Surviving Dominant Planes in Uncalibrated Structure and Motion Recovery

Marc Pollefeys, Frank Verbiest, and Luc Van Gool

Center for Processing of Speech and Images (PSI), K.U.Leuven, Belgium,
firstname.lastname@esat.kuleuven.ac.be

Abstract. In this paper we address the problem of uncalibrated structure and motion recovery from image sequences that contain dominant planes in some of the views. Traditional approaches fail when the features common to three consecutive views are all located on a plane. This happens because in the uncalibrated case there is a fundamental ambiguity in relating the structure before and after the plane. This is, however, a situation that is often hard to avoid in man-made environments. We propose a complete approach that detects the problem and defers the computation of parameters that are ambiguous in projective space (i.e. the registration between partial reconstructions only sharing a common plane and poses of cameras only seeing planar features) till after self-calibration. Also a new linear self-calibration algorithm is proposed that couples the intrinsics between multiple subsequences. The final result is a complete metric 3D reconstruction of both structure and motion for the whole sequence. Experimental results on real image sequences show that the approach yields very good results.

1 Introduction

There has been a lot of progress in uncalibrated structure and motion (USaM) recovery over the last decade. Faugeras [3] and Hartley et al. [9] have shown that starting from an uncalibrated image pair a projective reconstruction was possible. The use of robust statistics for the computation of the epipolar geometry made it possible to obtain good results on real image data [27,21]. These approaches were later extended to image sequences (e.g. [2,15]). In parallel with these developments the possibility to upgrade a projective reconstruction to metric (i.e. Euclidean up to scale) based on constraints on the intrinsics was shown [4]. Over the years many different methods have been proposed for constant [11,16,24] and varying intrinsics [17]. Therefore, starting from an uncalibrated image sequence it became possible to retrieve a metric 3D reconstruction. Compared to the more traditional structure and motion recovery approaches where the camera is calibrated separately, USaM recovery offers an important increase in flexibility.

However, an important –but often ignored– problem of the uncalibrated approach is that it breaks down in the case of a planar scene. The relative pose between views can not be determined when all common features are located in a plane. In fact this is a specific case of the more general problem of critical surfaces (e.g. [12]). However, other cases are much less probable to be encountered in practice. Recently, there has been quite some work on dealing with planes in USaM recovery. Liu et al. [14] and Bartoli et al. [1] looked at architectural scenes containing planes. Note, however, that these techniques

A. Heyden et al. (Eds.): ECCV 2002, LNCS 2351, pp. 837–851, 2002.

require multiple planes or general structure and would therefore fail on the cases treated by this paper. Another interesting approach was proposed by Rother and Carlsson [19]. In this case a linear solution is obtained when a single plane can be seen in all views. Nevertheless, in each view at least two points not located on the plane are required.

In fact, the work that comes closest to solving the problem was carried out by Torr et al. In [23] a robust model selection criterion to differentiate between general 3D structure and planar structure was proposed. This allows to automatically identify the views where the structure is not sufficiently general and to deal with them accordingly (i.e. estimating a homography instead of the epipolar geometry). Although some possibilities were sketched on how this could be used to solve the planarity induced ambiguities in the recovery of USaM, the paper mostly focuses on the model selection and feature tracking issues. No general solution is provided to solve the ambiguity between the structure and motion of subsequences only sharing a single plane.

The main subject of this paper consists of proposing a complete approach to uncalibrated structure and motion recovery that can deal with dominant planes. The approach starts by extending the work by Torr et al. [23] to the 3-view case (which is necessary as will be seen later) so that the difference can be made between subsequences observing sufficiently general structure and subsequences where the tracked features are all located on a single plane. The next step consists of independently recovering the projective structure of the different 3D subsequences. Then the reconstruction for the 3D subsequences is extended with the reconstruction of the planes. Once this is done self-calibration is used to recover the metric structure. To improve the accuracy and robustness the approach couples the intrinsics between the different subsequences. This is especially important to allow successful self-calibration of shorter subsequences. These results are refined using a bundle adjustment that couples the intrinsics for all the subsequences. At this stage a pose estimation algorithm can be used to determine the motion of the camera over the planar parts. The different parts are also assembled (by aligning the overlapping planes). Finally, a global bundle adjustment is carried out to obtain a maximum likelihood estimation of the metric structure and motion for the whole sequence.

The paper is organized as follows. In the next section a traditional uncalibrated structure and motion approach is reviewed. Then, the problem caused by dominant planes is described and the approach for detecting the problem is described. The actual approach to solve it is described in Section 5 (partial projective USaM recovery), Section 6 (coupled self-calibration) and Section 7 (complete metric SaM recovery). In the final sections results and conclusions are presented.

1.1 Notations

Points are represented by homogeneous 4-vectors $\mathtt{M}$ in 3-space, and by homogeneous 3-vectors $\mathtt{m}$ in the image. A plane is represented by a homogeneous 4-vector Π and a point $\mathtt{M}$ is on a plane if $\Pi^\top \mathtt{M} = 0$. A point $\mathtt{M}$ is mapped to its image $\mathtt{m}$ through perspective projection, represented by a 3×4 projection matrix $\mathbf{P}$ as $\mathtt{m} \sim \mathbf{P}\mathtt{M}$. The symbol $\sim$ indicates equality up to a non-zero scale factor. In a metric coordinate system the matrix $\mathbf{P}$ can be factorized in intrinsic and extrinsic camera parameters: $\mathbf{P} = \mathbf{K}[\mathbf{R}\,\mathtt{t}]$ where the upper-triangular matrix $\mathbf{K}$ is given by the following equation:

$$\mathbf{K} = \begin{bmatrix} f & s & u \\ & rf & v \\ & & 1 \end{bmatrix} \tag{1}$$

with f the focal length (measured in pixels), r the aspect ratio, (u, v) the coordinates of the principal point and s a factor that is zero when the pixels are rectangular. To deal with radial distortion, the perspective projection model is extended to $\mathbf{K}\mathcal{R}([\mathbf{R}\,\mathtt{t}]\mathtt{M})$ with $\mathcal{R}([x\,y\,1]^\top) \sim [x\,y\,w]^\top$ and $w^{-1} = (1 + k_1 r^2 + k_2 r^4)$ and $r^2 = x^2 + y^2$ and k_1 and k_2 are parameters of radial distortion. The fundamental matrix $\mathbf{F}$ and the two image homography $\mathbf{H}$, are both 3×3 homogeneous matrices. A point $\mathtt{m}$ located in the plane corresponding to the homography $\mathbf{H}$ is transferred from one image to the other according to $\mathtt{m}' \sim \mathbf{H}\mathtt{m}$. A more complete description of these concepts can be found in [10].

2 General Projective Structure and Motion Recovery

Starting from an uncalibrated image sequence the first step consists of relating the different images to each other. This is not an easy problem. In general a restricted number of corresponding points is sufficient to determine the epipolar geometry between the images. Since not all points are equally suited for matching (e.g. pixels in a homogeneous region), the first step consist of selecting feature points [8] that are suited for automated matching. Features of consecutive views are compared and a number of potential correspondences are obtained. From these the epipolar geometry can be computed. However, the initial set of corresponding points is typically contaminated with an important number of outliers. In this case, a traditional least-squares approach will fail and therefore a robust method is used [21,27,5]. Once the epipolar geometry has been obtained it can be used to guide the search for additional correspondences. These can then in turn be used to further refine the epipolar geometry.

The relation between the views and the correspondences between the features can then be used to retrieve the structure of the scene and the motion of the camera. The approach that is used is related to [2] but is fully projective and therefore not dependent on any approximation. This is achieved by strictly carrying out all measurements in the images, i.e. using only reprojection errors. At first two images are selected and an initial projective reconstruction frame is set-up [3,9]. Then the pose of the camera for the other views is determined in this frame and for each additional view the initial reconstruction is refined and extended. Once the structure and motion has been determined for the whole sequence, the results are refined through a projective bundle adjustment [26]. To minimize the presence of a consistent bias in the reconstruction, this bundle adjustment takes into account radial distortion (around the image center). Then the ambiguity is restricted to metric through self-calibration. A modified version of [17] is used (see Section 6). Finally, a metric bundle adjustment is carried out to obtain an optimal estimation for both structure and motion.

3 Problems with Planes

The projective structure and motion approach described in the previous section assumes that both motion and structure are general. When this is not the case, the approach can fail.

In the case of motion this will happen when the camera is purely rotating. A solution to this problem was proposed in [23]. Here we will assume that care is taken during acquisition to not take multiple images from the same position so that this problem doesn't occur[1].

Scene related problems occur when (part of) the scene is purely planar. In this case it is not possible anymore to determine the epipolar geometry uniquely. If the scene is planar, the image motion can be fully described by a homography. Since $\mathbf{F} = [\mathsf{e}']_\times \mathbf{H}$ (with $[\mathsf{e}']_\times$ the vector product with the epipole e'), there is a 2 parameter family of solutions for the epipolar geometry. In practice robust techniques would pick a random solution based on the inclusion of some outliers.

Assuming we would be able to detect this degeneracy, the problem is not completely solved yet. Obviously, the different subsequences containing sufficient general 3D structure could be reconstructed separately. The structure of subsequences containing only a single plane could also be reconstructed as such. These planar reconstructions could then be inserted into the neighboring 3D projective reconstructions. However, there remains an ambiguity on the transformation relating two 3D projective reconstruction only sharing a common plane. The plane shared by the two reconstructions can be uniquely parameterized by three 3D points (3×3 parameters) and a fourth point in the plane (2 free parameters) to determine the projective basis within the plane. The ambiguity therefore has 15-11=4 degrees of freedom. An illustration is given on the left side of Figure 1. Note also that it can be very hard to avoid this type of degeneracy as can be seen from the right side of Figure 1. Many scenes have a configuration similar to this one.

4 Detecting Dominant Planes

The first part of the solution consists of detecting the cases where only planar features are being matched. The Geometric Robust Information Criterion (GRIC) model selection approach proposed in [22] is briefly reviewed. The GRIC selects the model with the lowest score. The score of a model is obtained by summing two contributions. The first one is related to the goodness of the fit and the second one is related to the parsimony of the model. It is important that a robust Maximum Likelihood Estimator (MLE) be used for estimating the different structure and motion models being compared through GRIC. GRIC takes into account the number n of inliers plus outliers, the residuals e_i, the standard deviation of the measurement error σ, the dimension of the data r, the number k of motion model parameters and the dimension d of the structure:

$$\text{GRIC} = \sum \rho(e_i^2) + (nd \ln(r) + k \ln(rn)) \quad . \tag{2}$$

where $\rho(e^2)$

$$\rho(e^2) = \min\left(\frac{e^2}{\sigma^2}, 2(r-d)\right) \quad . \tag{3}$$

[1] Note that the approach would still work if the pure rotation takes place while observing a planar part.

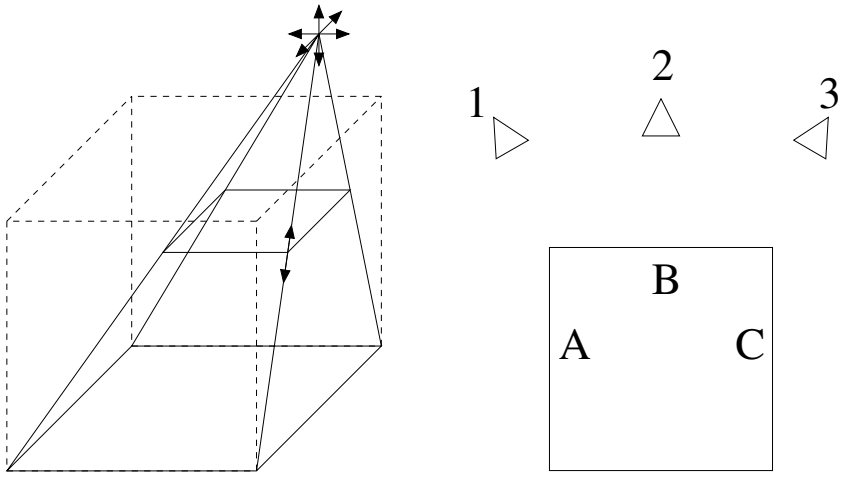

Fig. 1. Left: Illustration of the four-parameter ambiguity between two projective reconstructions sharing a common plane. If the base of the cube is shared, a projective transformation can still affect the height of the cube and the position of the third vanishing point. Right: A fundamental problem for many (man-made) scenes is that it is not possible to see A,B and C at the same time and therefore when moving from position 1 to position 3 the planar ambiguity problem will be encountered.

In the above equation $nd\ln(r)$ represents the penalty term for the structure having n times d parameters each estimated from r observations and $k\ln(rn)$ represents the penalty term for the motion model having k parameters estimated from rn observations.

For each image pair GRIC($\mathbf{F}$) and GRIC($\mathbf{H}$) can be compared. If GRIC($\mathbf{H}$) yields the lowest value it is assumed that most matched features are located on a dominant plane and that a homography model is therefore appropriate. On the contrary, when GRIC($\mathbf{F}$) yields the lowest value one could assume, as did Torr [23], that standard projective structure and motion recovery could be continued. In most cases this is correct, however, in some cases this might still fail. An illustration of the problem is given on the left side of Figure 2 where both $\mathbf{F}_{12}$ and $\mathbf{F}_{23}$ could be successfully computed, but where structure and motion recovery would fail because all features common to the three views are located on a plane. Estimating the pose of camera 3 from features reconstructed from views 1 and 2 or alternatively estimating the trifocal tensor from the triplets would yield a three-parameter family of solutions. However, imposing reconstruction 1–2 and reconstruction 2–3 to be aligned (including the center of projection for view 2) would reduce the ambiguity to a one-parameter family of solutions. This ambiguity is illustrated on the right side of Figure 2. Compared to the reference frame of cameras 1 and 2 the position of camera 3 can change arbitrarily as long as the epipole in image 2 is not modified (i.e. motion along a line connecting the center of projections of image 2 and 3). Since intersection has to be preserved and the image of the common plane also has to be invariant, the transformation of the rest of space is completely determined. Note –as seen in Figure 2– that this remaining ambiguity could still cause an important distortion.

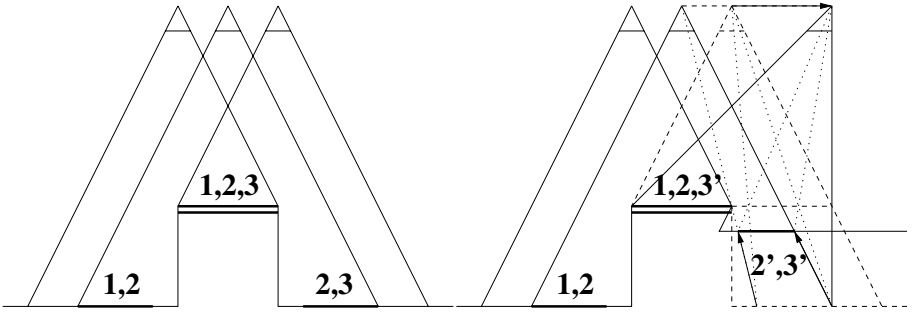

Fig. 2. Left: Although each pair contains non-coplanar features, the three views only have coplanar points in common. Right: Illustration of the remaining ambiguity if the position of the center of projection for view 2 corresponds for structure 1–2 and 2–3.

For the reason described above we propose to use the GRIC criterion on triplets of views ($r = 6$). On the one hand we have GRIC(**PPP**) based on a model containing 3 projection matrices (up to a projective ambiguity) with $k = 3 \times 11 - 15 = 18$ and $d = 3$ (note that using a model based on the trifocal tensor would be equivalent), on the other hand we have GRIC(**HH**) based on a model containing 2 homographies with $k = 2 \times 8 = 16$ and $d = 2$. To efficiently compute the MLE of both **PPP** and **HH** the sparse structure of the problem is exploited (similar to bundle adjustment). We can now differentiate between two different cases: Case A: GRIC(**PPP**)<GRIC(**HH**): three views observe general 3D structure. Case B: GRIC(**PPP**)>GRIC(**HH**): common structure between three views is planar. Note that it does not make sense to consider mixed cases such as **HF** or **FH** since for structure and motion recovery triplets are needed which in these cases would all be located on a plane anyway.

Note that in addition, one should verify that a sufficient number of triplets remain (say more than 50) to allow a reliable estimation. When too few points are seen in common over three views, the sequence is also split up. In a later stage it can be reassembled (using the procedure laid out in Section 7). This avoids the risk of a (slight) change of projective basis due to an unreliable estimation based on too few points. Note that it is important to avoid this, since this would mean that different transformations would be required to bring the different parts of the recovered structure and motion back to a metric reference frame. In practice this causes self-calibration to fail and should therefore be avoided.

5 Partial Projective Structure and Motion Recovery

The sequence is first traversed and separated in subsequences. For subsequences with sufficient 3D structure (case A) the approach described in Section 2 is followed so that the projective structure and motion is recovered. When a triplet corresponds to case B, only planar features are tracked and reconstructed (in 2D). A possible partitioning of an image sequence is given in Table 1. Note that the triplet 3-4-5 would cause an approach based on [23] to fail.

Table 1. Example on how a sequence would be partitioned based on the different cases obtained in the model selection step. Underlined F correspond to cases that would not be dealt with appropriately using a pairwise analysis.

case	AABAABBBBBAAA
3D	PPPP PPPPP
2D	HH HHHHHH
3D	PPPP
	FFF̲F̲FFHHHHFFFF

Suppose the plane Π is labeled as a dominant plane from view i based on features tracked in views $(i-1, i, i+1)$. In general, some feature points $\mathtt{M}_\Pi$ located on Π will have been reconstructed in 3D from previous views (e.g. i and $(i-1)$). Therefore, the coefficients of Π can be computed from $\mathtt{M}_\Pi^\top \Pi = 0$. Define $\mathbf{M}_\Pi$ as the right null space of $\Pi^\top$ (4×3 matrix). $\mathbf{M}_\Pi$ represents 3 supporting points for the plane Π and let $\mathbf{m}_{\Pi\mathtt{i}} = \mathbf{P}_i \mathbf{M}_\Pi$ be the corresponding image projections. Define the homography $\mathbf{H}_{i\Pi} = \mathbf{m}_{\Pi\mathtt{i}}^{-1}$, then the 3D reconstruction of image points located in the plane Π are obtained as follows:

$$\mathtt{M}_i = \mathbf{M}_\Pi \mathbf{H}_{i\Pi} \mathtt{m}_i \tag{4}$$

Similarly, a feature $\mathtt{m}_j$ seen in view $j(> i)$ can be reconstructed as:

$$\mathtt{M}_j = \mathbf{M}_\Pi \mathbf{H}_{i\Pi} (\mathbf{H}_{ij}^\Pi)^{-1} \mathtt{m}_j \tag{5}$$

where $\mathbf{H}_{ij}^\Pi = \mathbf{H}_{i(i+1)}^\Pi \cdots \mathbf{H}_{(j-1)j}^\Pi$.

6 Coupled Self-Calibration

Once the projective structure and motion has been computed for each subsequence, standard self-calibration approaches could be used on the subsequences. However, some of these could be too short to obtain good results.

In this section a self-calibration approach is proposed that couples the camera intrinsics for the different subsequences containing general 3D structure. The approach is based on the approach proposed in [17], but was adapted to better reflect a priory expectations for the unknowns. The approach is based on the projection equation for the absolute quadric [24]:

$$\mathbf{K}\mathbf{K}^\top \sim \mathbf{P}\boldsymbol{\Omega}^*\mathbf{P}^\top \tag{6}$$

where $\boldsymbol{\Omega}^*$ represents the absolute quadric. In metric space $\boldsymbol{\Omega}^* = \mathrm{diag}(1,1,1,0)$, in projective space $\boldsymbol{\Omega}^*$ is a 4×4 symmetric rank 3 matrix representing an imaginary disc-quadric. By transforming the image so that a typical focal length (e.g. 50mm) corresponds to unit length in the image and that the center of the image is located at the origin, realistic expectations for the intrinsics are $\log(f) = \log(1) \pm \log(3)$ (i.e. f is typically in the range $[17\mathrm{mm}, 150\mathrm{mm}]$), $r = log(1) \pm \log(1.1), u = 0 \pm 0.1, v = 0 \pm 0.1, s = 0$. These expectations can be used to obtain a set of weighted self-calibration equations

from Equation (6):

$$\begin{array}{ll} \frac{1}{9\nu}\left(P_1\mathbf{\Omega}^*{P_1}^\top - P_3\mathbf{\Omega}^*{P_3}^\top\right) = 0 & \frac{1}{0.01\nu}\left(P_1\mathbf{\Omega}^*{P_2}^\top\right) = 0 \\ \frac{1}{9\nu}\left(P_2\mathbf{\Omega}^*{P_2}^\top - P_3\mathbf{\Omega}^*{P_3}^\top\right) = 0 & \frac{1}{0.1\nu}\left(P_1\mathbf{\Omega}^*{P_3}^\top\right) = 0 \\ \frac{1}{0.2\nu}\left(P_1\mathbf{\Omega}^*{P_1}^\top - P_2\mathbf{\Omega}^*{P_2}^\top\right) = 0 & \frac{1}{0.1\nu}\left(P_2\mathbf{\Omega}^*{P_3}^\top\right) = 0 \end{array} \tag{7}$$

where P_i is the i-th row of a projection matrix and ν a scale factor that is initially set to 1 and later on to $P_3\tilde{\mathbf{\Omega}}^*{P_3}^\top$ with $\tilde{\mathbf{\Omega}}^*$ the result of the previous iteration. In practice iterating is not really necessary, but a few iterations can be performed to refine the initial result. Experimental validation has shown that this approach yields much better results than the original approach described in [17]. This is mostly due to the fact that constraining all parameters (even with a small weight) allows to avoid most of the problems due to critical motion sequences [20,13] (especially the specific additional case for the linear algorithm [18]).

When choosing $\mathbf{P} = [\mathbf{I}|\mathbf{0}]$ for one of the projection matrices it can be seen from Equation (6) that $\mathbf{\Omega}^*$ can be written as:

$$\mathbf{\Omega}^* = \begin{bmatrix} \mathbf{K}\mathbf{K}^\top & \mathbf{a} \\ \mathbf{a}^\top & b \end{bmatrix} \tag{8}$$

Now the set of equations (7) can thus be written as:

$$\begin{bmatrix} \mathbf{C} & \mathbf{D} \end{bmatrix} \begin{bmatrix} \mathbf{k} \\ \begin{bmatrix} \mathbf{a} \\ b \end{bmatrix} \end{bmatrix} \tag{9}$$

where $\mathbf{k}$ is a vector containing 6 coefficients representing the matrix $\mathbf{K}\mathbf{K}^\top$, $\mathbf{a}$ is a 3-vector and b a scalar and $\mathbf{C}$ and $\mathbf{D}$ are matrices containing the coefficients of the equations. Note that this can be done independently for every 3D subsequence.

If the sequence is recorded with constant intrinsics, the vector $\mathbf{k}$ will be common to all subsequences and one obtains the following coupled self-calibration equations:

$$\begin{bmatrix} \mathbf{C}_1 & \mathbf{D}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{C}_2 & \mathbf{0} & \mathbf{D}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{C}_n & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{D}_n \end{bmatrix} \begin{bmatrix} \mathbf{k} \\ \begin{bmatrix} \mathbf{a}_1 \\ b_1 \end{bmatrix} \\ \begin{bmatrix} \mathbf{a}_2 \\ b_2 \end{bmatrix} \\ \vdots \\ \begin{bmatrix} \mathbf{a}_n \\ b_n \end{bmatrix} \end{bmatrix} \tag{10}$$

As will be seen in the experiments this approach is very successful. The most important feature is that through the coupling it allows to get good results even for the shorter subsequences. For each subsequence a transformation to upgrade the reconstruction from projective to metric can be obtained from the constraint $\mathbf{T}_i\mathbf{\Omega}_i^*\mathbf{T}_i^\top = \mathrm{diag}(1,1,1,0)$ (through Cholesky factorization). This result is then further refined through a metric bundle adjustment that also couples the intrinsics of the different subsequences.

7 Combined Metric Structure and Motion Recovery

Now that the metric structure of the subsequences has been recovered, the pose of the camera can also be determined for the viewpoints observing only planar points. Since the intrinsics have been computed, a standard pose estimation algorithm can be used. We use Grunert's algorithm as described in [7]. To deal with outliers a robust approach was implemented [5].

Finally, it becomes possible to align the structure and motion recovered for the separate subsequences based on common points. Note that these points are all located in a plane and therefore some precautions have to be taken to obtain results using linear equations. However, since 3 points form a basis in a metric 3D space, additional points out of the plane can easily be generated (i.e. using the vector product) and used to compute the relative transform using linear equations. Here again a robust approach is used.

Now that all structure and motion parameters have been estimated for the whole sequence. A final bundle adjustment is carried out to obtain a globally optimal solution.

Fig. 3. Some of the 64 images of the *corner* sequence.

Figure 5 shows results for different stages of our approach. At the top-left the recovered metric structure and motion for the two subsequences that contain sufficiently general structure is given (after coupled self-calibration). Then, both structure and motion are extended over the planar parts. This can be seen in the middle-left part of the figure. At the bottom-left the complete structure and motion for the whole sequence is shown after bundle adjustment. On the right side of the figure orthogonal top and front views are shown.

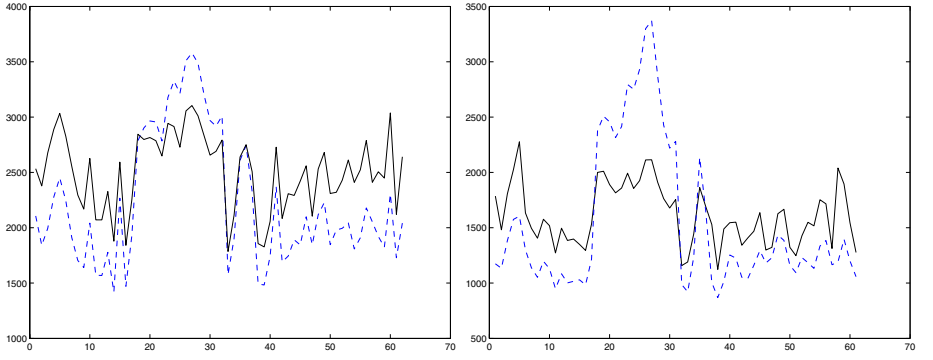

Fig. 4. Left: GRIC(**F**) (solid/black line) and GRIC(**H**) (dashed/blue line). Right: GRIC(**PPP**) (solid/black line) and GRIC(**HH**) (dashed/blue line).

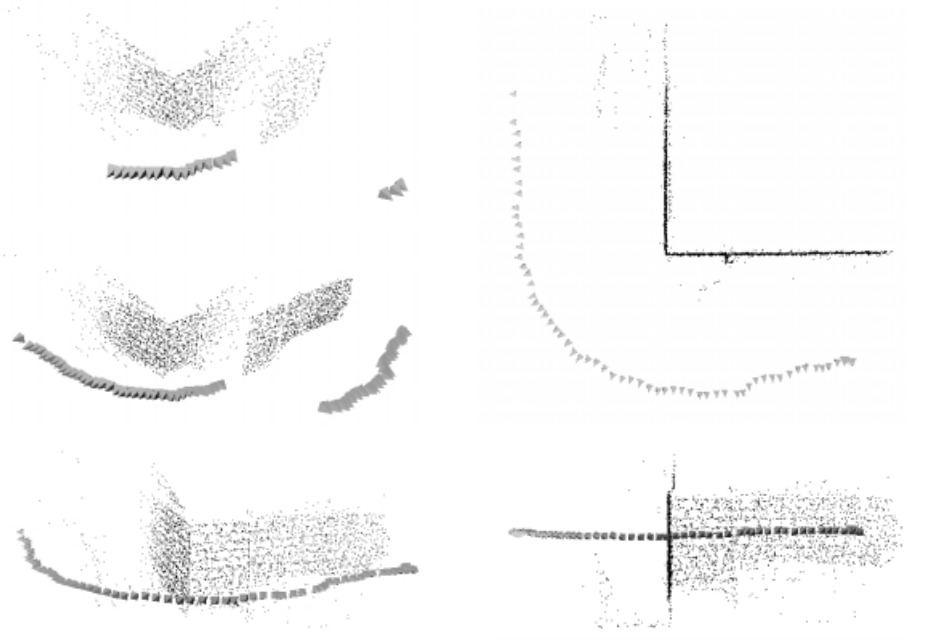

Fig. 5. Left: different stages of the structure and motion recovery, Right: orthogonal views of the final result.

8 Results

In this section results of our approach on two real image sequences are shown. The first image sequence was recorded from a corner of our institute. The *corner* sequence contains 64 images recorded using a Sony TRV900 digital camcorder in progressive scan mode. The images therefore have a resolution of 720×576 (PAL). Some of the images are shown in Figure 3. Note that the images contain quite some radial distortion.

In Figure 4 the GRIC values are given for $\mathbf{F}$ and $\mathbf{H}$ as well as for $\mathbf{PPP}$ and $\mathbf{HH}$. It can clearly be seen that –besides dealing with additional ambiguities– the triplet based analysis in general provides more discriminant results. It is also interesting to note that triplet 34-35-36 is clearly indicated as containing sufficiently general structure for the triplet-based approach while the pair-based approach marginally prefers to use the plane based model. The USaM approach reconstructs the structure for this triplet (including some points seen in the background of the lower left picture of Figure 3) and successfully integrates them with the rest of the recovered structure and motion.

The second sequence consists of 150 images of an old *farmhouse*. It was recorded with the same camera as the first sequence. In Figure 6 the GRIC values are plotted and for some of them the corresponding images are shown. As can be seen the approach successfully discriminates between the planar parts and the others. In Figure 7 the computed

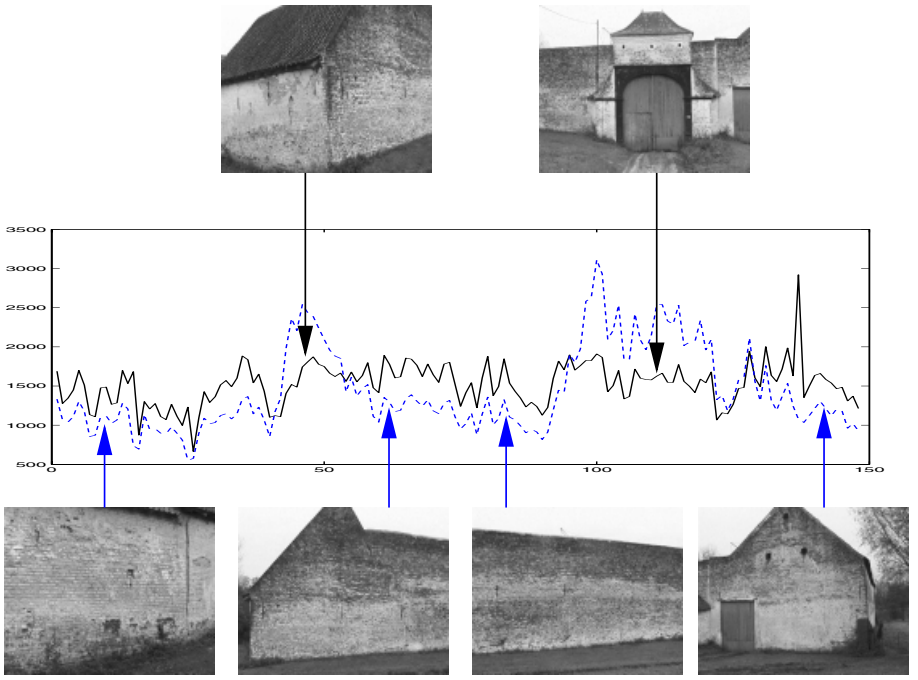

Fig. 6. Some images of the *farmhouse* sequence together with GRIC(**PPP**) (solid/black line) and GRIC(**HH**) (dashed/blue line).

structure and motion is shown. In Figure 8 some views of a dense textured 3D model are shown. This model was obtained by computing some depth maps using a stereo algorithm and the obtained metric structure and motion. Note that the whole approach from image sequence to complete 3D model is fully automatic.

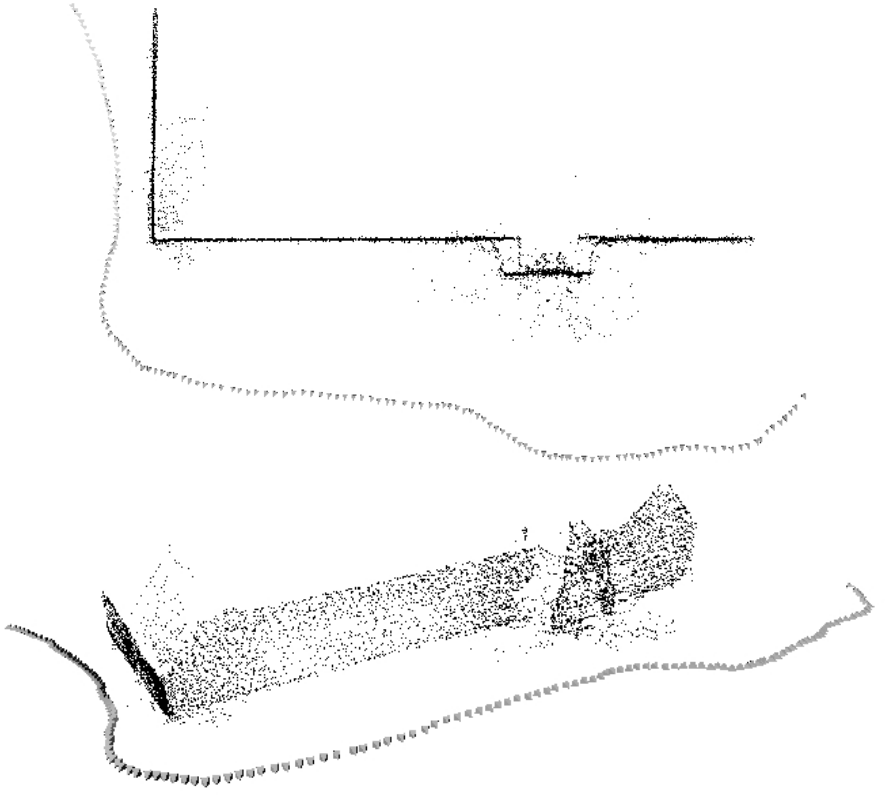

Fig. 7. Combined structure and motion for the whole *farmhouse* sequence.

9 Conclusion

In this paper we have presented an approach that successfully deals with dominant planes in uncalibrated structure and motion recovery. This is an important problem that limited the practical applicability of uncalibrated approaches, especially in man-made environments. The solution proposed in this paper yields very good results on real image sequences. The approach uses the Geometric Robust Information Criterion to detect if features seen in common by three views are all in a plane. Subsequences containing sufficiently general structure are reconstructed and then extended with the planar

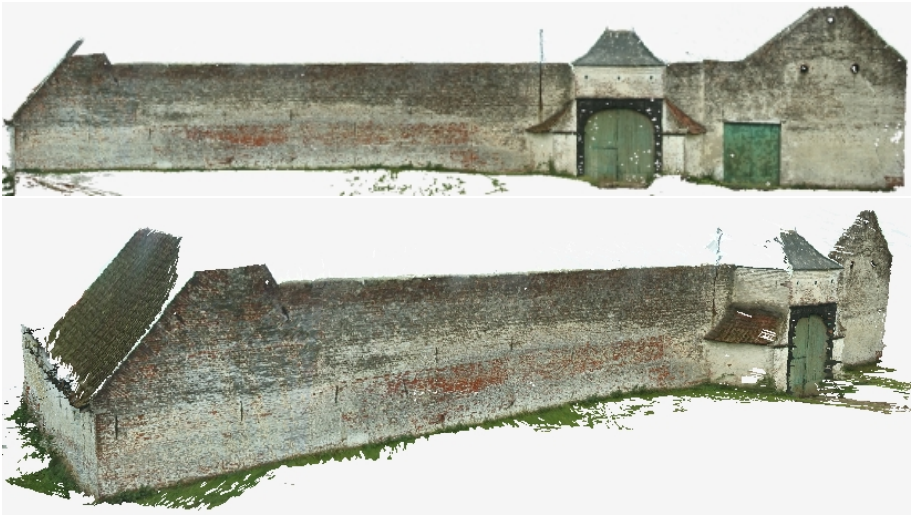

Fig. 8. Textured 3D model of the *farmhouse*

parts. A new linear self-calibration algorithm couples the intrinsics between the different subsequences so that even for short sequences good results can be obtained. Once the reconstruction has been upgraded to metric, the pose is estimated for the cameras observing planar parts and the reconstructions for the different subsequences are assembled. Finally a global bundle adjustment provides an optimal estimate of both structure and motion. A key factor for the success of the proposed approach is the consistent use of robust maximum likelihood estimation through efficient bundle adjustment and robust estimation (i.e. RANSAC) at most of the stages of the computations.

Acknowledgment. Marc Pollefeys is a post-doctoral fellow of the Fund for Scientific Research - Flanders (Belgium). The financial support of the FWO project G.0223.01 and the IST projects ATTEST and INVIEW are also gratefully acknowlegded.

References

1. A. Bartoli and P. Sturm, "Constrained Structure and Motion from N Views of a Piecewise Planar Scene", *VAA'01 - In Proceedings of the International Symposium on Virtual and Augmented Architecture*, Dublin, Ireland, pp. 195-206, June 2001.
2. P. Beardsley, P. Torr, and A. Zisserman. "3D model acquisition from extended image sequences". In *Proc. European Conf. on Computer Vision*, LNCS 1064, Vol. 2, Springer-Verlag, pages 683–695, 1996.
3. O. Faugeras, "What can be seen in three dimensions with an uncalibrated stereo rig", *Computer Vision - ECCV'92*, Lecture Notes in Computer Science, Vol. 588, Springer-Verlag, pp. 563-578, 1992.
4. O. Faugeras, Q.-T. Luong and S. Maybank. "Camera self-calibration: Theory and experiments", *Computer Vision - ECCV'92*, Lecture Notes in Computer Science, Vol. 588, Springer-Verlag, pp. 321-334, 1992.

5. M. Fischler and R. Bolles, "RANdom SAmpling Consensus: a paradigm for model fitting with application to image analysis and automated cartography", *Commun. Assoc. Comp. Mach.*, 24:381-95, 1981.
6. A. Fitzgibbon and A. Zisserman, "Automatic camera recovery for closed or open image sequences", *Computer Vision – ECCV'98*, vol.1, Lecture Notes in Computer Science, Vol. 1406, Springer-Verlag, 1998. pp.311-326, 1998.
7. R. Haralick, C. Lee, K. Ottenberg, and M. Nolle, "Review and Analysis of Solutions of the Three Point Perspective Pose Estimation Problem", *International Journal of Computer Vision*, Vol.13, No.3, 1994, pp. 331-356.
8. C. Harris and M. Stephens, "A combined corner and edge detector", *Fourth Alvey Vision Conference*, pp.147-151, 1988.
9. R. Hartley, R. Gupta, T. Chang, "Stereo from uncalibrated cameras". In *Proc. Conf. on Computer Vision and Pattern Recognition*, 1992.
10. R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.
11. A. Heyden and K. Åström, "Euclidean Reconstruction from Constant Intrinsic Parameters" *Proc. 13th International Conference on Pattern Recognition*, IEEE Computer Soc. Press, pp. 339-343, 1996.
12. W. Hofmann. *Das problem der "Gefährlichen Flächen" in Theorie und Praxis -Ein Beitrag zur Hauptaufgabe der Photogrammetrie*. PhD Thesis, Fakultät für Bauwesen, Technische Universityät München, Germany, 1953.
13. F. Kahl, B. Triggs, K. Åström, "Critical Motions for Auto-Calibration When Some Intrinsic Parameters Can Vary", *Journal of Mathematical Imaging and Vision* 13,131-146,2000.
14. Y. Liu, H.-T. Tsui and A. Heyden, "3D Reconstruction of Buildings from an Uncalibrated Image Sequence – A Scene Based Strategy", *Proc. Virtual and Augmented Architecture (VAA'01)*, pp. 231–242, Springer-Verlag, 2001.
15. D. Nister, *Automatic Dense Reconstruction from Uncalibrated Video Sequences*, Ph. D. dissertation, Dept. of Numerical Analysis and Computing Science, KTH Stockholm, 2001.
16. M. Pollefeys and L. Van Gool, "Stratified Self-Calibration with the Modulus Constraint", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol 21, No.8, pp.707-724, 1999.
17. M. Pollefeys, R. Koch and L. Van Gool. "Self-Calibration and Metric Reconstruction in spite of Varying and Unknown Internal Camera Parameters", *International Journal of Computer Vision*, 32(1), 7-25, 1999.
18. M. Pollefeys, *Self-calibration and metric 3D reconstruction from uncalibrated image sequences*, Ph.D. Thesis, ESAT-PSI, K.U.Leuven, 1999.
19. C. Rother and S. Carlsson. "Linear Multi View Reconstruction and Camera Recovery", Proc. *Eight IEEE International Conference on Computer Vision*, pp. 42-49 ,2001.
20. P. Sturm. "Critical Motion Sequences for Monocular Self-Calibration and Uncalibrated Euclidean Reconstruction", *IEEE Conference on Computer Vision and Pattern Recognition*, pp.1100-1105, 1997.
21. P. Torr, *Motion Segmentation and Outlier Detection*, PhD Thesis, Dept. of Engineering Science, University of Oxford, 1995.
22. P. Torr. "An assessment of information criteria for motion model selection". In *CVPR97*, pages 47–53, 1997.
23. P. Torr, A. Fitzgibbon and A. Zisserman, "The Problem of Degeneracy in Structure and Motion Recovery from Uncalibrated Image Sequences", *International Journal of Computer Vision*, vol. 32, no. 1, pages 27-44, August, 1999.
24. B. Triggs, "The Absolute Quadric", *Proc. 1997 Conference on Computer Vision and Pattern Recognition*, IEEE Computer Soc. Press, pp. 609-614, 1997.

25. B. Triggs, "Autocalibration from planar scenes", *Computer Vision – ECCV'98*, vol.1, Lecture Notes in Computer Science, Vol. 1406, Springer-Verlag, pp 89-105, 1998.
26. B. Triggs, P. McLauchlan, R. Hartley, and A. Fitzgibbon. "Bundle adjustment – A modern synthesis". In B. Triggs, A. Zisserman, and R. Szeliski, editors, *Vision Algorithms: Theory and Practice*, LNCS, pages 298–375. Springer Verlag, 2000.
27. Z. Zhang, R. Deriche, O. Faugeras and Q.-T. Luong, "A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry", *Artificial Intelligence Journal*, Vol.78, pp.87-119, October 1995.