ICCV'01 - EIGHTH IEEE INTERNATIONAL CONFERENCE ON COMPUTER VISION, VANCOUVER, BC, CANADA, PP. 593-598, JULY 2001.

Projective Structure and Motion from Two Views of a Piecewise Planar Scene

Adrien Bartoli Peter Sturm Radu Horaud

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

Abstract

In this paper, we address the problem of structure and motion recovery from two views of a scene containing planes, i.e. sets of coplanar points. Most of the existing works do only exploit this constraint in a sub-optimal manner. We propose to parameterize the structure of such scenes with planes and points on planes and derive the MLE (Maximum Likelihood Estimator) using a minimal parameterization based on 2D entities. The result is the estimation of camera motion and 3D structure in projective space, that minimizes reprojection error, while satisfying the piecewise planarity. We propose a quasi-linear estimator that provides reliable initialization values for plane equations. 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. 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, using 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.

On ther other hand, the algebraic entities on the image level are easier to minimally parameterize (e.g. an image point for a point lying on a plane).

The main result of this paper, §2 is a *consistent parameterization on the image level* expressing the entire scene geometry. This provides the material to devise the *MLE* in §3. A method to obtain reliable initialization values for plane parameters is described in §4. Experimental results are presented in §§3 and 4 and §5 for simulated data and real images respectively followed by our conclusions and perspectives. In the following two paragraphs, we review existing work and give some preliminaries.

1.1. Previous Work

Maybe the first attempt 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 [16] 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 for the case of calibrated images. They do not include an image level representation and do not yield an *MLE*.

This work was supported by the project IST-1999-10756, VISIRE.

The constraint of coplanarity has been studied in [11, 15]. 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 $\mathsf{F}\mathbf{e} = \mathsf{F}^{\mathsf{T}}\mathbf{e}' = \mathbf{0}$. The epipolar geometry can be decomposed as $\mathbf{e}$, $\mathbf{e}'$ and the epipolar transformation $\widetilde{\mathsf{h}}$, a $\mathbb{P}^1$ homography relating corresponding epipolar lines [7, 16].

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 representing a plane homography. 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 π 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 coordinates of a point $\mathbf{X}$ lying on the plane π, given e.g. its projection $\mathbf{x}$ in the first image, are:

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

We define an *algebraic distance* d_a between two points $\mathbf{x}$ and $\mathbf{y}$ of the Euclidean plane $\mathbb{E}^2$ by $d_a^2(\mathbf{x}, \mathbf{y}) = ||\mathsf{S}(\mathbf{x} \times \mathbf{y})||^2$ where $\mathsf{S} = (\mathrm{I}_2|\mathbf{0}_2)$. The relation with the *Euclidean distance* d is via a scalar w depending only on the last elements of $\mathbf{x}$ and $\mathbf{y}$:

$$d^2(\mathbf{x}, \mathbf{y}) = w\, d_a^2(\mathbf{x}, \mathbf{y}) \text{ where } w = \frac{1}{(x_3 y_3)^2}. \tag{3}$$

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. More details can be found in [1].

For that purpose, we first introduce what we call the *extended epipolar transformation*, a singular $\mathbb{P}^2$ homography relating particular points on corresponding epipolar lines.

We then construct a parameterization of a reference homography and of the fundamental matrix. Finally, we consider the problem of selecting the most appropriate map.

2.1. The Extended Epipolar Transformation

Let us define points $\mathbf{p}$ and $\mathbf{p}'$ as the intersections of two corresponding epipolar lines and respectively two lines $\mathbf{d}$ and $\mathbf{d}'$ not containing the epipoles. These points, see figure

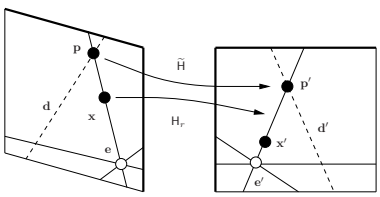


Figure 1. The role of the extended epipolar transformation.

1, are related by the extended epipolar transformation $\widetilde{\mathsf{H}}$, whose structure depends on both $\mathbf{d}$ and $\mathbf{d}'$. The choice for $\mathbf{d}$ and $\mathbf{d}'$ is explained below.

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 $\mathbf{p}' \sim (\bar{\mathbf{p}}', 0)^{\mathsf{T}}$. Consequently, the extended epipolar transformation writes $\widetilde{\mathsf{H}} \sim \begin{pmatrix} \widetilde{\mathsf{h}} & \mathbf{q} \\ \mathbf{0}_2^{\mathsf{T}} & \end{pmatrix}$ where $\mathbf{q}$ is any 3-vector, chosen here as $\mathbf{q} = \mathbf{0}$ and $\widetilde{\mathsf{h}}$ the epipolar transformation.

There also exist particular two view configurations [14] where the extended epipolar transformation might take a simplified form. The principle of parameterization developed here can easily be applied to them.

Table 1 summarizes the 3 basic cases that have to be parameterized. 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. The minimal set of parameters is designated by ν. When an epipole is at infinity, e.g. $\mathbf{e}$, we ensure that it can be parameterized using its second element e_2 by using an appropriate image rotation (which leaves the residual invariant).

case	$\mathbf{e}$	$\mathbf{e}'$	#dof	ν
1	∞	∞	7	$\mathsf{h}, e_1, e_2, e'_1, e'_2$
2	∞	∞	6	$\mathsf{h}, e_1, e_2, e'_2$
3	∞	∞	5	h, e_2, e'_2

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

2.2. 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 via $\widetilde{\mathsf{H}}$. 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:

$$\mathsf{H}_\mathbf{a} \sim \widetilde{\mathsf{H}}[\mathbf{d}]_\times[\mathbf{e}]_\times + \mathbf{e}'\mathbf{a}^\mathsf{T}. \tag{4}$$

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 (4):

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

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.

2.3. 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 coincides with that given in [7, 16].

2.4. Choosing the Best Map

A method to choose the best map for the fundamental matrix is given in [16]. 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. Consequently, it might be very appropriate to use a model selection criterion such as *MDL* or *AIC* [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 $\widetilde{\mathsf{h}}$.

3. Optimal Estimation

In this section, we derive a criterion to compute the optimal structure and motion of a piecewise planar scene. 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*

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 $\mathcal{R}_{\mathsf{H}_j} = \sum_{\{(\mathbf{x}\leftrightarrow\mathbf{x}')\in\pi_j\}} \left(d^2(\mathbf{x},\widehat{\mathbf{x}}) + d^2(\mathbf{x}',\mathsf{H}_j\widehat{\mathbf{x}})\right)$, where π_j denotes the plane of equation $\mathbf{a}_j$ corresponding to the plane homography H_j.

The optimal parameters are obtained as:

$$\{\mathsf{F}, \mathbf{a}_1, \ldots, \mathbf{a}_m, \{\widehat{\mathbf{x}}\}\} = \underset{\nu\cup\{\mathbf{a}_1,\ldots,\mathbf{a}_m\}\cup\{\widehat{\mathbf{x}}\}}{argmin} \sum_{j=1}^{m} \mathcal{R}_{\mathsf{H}_j}, \tag{5}$$

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 $\mathbf{x} \in \pi_1$ and $\mathbf{x} \in \pi_2$, we can not guarantee that $\mathsf{H}_1\widehat{\mathbf{x}} \sim \mathsf{H}_2\widehat{\mathbf{x}}$). The epipolar geometry is implicitly estimated via plane homographies.

Note that this is very different from estimating individual plane homo-

It is also possible to make points $\{\mathbf{y} \leftrightarrow \mathbf{y}'\}$ that do not belong to any plane, contribute to the estimation by adding the residual $\mathcal{R}_{\mathsf{F}} = \sum_{\{\mathbf{y}\leftrightarrow\mathbf{y}'\}} \left(d^2(\mathbf{y}, \widehat{\mathbf{y}}) + d^2(\mathbf{y}', \widehat{\mathbf{y}}')\right)$ to equation (5), plus the constraints $\widehat{\mathbf{y}}'^{\mathsf{T}}\mathsf{F}\widehat{\mathbf{y}} = 0$ so that points satisfy exactly the epipolar geometry.

3.2. Experimental Results

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

The test bench consists of a one meter cube at various distances from two cameras. A number of 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 non-linear minimization of E_3.

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

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

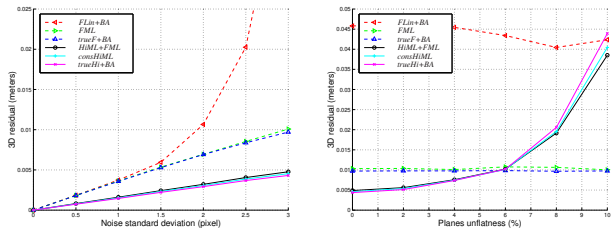
Plane-based estimators, *Methods H*:

- *HiML+FML*: maximum likelihood estimation of plane homographies [5] and then the method *FML*;
- *consHiML*: the consistent approach developed in this paper (equation (5));
- *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 2a) 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.

graphies, and subsequently F from these [8], which is known to be rather unstable. Here, the epipolar geometry and all plane homographies are estimated simultaneously.



(a) (b)

Figure 2. Comparison of the different methods using the *3D* residual E_3, for a distance scene/cameras of 10 meters and versus (a): image noise and (b): planes unflatness under a 3 pixels image noise.

	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.

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 and the size of the simulated planar surface where *Methods H* begin to perform worse than *Methods F*, e.g. for figure 2b, 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.

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 an efficient way of initializing the values of the plane parameters for the *MLE*.

4. Initialization of Plane Equations

In this section, we aim at finding a suitable 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). The choice of an accurate estimator is important since the result will be used to detect and add point correspondences that belong to the observed planes. Moreover, this is essential to prevent the non-linear optimization process of the *MLE* falling into a local optimum.

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.

We have chosen to follow the reduced approach [13] which consists in minimizing a geometric residual in both images. This criterion is known to give stable results and involves only the parameters of the model (here a plane homography) into the optimization whereas the direct approach [13] also involves the data. The drawback of such a criterion is that it is non-linear due to the use of the Euclidean distance and the need to compute the inverse homography. We show that it can be quasi-linearly optimized.

Let us start from the inconsistent criterion corresponding to any *2D* homography for the reduced approach: $\mathsf{H} \sim argmin_{\mathsf{H}} \mathcal{R}(\mathsf{H})$ where the residual $\mathcal{R}$ is given by $\mathcal{R}(\mathsf{H}) = \sum_{\{\mathbf{x}\leftrightarrow\mathbf{x}'\}} \left(d^2(\mathbf{x}', \mathsf{H}\mathbf{x}) + d^2(\mathbf{x}, \mathsf{H}^{-1}\mathbf{x}')\right)$. The scale ambiguity of H is fixed using $||\mathsf{H}||^2 = 1$.

Now, we want to ensure that the result is a plane homography. This can be done by introducing equation (1) in the inconsistent criterion:

$$\mathsf{H} \sim \mathsf{H}_r + \mathbf{e}'(\underset{\mathbf{a}}{argmin}\, \mathcal{R}(\mathsf{H}_r + \mathbf{e}'\mathbf{a}^\mathsf{T}))^\mathsf{T}, \qquad (6)$$

where $\mathbf{a}$ corresponds to the plane equation.

The next step consists in finding an expression for $\mathcal{R}(\mathsf{H})$ that is linear in the unknowns $\mathbf{a}$. In our consistent situation, there exists a linear expression for the inverse homography:

$$\mathsf{H}^{-1} \sim \mathbf{e}\mathbf{a}^\mathsf{T}\mathsf{H}_r^{-1} - (1 + \mathbf{a}^\mathsf{T}\mathbf{e})\mathsf{H}_r^{-1}. \qquad (7)$$

The reader can easily check its validity by doing the product $(\mathsf{H}_r + \mathbf{e}'\mathbf{a}^\mathsf{T})(\mathbf{e}\mathbf{a}^\mathsf{T}\mathsf{H}_r^{-1} - (1+\mathbf{a}^\mathsf{T}\mathbf{e})\mathsf{H}_r^{-1}) \sim \mathrm{I}_3$. Note that $\mathbf{e}' = \mathsf{H}_r\mathbf{e}$ and not just $\mathbf{e}' \sim \mathsf{H}_r\mathbf{e}$ is required and that H_r has to be full-rank. This expression is derived from the Sherman-Morrison non-linear formula [5].

The non-linear distance d can be replaced by the bilinear one d_a, weighted as indicated by equation (3). The two terms inside the sum in the residual $\mathcal{R}$ then become, after some algebraic manipulations:

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

$$d^2(\mathbf{x}, \mathsf{H}^{-1}\mathbf{x}') = w'||\mathsf{S}[\mathbf{x}]_\times \left([[\mathsf{H}_r^{-1}\mathbf{x}']_\times \mathbf{e}]_\times \mathbf{a} - \mathsf{H}_r^{-1}\mathbf{x}'\right)||^2.$$

The residual $\mathcal{R}$ can then be rewritten to form a linear system in terms of $\mathbf{a}$, where each point correspondence accounts for two linearly dependent equations. This allows to iteratively solve for $\mathbf{a}$ using the following quasi-linear optimization process:

1. *initialization:* set $w = w' = 1$ for all matches;
2. *estimation:* estimate $\mathbf{a}$ then H from equation (6) using standard weighted least squares;
3. *weighting:* use H to update weights w and w' according to equation (3);
4. *iteration:* iterate steps 2. and 3. until convergence (see below).

Convergence is achieved when the relative difference between two consecutive residuals vanishes. This estimator can be used with a minimum of three point correspondences.

We have conducted a set of experiments to compare three estimators:

- *NLin*: non-linear minimization of the residual $\mathcal{R}$. The resulting homography does in general not correspond to a world plane;
- *QLinC*: the above quasi-linear algorithm (*C* stands for *Consistent*, i.e. the resulting homography corresponds to a world plane);
- *NLinC*: non-linear minimization of the residual $\mathcal{R}$ with the consistency constraint enforced.

The test bench is the same as in §3. A number of 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. The maximum likelihood estimate of the epipolar geometry is computed from all points and an homography is estimated from the points of only one face of the cube.

A first experiment shows that all methods minimize roughly equally well the residual $\mathcal{R}$ for various noise levels and distances scene/cameras.

We then measure the Frobenius norm between the normalized versions of the estimated homography and the true one. The values are obtained as the median for 100 trials. Figure 3 shows that the quality achieved by the inconsistent method *NLin* is worse compared to those achieved by the consistent ones. The best quality is achieved by the estimator *NLinC*, closely followed by *QLinC*.

Other experiments not shown here permit to say that the quality increases with the number of points or when using the true fundamental matrix instead of an estimated one.

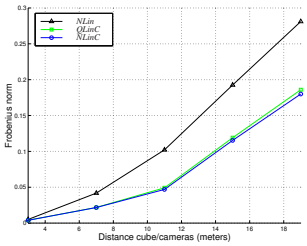


Figure 3. Comparison of the different methods using the Frobenius norm, for different distances scene/cameras and a 1 pixel noise level.

The number of linear systems solved during the optimization process reflects the cost of each algorithm, $[2 \ldots 8]$ for *QLinC* and $[2 \ldots 20]$ for *NLin* and *NLinC*.

The other advantages that make us choose *QLinC* are its easy implementation and its low cost.

5. Results Using Real Images

In this section, we present the reconstruction results that we obtained using the real images of figure 4a. We use the technique given in [16] to obtain a first guess for structure and motion in terms of points and a *RANSAC*-like [4] algorithm to semi-automatically cluster these points into coplanar groups, provided the *LinC* estimator (see §4) for the random sampling and *QLinC* to further refine plane equations. The *MLE* of §3 is then run. Finally, the user provides a delineation of each plane in an image so that texture maps can be extracted and perspectively corrected. Textured renderings are visible in figure 4b and 4c for the perspective and metric models respectively. The latter is obtained via the autocalibration process described in [2].

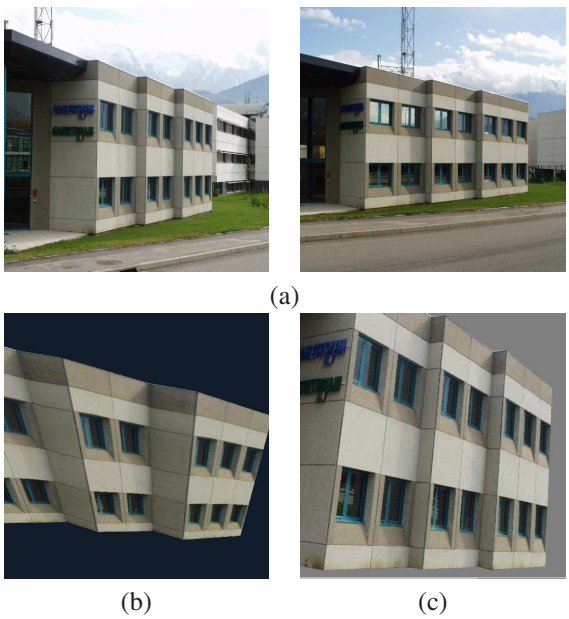


Figure 4. (a)**: the** *Game System* **stereo image pair and textured rendering of the recovered model,** (b)**: projective and** (c)**: metric.**

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 via *2D* homographies constrained to correspond to a world plane. The estimator is quasi-linear and follows the reduced approach [13].

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. Structure and motion from two uncalibrated views using points on planes. In *3DIM*, 2001.

[2] S. Bougnoux. From projective to euclidean space under any practical situation, a criticism of self-calibration. In *ICCV*, 1998.

[3] O. Faugeras and F. Lustman. Motion and structure from motion in a piecewise planar environment. *IJPRAI*, 1988.

[4] M. Fischler, R. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communication of the ACM*, 1981.

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

[6] K. Kanatani. Geometric information criterion for model selection. *IJCV*, 1998.

[7] Q. Luong. *Matrice fondamentale et autocalibration en vision par ordinateur*. PhD thesis, Université de Paris-Sud, France, 1992.

[8] Q. Luong, O. Faugeras. Determining the fundamental matrix with planes: Instability and new algorithms. In *CVPR*, 1993.

[9] Q. Luong and T. Vieville. Canonic representations for the geometries of multiple projective views. *CVIU*, 1996.

[10] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery. *Numerical Recipes in C* . Cambridge University Press, 1992.

[11] R. Szeliski and P. Torr. Geometrically constrained structure from motion : Points on planes. In *SMILE'98*, June 1998.

[12] J.-P. Tarel and J.-M. Vézien. A generic approach for planar patches stereo reconstruction. In *SCIA*, 1995.

[13] B. Triggs. Optimal estimation of matching constraints. In *SMILE'98*, June 1998.

[14] T. Viéville and D. Lingrand. Using singular displacements for uncalibrated monocular visual systems. In *ECCV*, 1996.

[15] G. Xu, J.-I. Terai, H.-Y. Shum. A linear algorithm for camera self-calibration, motion and structure recovery for multi-planar scenes from two perspective images. In *CVPR*, 2000.

[16] Z. Zhang. Determining the epipolar geometry and its uncertainty: A review. *IJCV*, March 1998.