{m}\mathcal{R}_{\mathsf{H}_j}\right) + \mathcal{R}_{\mathsf{F}}\right),$$

using the same constraints as for the optimization of criterion (4) plus $\widehat{\mathbf{y}}'^{\mathsf{T}}\mathsf{F}\widehat{\mathbf{y}} = 0$ so that points satisfy exactly the epipolar geometry. The residual $\mathcal{R}_{\mathsf{F}}$ is given by $\sum_{\{\mathbf{y}\leftrightarrow\mathbf{y}'\}} \left(d^2(\mathbf{y},\widehat{\mathbf{y}}) + d^2(\mathbf{y}',\widehat{\mathbf{y}}')\right)$.

3.2. Experimental Results

In this section, we compare our *MLE* estimator to various others that use or do not use coplanarity information, using simulated data. Our experimental results concern the case 1 (both epipoles are finite) of §2.

[1]Note that this is very different from estimating individual plane homographies, and subsequently F from these [8], which is known to be rather unstable. Here, the epipolar geometry and all plane homographies are estimated simultaneously.

The test bench consists of a one meter cube at various distances from two cameras. A number of n =50 points lying on each of three faces of the cube are projected onto the images. Gaussian centered noise is added to the image points. We evaluate the methods by assessing the quality of *3D* reconstructions that are based on the image level estimation results. *3D* reconstruction is achieved using triangulation [5] in the general case, and equation (2) for points on planes. The quality measure is the *RMS 3D* Euclidean distance $E_3 = \sqrt{\frac{1}{n}\sum_{\{\mathbf{X}\leftrightarrow\bar{\mathbf{X}}\}} d^2(\mathsf{H}_3\mathbf{X},\bar{\mathbf{X}})}$, where $\{\mathbf{X}\}$ is the estimated projective reconstruction and $\{\bar{\mathbf{X}}\}$ the true Euclidean one. The *3D* homography H_3 is estimated via nonlinear minimization of E_3.

The estimators compared are divided into two sets. Epipolar geometry-based estimators, *Methods F* [5]:

- *FLin+BA*: normalized 8 point algorithm for the epipolar geometry followed by a bundle adjustment of points;
- *FML*: *MLE* for the epipolar geometry and the image points;
- *trueF+BA*: bundle adjustment of points using the true epipolar geometry.

Plane-based estimators, *Methods H*:

- *HiML+FML*: maximum likelihood estimation of plane homographies [5] and then the method *FML*;
- *consHiML*: the consistent approach developed in this paper (equation (4));
- *trueHi+BA*: bundle adjustment of points [5] using the true plane homographies.

We have carried out two sets of experiments, where *3D* points lie in either perfectly or nearly coplanar groups.

The first set of experiments (see figure 2) show that when *3D* points are perfectly coplanar, *Methods H* perform better (the residual is two times lower) than *Methods F*. In more detail, we can say for *Methods F* that *trueF+BA* performs better than *FML* which itself performs better than *FLin+BA*, and the same for *Methods H*, *trueHi+BA* performs better than *consHiML* which itself performs better than *HiML+FML*. When the distance scene/cameras or noise level increase, *Methods F* diverge, whereas *Methods H* do not.

Now, let us investigate the second set of experiments. In this case, the *3D* points are offset vertically from their planes by a random distance (Gaussian noise with standard deviation between 0 and 0.1 meters).

Once again, the behaviour of the tested methods can be divided into the same two sets as above. Let us denote the breakdown ratio ε as the ratio between the planes unflatness

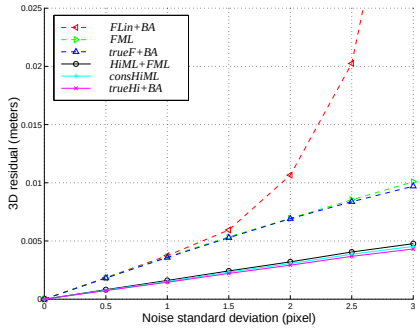

Figure 2. Comparison of the different methods using the *3D* residual E_3, for different noise levels and a distance scene/cameras of 10 meters.

and the size of the simulated planar surface where *Methods H* begin to perform worse than *Methods F*, e.g. for figure 3, ε=6%. Table 4 shows the value of ε established experimentally for different cases. The less stable the configuration is (large noise and/or a high distance scene/cameras), the higher is ε, i.e. the more important is the consistent incorporation of coplanarity constraints, even if the scene is not perfectly piecewise planar.

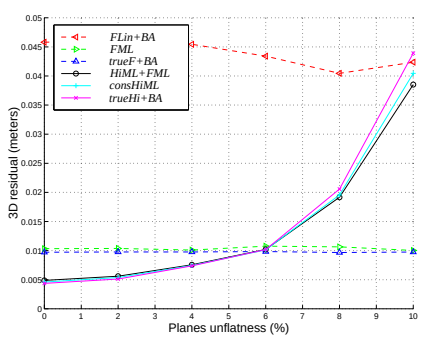

Figure 3. Comparison of the different methods using the *3D* residual E_3, for different planes unflatness, a distance scene/cameras of 10 meters and a 3 pixels standard deviation noise.

The values of one or several percent in table 4 represent relatively large variations which are superior to those of a great majority of approximately planar real surfaces. Consequently, we can say that there are a lot of cases when a plane-based method will perform better than any method based only on points.

Now, let us see how to initialize the values of the plane

	3 m.	10 m.	20 m.
1 pixel	0.5%	2%	4%
3 pixels	2%	6%	9%

Table 4. Breakdown ratio ε for different combinations of distance scene/cameras and noise level.

parameters for the *MLE* described in this section.

4. Initialization of Plane Equations

In this section, we aim at finding an initialization of the modeled scene planes required by the *MLE* of §3. At this point, we assume that the image points are clustered into coplanar groups (see §5).

Given the epipolar geometry, one can extract a reference plane homography H_r and the right epipole $\mathbf{e}'$ as proposed in §2. The plane equations are then given from plane homographies using equation 1. However, an estimated homography does not correspond to a world plane in general. We show how to constrain the estimation so that the estimated homography really corresponds to a world plane.

Each point correspondence $\mathbf{x} \leftrightarrow \mathbf{x}'$ arising from a point lying on a plane $\boldsymbol{\pi}$ is subject to $\mathbf{x}' \sim \mathsf{H}\mathbf{x}$. Introducing equation (1) yields $\mathbf{x}' \sim (\mathsf{H}_r + \mathbf{e}'\mathbf{a}^\mathsf{T})\mathbf{x}$, where $\mathbf{a}$ corresponds to the plane equation. By nullifiying the cross product of these vectors, and after some minor algebraic manipulations, we obtain $[\mathbf{x}']_\times \mathbf{e}'\mathbf{x}^\mathsf{T}\mathbf{a} = [\mathbf{x}']_\times [\mathbf{e}']_\times \mathsf{F}\mathbf{x}$. The equations corresponding to each point correspondence can then be rewritten to form a linear system for $\mathbf{a}$. This method can be used with a minimum of three point correspondences, which corresponds to the fact that a plane is defined by three points.

A quasi-linear estimator for the plane homography is proposed in [1] and could also be used to estimate the plane equation. However, it is not of a signifiant interest since this does not affect the result provided by the final *MLE*.

5. Results Using Real Images

In this section, we present the reconstruction results that we obtained using the real images of figure 4. Similar results have been obtained with other images. We describe the different steps necessary to perform a complete reconstruction, from the images to the *3D* textured model. We then compare our plane-based approach to the point-based method given in [5].

Estimation of the epipolar geometry: More than 450 interest points have been detected and slightly less than 100

Figure 4. The *Game System* stereo image pair.

matches have been automatically established while estimating the epipolar geometry between the two images [17]. Around 20% of these matches are outliers.

This initial estimate allows to compute a reference homography and the epipoles, as indicated in §2.

Segmentation into planes: This step provides the initilization to the optimal reconstruction process.

Planar structures are segmented semi-automatically using a *RANSAC*-like [4] algorithm. Provided the plane estimator (see §4) for the random sampling, the algorithm estimates recursively the successive dominant planes and their associated point correspondences. The user may interact with the system to add, remove or split planar structures.

Once all planes are modeled, we attempt to match previously unmatched interest points using plane homographies. This step is iterated until the convergence of point correspondences.

Texture maps: This requires the user to provide the polygonal contour of planar facets in one image.

Projective reconstruction: A projective reconstruction is estimated using the plane-based algorithm of §3. Textured rendering are visible on figure 5.

Figure 5. Textured rendering of the recovered projective model for the *Game System* stereo image pair.

Metric reconstruction: The metric structure is obtained via an autocalibration process using the fundamental matrix [2]. It is also possible to enhance this result using the technique based on multi-planes described in [16]. Textured rendering are visible on figure 6.

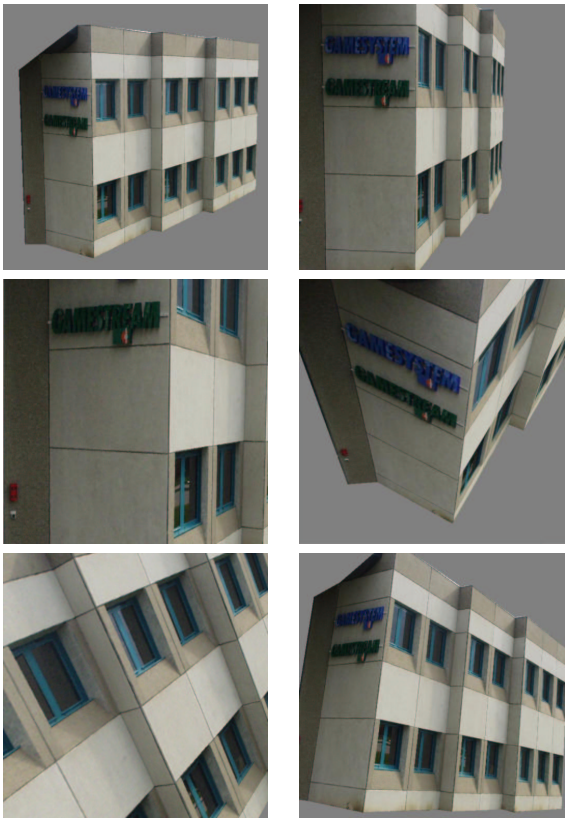

Figure 6. Textured rendering of the recovered metric model for the *Game System* stereo image pair.

Quality assessment: We have performed several measures on the metric reconstruction obtained using the method based on planes described above and on the one obtained using a method based only on points [5].

Two kinds of quantity are significant: length ratios and angles. Table 5 shows measures of such quantities. In this table, σ_1 and σ_2 are the variances of the length of respectively the 6 vertical edges and the 6 horizontal edges of equal length, whereas μ is the mean of $1 - 2\alpha_i/\pi$ where α_i are the measures of right angles.

The values given in table 5 show that the metric reconstruction obtained with the consistent plane-based method described in this paper is clearly of superior quality than the one obtained with the method based only on points.

	σ_1	σ_2	μ
point-based	0.0146	0.0543	0.0633
plane-based	0.0082	0.0267	0.0413

Table 5. Metric measures on the Euclidean reconstruction using a point-based or our plane-based method. The lower σ_1, σ_2 and μ (see text) are, the better the reconstruction is.

6. Conclusions and Perspectives

We have presented an *MLE* for the complete structure and motion from two uncalibrated views of a piecewise planar scene. The geometric structures are consistently represented on the image level by a fundamental matrix, a set of plane equations and points on planes.

The initialization of the *MLE* is provided by the 8 point algorithm for the epipolar geometry. The plane equations are then estimated image-based.

Experimental results on both simulated data and real images show that the reconstruction quality obtained with our consistent plane-based approach is clearly superior to those of methods that only reconstruct the individual points, even if the scene is not perfectly piecewise planar.

We are currently investigating the use of model selection criteria for the choice of the most appropriate map for a given fundamental matrix. We also plan to extend the approach to more than two images.

References

[1] A. Bartoli, P. Sturm, and R. Horaud. A projective framework for structure and motion recovery from two views of a piecewise planar scene. Research Report 4070, INRIA, Grenoble, France, October 2000.

[2] S. Bougnoux. From projective to euclidean space under any practical situation, a criticism of self-calibration. In *Proceedings of the International Conference on Computer Vision*, pages 790–796, 1998.

[3] O. Faugeras and F. Lustman. Motion and structure from motion in a piecewise planar environment. *International Journal of Pattern Recognition and Artificial Intelligence*, 2(3):485–508, September 1988.

[4] 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, 1981.

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

[6] K. Kanatani. Geometric information criterion for model selection. *International Journal of Computer Vision*, 26(3):171–189, 1998.

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

[8] Q.T. Luong and O.D. Faugeras. Determining the fundamental matrix with planes: Instability and new algorithms. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, 1993.

[9] Q.T. Luong and T. Vieville. Canonic representations for the geometries of multiple projective views. *Computer Vision and Image Understanding*, 64(2):193–229, 1996.

[10] 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, 1992.

[11] R. Szeliski and P.H.S. Torr. Geometrically constrained structure from motion : Points on planes. In *3D Structure from Multiple Images of Large-scale Environments SMILE*. 1998.

[12] J.-P. Tarel and J.-M. Vézien. A generic approach for planar patches stereo reconstruction. In *Proceedings of the Scandinavian Conference on Image Analysis*, pages 1061–1070, 1995.

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

[14] B. Triggs. Optimal estimation of matching constraints. In *3D Structure from Multiple Images of Large-scale Environments SMILE*. 1998.

[15] T. Viéville and D. Lingrand. Using singular displacements for uncalibrated monocular visual systems. In *Proceedings of the 4th European Conference on Computer Vision*, pages 207–216. 1996.

[16] G. Xu, J.-I. Terai, and H.-Y. Shum. A linear algorithm for camera self-calibration, motion and structure recovery for multi-planar scenes from two perspective images. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, 2000.

[17] Z. Zhang. Determining the epipolar geometry and its uncertainty: A review. *International Journal of Computer Vision*, 27(2):161–195, 1998.