Constrained Structure and Motion From N Views of a Piecewise Planar Scene

Adrien Bartoli, Peter Sturm

VAA'01 - International Symposium on Virtual and Augmented Architecture, Dublin, Ireland, pp. 195-206, June 2001

Constrained Structure and Motion from N Views of a Piecewise Planar Scene

Adrien Bartoli and Peter Sturm

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

Summary. This paper is about the problem of structure and motion recovery from N views of a rigid scene. Especially, we deal with the case of scenes containing planes, i.e. sets of coplanar points and focus on structure and motion estimation, which is one of the major steps of any reconstruction system. In more detail, we explore the strong constraint of multi-coplanarity, exploited only in a sub-optimal manner by most existing works. A typical example is to estimate an unconstrained structure and motion 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 minimizes reprojection error, while satisfying the multi-coplanarity. The estimation problem is consistently parameterized using mixed *2D*/*3D* entities. 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, and more accurate compared to other plane-based methods.

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.). Especially, they are characteristic of architectural scenes. Piecewise planar structures constitute very strong geometric constraints from which we can expect better reconstruction results than from *unconstrained* (i.e. general) structure and motion methods.

Among all the different modules that compose such a specific reconstruction system (image point matching, plane detection, structure and motion recovery, etc.), we focus in this paper on that of *constrained* structure and motion recovery. In other words, we study a *consistent* representation of structure which incorporates the geometric constraints of piecewise planarity. The set of constraints given by a piecewise planar configuration is the *multi-coplanarity* of points, which means that a point is subject to lie on none, one or several planes. Most of the existing works [9,3,11,8] do not take these constraints into account using a specific representation of structure which results in sub-optimal reconstruction estimations.

We propose a projective framework including the *MLE* (Maximum Likelihood Estimator) for constrained structure and motion recovery from N views of a piecewise planar scene, which provides the flexibility of working with uncalibrated or partially calibrated images. This estimator is entirely based on the consistent representation of structure defined in this paper.

The consistency of a representation reflects its property to express exactly the number of essential dof (degrees of freedom) of what it represents. In overparameterized cases, this is achieved provided an appropriate number of additional constraints. In our case, the analysis of the algebraic entities reveals that in general the number of algebraic dof is higher than that of essential dof while the additional constraints can not be easily incorporated, which results in overparameterizations. For example, a coplanarity constraint leaves only 2 dof to a *3D* projective point but is not directly incorporable to the usual homogeneous 4-vector that represents this point.

The same problem arises when representing the motion. For example, the projective motion between two views has 7 essential dof but 9 algebraic ones when expressed via the fundamental matrix. The additional homogeneity and null determinant contraints that might be used can not be incorporated to obtain a unique expression for the fundamental matrix.

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

Previous work: The method proposed in [9] consists in reconstructing planar patches from two images. The structure obtained is not optimal (in the sense of maximum likelihood) since the proposed criterion, photometric- and geometric-based, does not correspond to that of the direct method [10]. Another algorithm proposed in [3] yields sub-optimal results for the same reason.

The solution established in [11] consists in modeling the structure as a general set of *3D* points and in minimizing the residual of the reprojection based on plane homographies induced by scene planes. Points may not satisfy the piecewise coplanarity, even if this constraint is used to estimate the residual.

This constraint is taken into account in [8] where the authors add penalty terms to a criterion close to that of [11]. These penalty terms are devised to make the general recovered structure converge to the desired piecewise planar configuration. The drawbacks of this approach is that the overparameterization leads only to an approximate piecewise planarity and that the penalty terms, heavily weighted, might affect numerical stability.

The major drawback of these works is that the piecewise coplanarity of the scene is not explicitly taken into account using a parameterization of the structure. Moreover, the images used are supposed calibrated.

The method given in [2] is based on geometric primitives more complex than planes (boxes, prisms and surfaces of revolution). Weaknesses of this method are that the set of real scenes that it can deal with is more restricted that when modeling basic planes, and that the recovered structure is not optimal. Moreover, only the minimum number of points if taken into account.

In [1] the parameterization of the structure takes explicitly into account the coplanarity (a point lying on one plane). This allows to devise an optimal estimator for structure and motion. The multi-coplanarity is not taken into account which results in artefacts on the obtained reconstruction.

Contributions, paper organization: The principal result given in this paper is a parameterization of the structure expressing the multi-coplanarity geometric constraints and being therefore consistent. This parameterization is based on mixed *3D/2D* entities. We address the cases where a point is unconstrained (does not lie on any modeled plane), or lies on one, two or three planes. The cases where a point lies on more than three planes are rare so not very interesting to address. We derive the *MLE* for constrained structure and motion and examine experimentally how reconstruction results behave when the observed surfaces are only approximately planar.

We first give some preliminaries in §2, mainly describing the motion parameterization we use. Structure parameterization is described in §3 and the corresponding *MLE* is derived and validated on synthetic data in §4. The global method is then validated using real images in §5.

2 Preliminaries and Notations

We describe the multiple view geometry, from the camera model for one view to the relationship between N views, and how this leads to a motion parameterization. Our main notations are explained throughout this section.

One view: We model a camera using the perspective projection described by a 3×4 homogeneous matrix P. Physical entities (points, planes, etc.) are typeset using italic fonts (X, π, etc.) and their corresponding homogeneous coordinate vectors using the same letters in bold fonts ($\mathbf{X}$, $\boldsymbol{\pi}$, etc.). Homogeneous ($\mathbf{x}$) and inhomogeneous coordinate vectors ($\bar{\mathbf{x}}$) are related by $\mathbf{x}^{\mathsf{T}} \sim (\bar{\mathbf{x}}^{\mathsf{T}} 1)^{\mathsf{T}}$.
A point X in projective 3-space, modeled using an homogeneous 4-vector is projected onto the image plane via $\mathbf{x} \sim \mathsf{P}\mathbf{X}$. The image point x is represented by an homogeneous 3-vector. The notation $\sim$ represents the equality up to a non-null scale factor. The projection matrix P has 11 essential dof.

Two views: The geometry of two views is described by the epipolar geometry, contained in the 3×3 rank-deficient fundamental matrix F. A pair of camera matrices define a projective basis for the reconstruction. Among the multiple pairs of camera matrices compatible with F, one can extract those corresponding to the *canonic projective basis*, given by $\mathsf{P} \sim (\mathrm{I}|\mathbf{0})$ and $\mathsf{P}' \sim (\mathsf{H}_{\mathbf{a}}|\mathbf{e}')$ [5].

The second epipole e' is the projection of the first camera center onto the second image plane and is defined by $\mathsf{F}^\mathsf{T}\mathbf{e}' = \mathbf{0}$ where $^\mathsf{T}$ is the transposition. The matrix $\mathsf{H}_\mathbf{a}$ is any plane homography charaterized by the inhomogeneous 3-vector $\mathbf{a}$ and given by $\mathsf{H}_\mathbf{a} \sim [\mathbf{e}']_\times \mathsf{F} + \mathbf{e}'\mathbf{a}^\mathsf{T}$ where $[.]_\times$ denotes the matrix associated with the cross product, i.e. $[\mathbf{v}]_\times \mathbf{q} = \mathbf{v} \times \mathbf{q}$. Let us define the *canonic plane homography* H for $\mathbf{a} = \mathbf{0}$ by $\mathsf{H} \sim \mathsf{H}_\mathbf{0} \sim [\mathbf{e}']_\times \mathsf{F}$. It can be extracted from any plane homography $\mathsf{H}_\mathbf{a}$ by $\mathsf{H} \sim [\mathbf{e}']_\times^2 \mathsf{H}_\mathbf{a}$ and is characterized by the constraint $\mathsf{H}^\mathsf{T}\mathbf{e}' = \mathbf{0}$ [5]. We use the canonic plane homography in conjunction with the second epipole to parameterize the motion between two views.

Three or more views: Two or more views completely fix the projective basis. Consequently, when adding a supplementary view, in the general case its camera matrix does not have any special form and has to be entirely parameterized.

2.1 Summary of Motion Parameterization

Our motion parameterization, denoted by μ is consistent but not minimal. The two view relation is described by the canonic plane homography and the epipole in the second view (the fundamental matrix can be constructed from these). Each supplementary view is then modeled by a complete perspective projection matrix. This is summarized in table 1. Gauge constraints [10] are necessary to fix the parameters' internal freedoms. Their number is the difference between the number of parameters of the model and the number of essential dof. They are used in the optimization process for the *MLE* described in §4.

#views	#dof	#param.	parameters	gauge constraints
N=2	7	12	H, $\mathbf{e}'$	$\lVert\mathsf{H}\rVert^2 = \lVert\mathbf{e}'\rVert^2 = 1$, $\mathsf{H}^\mathsf{T}\mathbf{e}' = \mathbf{0}$
$N \geq 3$	$7 + 11(N-2)$	$12(N-1)$	$+\mathsf{P}_k$	$+\lVert\mathsf{P}_k\rVert^2 = 1$, $k \in 3 \ldots N$

Table 1. Motion parameterization.

3 Structure Parameterization

In this section, we give our parameterization of the structure. We assume that it is composed of planar patches and model it by a set of planes and a set of points, each one represented according to the number of modeled planes it lies on (from zero up to three).

We first give a representation of planes, in *3D* and then show that minimally representing points on planes can be achieved in *2D*.

3.1 Planes

A plane π is modeled using its equation $\boldsymbol{\pi}$ in the canonical basis. As $\boldsymbol{\pi}$ is an homogeneous 4-vector, an additional constraint is needed to reflect well the 3 dof of the plane and to be therefore a consistent parameterization. A solution is to fix the scale factor of $\boldsymbol{\pi}$ using the additional constraint $||\boldsymbol{\pi}||^2 = 1$.

3.2 Points

Unconstrained points: In our case, unconstrained means that the point X considered does not belong to any modeled plane. Consequently, its number of dof is 3. In the projective space $\mathbb{P}^3$, a general representation of X with only 3 parameters is not possible without restriction. On the other hand, parameterizing X on the image level with 3 parameters is possible. Let us see how this can be done.

Let x and x' be two reprojections of X. They are not independent but related by the epipolar constraint, which emphasizes their 4-1=3 dof. This constraint is represented by a fundamental matrix F between the two images, derived from their projection matrices [5]. Unfortunately, this constraint is bilinear in the points coordinates and can not be used easily to reduce the number of dof of the points.

On the other hand, we can use the interpretation of the epipolar constraint. Indeed x lies on l, its associated epipolar line in the first image, given by $\mathbf{l} \sim [\mathbf{e}]_\times \mathbf{x} \sim \mathsf{F}^\mathsf{T}\mathbf{x}'$. This reduces the number of dof of x from 2 to 1.

The problem is now to parameterize a finite point on a given line, different from the line at infinity, with 1 parameter. This can be done using the construction of figure 1: the point x is represented by its signed distance λ from a reference point b of l. The correspondence between λ and x is then given by:

$$\mathbf{x} \sim \begin{pmatrix} \bar{\mathbf{b}} \\ 1 \end{pmatrix} + \lambda \boldsymbol{\Delta}_l \sim \begin{pmatrix} \bar{\mathbf{b}} + \frac{\lambda}{l_1^2+l_2^2}\begin{pmatrix} l_2 \\ -l_1 \end{pmatrix} \\ 1 \end{pmatrix},$$

where Δ_l is the normalized direction of l given by $\boldsymbol{\Delta}_l = \frac{\mathbf{l} \wedge \mathbf{l}_\infty}{||\mathbf{l} \wedge \mathbf{l}_\infty||}$. The point b can be freely chosen along l. A convenient one is given by the orthogonal projection of o, the center of the image coordinate system, which yields:

$$\mathbf{b} \sim \begin{pmatrix} \bar{\mathbf{o}} - \frac{\mathbf{o}^\mathsf{T}\mathbf{l}}{||\bar{\mathbf{l}}||^2}\bar{\mathbf{l}} \\ 1 \end{pmatrix} \sim \begin{pmatrix} -\frac{l_3}{l_1^2+l_2^2}\begin{pmatrix} l_1 \\ l_2 \end{pmatrix} \\ 1 \end{pmatrix} \text{ where } \mathbf{o}^\mathsf{T} \sim (\bar{\mathbf{o}}, 1)^\mathsf{T} \sim (0,0,1)^\mathsf{T}.$$

An unconstrained point X is represented by its reprojections x and x' into two images. The reprojection x is parameterized on its epipolar line using one parameter, λ and x' using its two image coordinates. Recovering the point X can then be achieved using any triangulation method [4].

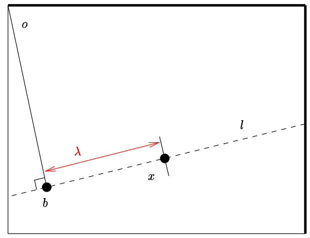

Fig. 1. Principle of parameterizing a point on a line in $\mathbb{E}^2$.

Single-coplanar points: Such a point, constrained to lie on a plane π has 2 dof. Given one of its reprojections x, the corresponding *3D* point X can be recovered using the plane equation $\boldsymbol{\pi}$ and the camera matrix P. Indeed, the following equations hold: $\mathbf{x} \sim \mathsf{P}\mathbf{X}$ and $\boldsymbol{\pi}^\mathsf{T}\mathbf{X} = 0$. This conducts to the following closed-form solution for X:

$$\mathbf{X} \sim \mathsf{A}_\pi^{-1} \begin{pmatrix} \mathbf{x} \\ 0 \end{pmatrix} \text{ where } \mathsf{A}_\pi \sim \begin{pmatrix} \mathsf{P} \\ \boldsymbol{\pi}^\mathsf{T} \end{pmatrix}. \tag{3}$$

Consequently, we choose to represent this kind of point by its reprojection in an image which has the adequat 2 dof. The *3D* point is directly obtained from the projection matrix and the plane equation.

Multi-coplanar points, two planes: Such a point X, constrained to lie on the intersection line of planes π and π' has only 1 dof. Let x be a reprojection in an image, P the corresponding camera matrix and l the reprojection of the intersection line of π and π'. The problem is similar to the unconstrained case: we have to represent an image point lying on an image line with 1 parameter. We choose the same solution as that established previously, i.e. x is represented by its signed distance with a reference point lying on l.

The corresponding *3D* point X is then obtained, as for the previous one plane case, using equation (3) (the result does not depend on which plane π or π' is used to reconstruct).

The line l can be obtained from π and π'. Using equation (3) for both planes, we obtain the equality:

$$\begin{pmatrix} \mathbf{x} \\ 0 \end{pmatrix} \sim \mathsf{B} \begin{pmatrix} \mathbf{x} \\ 0 \end{pmatrix} \forall x \in l \text{ where } \mathsf{B} \sim \mathsf{A}_{\pi'} \mathsf{A}_\pi^{-1} \sim \begin{pmatrix} \mathsf{C}_{3\times 3} & \mathbf{d}_3 \\ \mathbf{c}_3^\mathsf{T} & \alpha \end{pmatrix}.$$

The line l can then be extracted directly from B as $\mathbf{l} \sim \mathbf{c}$. This can be demonstrated using the fact that $\mathsf{B}(\mathbf{x}^\mathsf{T}\ 0)^\mathsf{T} \sim (\mathbf{x}^\mathsf{T}\mathsf{C}^\mathsf{T}\ \mathbf{c}^\mathsf{T}\mathbf{x})^\mathsf{T} \sim (\mathbf{x}^\mathsf{T}\ 0)^\mathsf{T}$ which implies $\mathbf{c}^\mathsf{T}\mathbf{x}{=}0$. Let x' be another point of l.We obtain similarly $\mathbf{c}^\mathsf{T}\mathbf{x}'{=}0$. These two equalities mean that $\mathbf{c} \sim \mathbf{x} \times \mathbf{x}'$ $\square$.

Multi-coplanar points, three planes: Such a point X does not have any dof. Indeed, the intersection of three planes is, in general, a point. Recovering the corresponding *3D* point is straightforward. Let the three planes containing X be π, π' and π''. We have the three equations $\boldsymbol{\pi}^\mathsf{T}\mathbf{X} = \boldsymbol{\pi}'^\mathsf{T}\mathbf{X} = \boldsymbol{\pi}''^\mathsf{T}\mathbf{X} = 0$ which yield $\mathbf{X} = ker(\mathsf{D})$ where $\mathsf{D}^\mathsf{T} \sim (\boldsymbol{\pi}|\boldsymbol{\pi}'|\boldsymbol{\pi}'')$.

3.3 Summary of Structure Parameterization

Our structure parameterization, denoted by ν, relies on mixed *2D/3D* entities. Planes are represented by their equations in the canonic projective basis whereas the representation of points is image-based, depending on the number of planes they lie on. This is summarized in table 2. For the image-

$X \in$	#dof	param.	comments
$\mathbb{P}^3$	3	λ, $\mathbf{x}$	signed distance and *2D* point
π	2	$\mathbf{x}$	*2D* point
π, π'	1	λ	signed distance
π, π', π''	0	$\emptyset$	intersection of three planes

Table 2. Structure parameterization for points.

based parameterization of points, we could have used any image, especially that with camera matrix $\mathsf{P} \sim (\mathrm{I}|\mathbf{0})$ to simplify expressions. However, it is important that the reprojection used for each point is in the image because of numerical stability and to avoid the special case of infinity. Consequently, each point is represented in an image where it is visible (two images for the unconstrained case).

4 Optimal Estimation

In this section, we derive the *MLE* for constrained structure and motion. We first describe how to initialize the previously given parameterization from a general structure and then give details about the *MLE*.

4.1 Constrained Structure Initialization

The initialization is done from a previously estimated unconstrained structure and motion. A clustering of points into multi-coplanar sets is also given.

Planes: A plane is fitted to the points of each coplanar group. Let X be a point of the plane π, the linear constraint $\mathbf{X}^\mathsf{T}\boldsymbol{\pi} = 0$ holds. By stacking the equations for all points lying on π, a linear system is obtained for $\boldsymbol{\pi}$ and solved using an *SVD* [6].

Unconstrained points: Such points are taken directly from the structure obtained previously. Indeed they are not subject to modeled geometric constraint relative to planes.

Single-coplanar points: We use the reprojection of each point in a single view without modification. This introduces a bias which, in general, will be small and without consequence after the final optimization.

Multi-coplanar points, two planes: We project orthogonally the reprojection of X on the reprojection of the line of intersection of the two planes containing X. This gives the initialization of λ.

Multi-coplanar points, three planes: Such a point does not have any dof and consequently, does not need any initialization.

4.2 Maximum Likelihood Estimation

This consists in minimizing the cost function denoted by $\mathcal{C}$ corresponding to the reprojection residual. This is also called the direct approach [10]. The value of the cost function $\mathcal{C}$ depends on measured (automatically detected or user-provided) image points x and on the reprojected structure points $\hat{x}$. It is defined by ($d(.,.)$ is the Euclidean distance):

$$\mathcal{C}(\mu,\nu) = \sum_{i=1}^{N}\sum_{j=1}^{M} w_{ij}\; d^2(x_{ij}, \hat{x}_{ij}),$$

where $w_{ij} = 1$ if and only if the j-th point appears in the i-th image and 0 otherwise and M is the number of points. The structure is extracted from the parameterization $\hat{\nu}$ (see §3.3) and reprojected using camera matrices extracted from $\hat{\mu}$ (see §2.1) to obtain the reprojections $\hat{x}$. The optimal structure and motion are then given by the minimization of $\mathcal{C}$ via the Levenberg-Marquardt algorithm [6] using numerical differentiation: $\{\hat{\mu}, \hat{\nu}\} = \underset{\mu,\nu}{argmin}\, \mathcal{C}(\mu, \nu)$. The image coordinates are normalized so that they lie in $[-1 \ldots 1]$. Gauge constraints on motion (see §2.1) and planes (see §3.1) are enforced at each step of the optimization process.

4.3 Experimental Results Using Simulated Data

We compare the results obtained using our parameterization to those obtained with other existing methods. The test bench consists in a cube of one meter side at different distances of a set of cameras. Points are generated on the cube according to the following distribution: 50 points on each face, 10 points on each edge and 1 point per vertex. These points are projected onto the image planes and a Gaussian centered noise is added.

The quality of a reconstruction X is measured using the *3D* residual of the Euclidean distance to the true one $\underline{X}$: $E_3 = \sqrt{\frac{1}{M}\sum_{j=1}^{M} d^2(\mathsf{H}_3\mathbf{X}, \underline{\mathbf{X}})}$, where

H_3 is a *3D* homography (mapping the projective to the Euclidean structure) estimated using non-linear minimization of E_3. In the following experiments, two cameras have been simulated.

The estimators compared can be divided into two sets. Those that do not use the coplanarity information. Their names begin with *Po-*, which stands for points. The others, using the piecewise planarity, i.e. based on planes and points on plane reconstruction. Their names begin with *Pl-*, which stands for planes. In more detail:

- *Po-ML*: optimal structure and motion [4];
- *Po-trueM*: optimal reconstruction based on a bundle adjustement using the true camera matrices;
- *Pl-wt*: use additional equations to model multi-coplanarity. This is method [8] adapted to a projective framework;
- *Pl-1*: use an explicit structure parameterization to model coplanarity [1];
- *Pl-3*: similar to *Pl-1*, but takes into account multi-coplanarity as described in this paper;
- *Pl-true*π: similar to *Pl-3*, but uses the true plane equations.

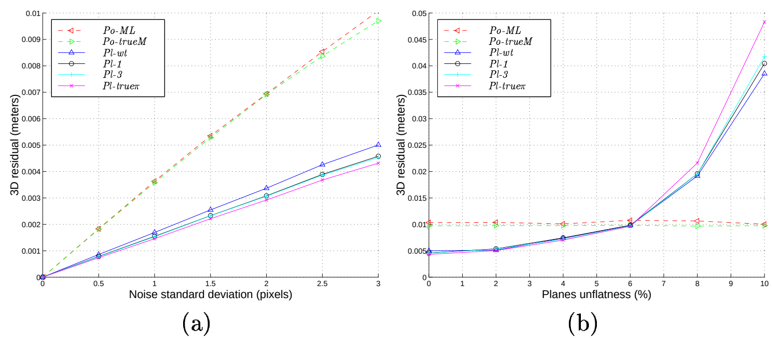

Fig. 2. Comparison of the different methods using the *3D* residual E_3, for perfectly (a) or approximately coplanar points (b).

The first experiment, figure 2a, shows that when points are perfectly coplanar, methods *Po-* based on individual points reconstruction give results of a quality lower than methods *Pl-* modeling also planes. In more detail, we can say that the explicit parameterization of the geometric constraints (methods *Pl-1* and *Pl-3*) improves the results obtained using additional equations (method *Pl-wt*). Taking into account the multi-coplanarity (method *Pl-3*) instead of only the coplanarity (method *Pl-1*) does not significantly improve the results.

The second experiment, figure 2b, shows the results obtained when simulated points are offset vertically from their planes by a random distance

(Gaussian noise with standard deviation between 0 and 0.1 meters). We observe that there is a threshold on the plane unflatness where methods *Pl-* begin to perform worse than methods *Po-*. At this point, it is interesting to measure the ratio between the plane unflatness and the size of the simulated planar surface. This value is called the *breakdown ratio* and is denoted by ε. In the case of figure 2b, ε=6%. Table 3 shows the value of ε established experimentally for different cases. The less stable the configuration is (large noise and/or high distance scene/cameras), the higher is ε, i.e. the more important is the incorporation of piecewise planarity constraints, even if the scene is not perfectly piecewise planar.

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

Table 3. Breakdown ratio ε for different combinations of distance scene/cameras and noise level.

The values of one or several percent in table 3 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 method using piecewise planarity will perform better than any method based on individual point reconstruction.

5 Results Using Real Images

In this section, we present the reconstruction results obtained using the images of figure 3. Similar results have been obtained with other images (on the images shown in [1] for example). We describe the different steps necessary to perform a complete reconstruction, from the images to the *3D* textured model. This process reflects the modular organization of our implementation. The computational time needed to reconstruct such a scene is about five minutes.

Fig. 3. Images (3 out of 4) used to validate the method.

Structure and motion initialization: This has been obtained using image point matches given manually. We perform a partial reconstruction from two images using the method [7] and incrementally add the others to obtain the complete structure and motion. We then run a bundle adjustment to minimize the reprojection error and to obtain the maximum likelihood estimate for an unconstrained structure.

Multi-coplanarity: These relationships are established semi-automatically using plane homographies. The user provides three image points matched in at least one other view to obtain a first guess for the plane. The other points lying on this plane are then automatically detected. The user may interact to correct badly clustered points and add points visible in only one view.

Constrained refinement of structure and motion: From the previous data, the structure is parameterized as described in this paper and the maximum likelihood estimate for constrained structure and motion is computed.

Structure completion: Points appearing in only one view but on at least one modeled plane are automatically reconstructed using equation (3).

Calibration: The metric structure is obtained via a calibration process relying on the definition of a Euclidean base, i.e. the user provides the Euclidean coordinates of five reconstructed points. Alternatively, self-calibration methods could be used, but this is not the topic of this paper.

Texture maps: The texture mapping requires the user to provide a polygonal delineation for each planar facet in one of the images. The texture maps are then extracted and perspectively corrected using calibrated projection matrices. Figure 4 shows different views of the recovered textured model.

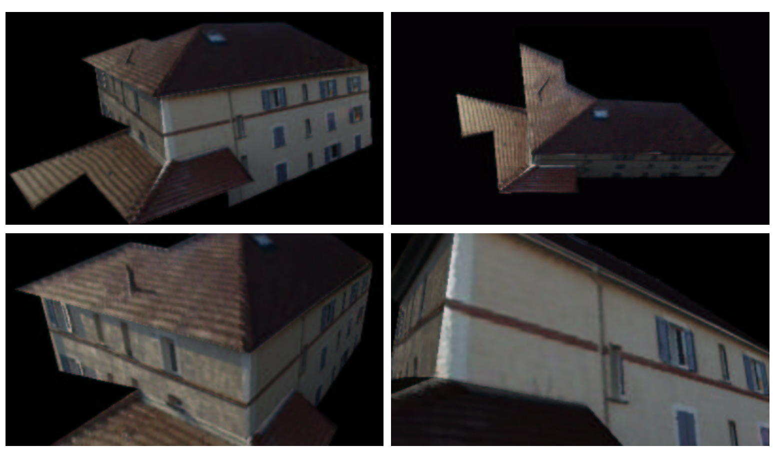

Fig. 4. Different views of the reconstructed model.

6 Conclusions and Perspectives

We have presented an *MLE* for the complete structure and motion from N uncalibrated views of a piecewise planar scene. The structure is consistently represented while incorporating multi-coplanarity relationships, corresponding to the geometric constraints given by the observation of such a scene.

Experimental results on simulated data show that the quality of the reconstruction obtained with plane-based methods is clearly superior to those of methods that reconstruct points individually and that this conclusion is true even if the surfaces are only approximately planar, up to unflatness ratios higher than those of approximately planar world surfaces. Experiments also show that our method improves the accuracy of reconstruction compared to other existing plane-based methods. Real images have been used to validate the approach.

We are currently investigating the complete automatization of the image-based steps of the reconstruction system, i.e. the plane detection and the plane-based image matching. We also plan to use an autocalibration process to upgrade the reconstruction to metric.

References

1. A. Bartoli, P. Sturm, and R. Horaud. Structure and motion from two uncalibrated views using points on planes. In *3DIM*, June 2001. to appear.
2. P.E. Debevec, C.J. Taylor, and J. Malik. Modeling and rendering architecture from photographs: a hybrid geometry-and image-based approach. In SIGGRAPH, August 1996.
3. O. Faugeras and F. Lustman. Motion and structure from motion in a piecewise planar environment. *IJPRAI*, September 1988.
4. R.I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision.* Cambridge University Press, June 2000.
5. Q.T. Luong and T. Vieville. Canonic representations for the geometries of multiple projective views. *CVIU*, 1996.
6. W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. *Numerical Recipes in C.* Cambridge University Press, 1992.
7. P. Sturm and B. Triggs. A factorization based algorithm for multi-image projective structure and motion. In *ECCV*, April 1996.
8. R. Szeliski and P.H.S. Torr. Geometrically constrained structure from motion : Points on planes. In *SMILE*, June 1998.
9. J.-P. Tarel and J.-M. Vézien. A generic approach for planar patches stereo reconstruction. In *SCIA*, 1995.
10. B. Triggs. Optimal estimation of matching constraints. In *SMILE*, June 1998.
11. 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 *CVPR*, June 2000.