Projective Factorization of Planes and Cameras in Multiple Views

Carsten Rother, Stefan Carlsson and Dennis Tell
Computational Vision and Active Perception Laboratory (CVAP)
Royal Institute of Technology (KTH), Stockholm, Sweden
{carstenr, stefanc, dennis}@nada.kth.se

Abstract

This paper proposes a novel method for the projective reconstruction of planes and cameras from multiple images by factorizing a matrix containing all planar homographies between a reference view and all other views. If some planes are not visible in all views an alternative method is presented which solves the problem in two steps: a) all camera centers are recovered simultaneously b) all planes are reconstructed. The key idea of both methods is to specify one of the planes, which is visible in all views, as the plane at infinity. The methods were applied to synthetic and real data, where VRML models can be created with a small amount of user interaction.

1 Introduction

A major property of a man-made environment are planes: buildings, walls, floors, streets, furniture, etc. The recovery of planes and cameras from an image sequence of such an environment could be utilized in a wealth of applications, e.g. image-based modeling and rendering, robotics. This problem can be addressed in two different ways: a) reconstruct 3D-features, e.g. points, including the constraint that features are coplanar b) reconstruct planes directly.

The key idea of the first approach is to *improve* feature-based, e.g. points, reconstructions with additional coplanarity constraints. These constraints can be incorporated in the reconstruction process directly, e.g. [11, 6, 1], or included in a final bundle adjustment stage of an initial reconstruction, e.g. [11]. However, such methods do not attempt to reconstruct planes independently of other features.

The second approach, which will be pursued in this paper, reconstruct the planes directly from the knowledge of *homographies*. A homography is a projective transformation matrix and encodes the information about a plane visible in two views (see [8, 3, 2]). A homography can be determined from image features such as points, e.g. [3], or directly from image brightness information, e.g. [4]. Methods to reconstruct multiple planes in *two uncalibrated views* have been presented in [5, 13]. Planes and cameras are reconstructed directly, however, separately on the basis of a generalized eigenvalue analyses of the respective homographies. A method for the recovery of multiple planes in *multiple uncalibrated views* has been briefly introduced in [12]. It requires all planes to be visible in all views and selects one of the views as a reference view. Cameras and "homography parallaxes", which describe the planes, are obtained from a rank-1 factorization of a matrix containing all homographies. A less compact rank-4 factorization method has been presented in [9, 14] for two view.

This paper presents two methods for the projective recovery of planes and cameras from multiple uncalibrated views. Both methods specify one of the planes, which is visible in all views, as the plane at infinity. This basic idea has been exploited in [7] to reconstruct 3D points and cameras simultaneously. The first method assumes all planes to be visible in all views. The simultaneous reconstruction of planes and cameras is achieved by a rank-1 factorization of a matrix containing all homographies between a reference view and all other views. In contrast to [12] this method is simpler and more direct. Furthermore, no multi-image constraints, e.g. epipolar constraints, are needed. Since planes become inevitably occluded when the camera's viewpoint changes, we present an alternative method which handles missing planes. The only requirement is that a reference plane is visible in all views. Firstly, all the cameras are recovered simultaneously, secondly all planes are reconstructed. Optionally, these two steps can be iterated.

2 Planes and Cameras in Projective Space

The general mapping of a 3D point X onto an image point x by camera P is defined in homogeneous coordinates as

$$x \sim PX, \tag{1}$$

where P denotes the 3×4 camera matrix and "$\sim$" means equal up to scale. A 3D point X lies on the 3D plane π_k if

$$\pi_k^T X = 0. \tag{2}$$

Suppose a plane is visible in two images. It induces a projective transformation between the two views, which is called a *homography* (see [8, 3, 2]). In particular, the image points x_i and x_j in view i and j of a 3D point X which lies on the plane π_k are related by the homography H_{ij}^k as

$$x_j \sim H_{ij}^k \, x_i, \tag{3}$$

where H_{ij}^k is a 3×3 matrix. The homography is well defined, i.e. has rank 3, if both camera centers do not lie on the plane. Homographies can be determined from point correspondences (at least 4) or directly from image intensities.

Let us consider a general multi view situation which consists of m unknown camera matrices $P_1, \ldots P_m$ and n unknown planes $\pi_1, \ldots, \pi_n$. The goal of this paper is to recover the m cameras and n planes on the basis of known homographies H_{ij}^k. Since in practice not all planes might be visible in all views we consider two cases: a) all planes are visible in all views (sec. 3) and b) at least one plane is visible in all views (sec. 4).

In a projective framework it is common to choose the camera matrices as (see [3, 2]):

$$P_1 \sim (I \,|0) \,, \; P_j \sim H_j^\infty \, (I \,| - \bar{Q}_j) \; \text{ for } j = 2 \ldots m \,, \tag{4}$$

where I is the identity matrix and $\bar{Q}_j$ represents the camera center of view j in non-homogeneous coordinates. Let us consider the homography H_j^∞ in more detail. A point at infinity $X = (x, y, z, 0)^T$ is mapped into view 1 and j as $x_1 = (x, y, z)^T$ and $x_j = H_j^\infty (x, y, z)^T$. This means that H_j^∞ represents the mapping from view 1 to view j via the plane at infinity, i.e. $x_j \sim H_j^\infty x_1$. Let us assume that one plane, without loss of generality π_1, is visible in all views. By applying a projective transformation, we may choose this plane as the plane at infinity in our specific projective space, i.e. $\pi_1 = \pi_\infty = (0, 0, 0, 1)^T$. This means that the only remaining unknowns of the camera matrices (see eqn. (4)) are the camera centers. Furthermore, 14 of the 15 degrees of freedom (dof) of our specific projective space are fixed (camera P_1 fixes 11 dof and π_1 fixes 3 dof).

3 All planes visible in all views

Let us choose the first view P_1 as a reference view. With the assumption that all n planes are visible in all views we can determine the homography H_{1j}^k (for simplicity denoted as H_j^k) between the reference view 1 and view j for a plane π_k. It is known (see [3, 2]) that this homography can be explicitly expressed as

$$H_j^k = \lambda \, H_j^\infty \, (I + \bar{Q}_j \, v_k^T) \,, \tag{5}$$

with $\pi_k^T = (v_k^T, 1)$ and λ as an arbitrary scalar. The fourth coordinate of π_k can be chosen as 1 since the camera center $Q_1 = (0, 0, 0, 1)^T$ must not lie on π_k, i.e. $\pi_k^T Q_1 \neq 0$.

We will now show that λ can be determined directly from the known homographies H_j^k and H_j^∞. Eqn. (5) can be rewritten as

$$H_j^{\infty -1} \, H_j^k - \lambda \, I = \lambda \, \bar{Q}_j \, v_k^T. \tag{6}$$

This matrix has rank 1 since it is the product of two rank 1 matrices. Therefore this matrix has the double eigenvalue 0, which means that the matrix $H_j^{\infty -1} \, H_j^k$ has the double eigenvalue λ. This result was previously presented in e.g. [5, 12] and relates to the fact that $H_j^{\infty -1} \, H_j^k$ is a planar homology [3]. However, it has never been exploited for the direct recovery of planes in multiple views.

With known λ eqn. (5) can be rewritten as

$$\hat{H}_j^k = \lambda^{-1} \, H_j^{\infty -1} H_j^k \, - I = \bar{Q}_j \, v_k^T \,. \tag{7}$$

We are now able to construct a block matrix of size $3m \times 3n$ which is the product of two vectors: one vector containing all the camera centers and another vector containing all the planes:

$$W = \begin{pmatrix} \hat{H}_2^2 & \ldots & \hat{H}_2^n \\ \vdots & & \vdots \\ \hat{H}_m^2 & \ldots & \hat{H}_m^n \end{pmatrix} = \begin{pmatrix} \bar{Q}_2 \\ \vdots \\ \bar{Q}_m \end{pmatrix} \left(v_2^T, \ldots, v_m^T \right). \tag{8}$$

The matrix W has rank at most 1 which corresponds to the 1 dof of our specific projective space. The final reconstruction for cameras and planes can be obtain from the Singular Value Decomposition (SVD) of $W = UDV^T$:

$$\left(\bar{Q}_2, \ldots, \bar{Q}_m \right)^T = u_1^T \text{ and } \left(v_2^T, \ldots, v_m^T \right) = \sigma_1 v_1^T \,, \tag{9}$$

where σ_1 is the first singular value of D and u_1, v_1 are the first columns of U and V respectively.

4 A reference plane visible in all views

In the previous section, only those homographies were utilized which include the reference view P_1, i.e. H_{1j}^k. In this section we will present a method which exploits all available homographies, i.e. H_{ij}^k. Furthermore, the only restriction will be that a reference plane, i.e. the plane at infinity, is visible in all views. The homography H_{ij}^k can be explicitly expressed as $H_{ij}^k =$

$$\lambda \, H_j^\infty \Big(I + (v_k^T \bar{Q}_i + v_{k4})^{-1} (\bar{Q}_j - \bar{Q}_i) \, v_k^T \Big) H_i^{\infty -1}, \tag{10}$$

where $\pi_k^T = (v_k^T, v_{k4})$. The derivation is similar to the derivation of eqn. (5) and was presented in [11]. As in the previous case, λ can be directly determined from the homographies H_{ij}^k, H_i^∞ and H_j^∞. We rewrite eqn. (10) as

$$H_j^{\infty -1} H_{ij}^k H_i^\infty - \lambda I \; = \; \lambda \, (v_k^T \bar{Q}_i + v_{k4})^{-1} (\bar{Q}_j - \bar{Q}_i) \, v_k^T.$$

Since this matrix is the product of two rank 1 matrices, the matrix $H_j^{\infty -1} H_{ij}^k H_i^\infty$ has the double eigenvalue λ. With known λ eqn. (10) can be rewritten as

$$\hat{H}_{ij}^k = (v_k^T \bar{Q}_i + v_{k4})^{-1} (\bar{Q}_j - \bar{Q}_i) \, v_k^T.$$

where $\hat{H}_{ij}^k = \lambda^{-1} H_j^{\infty -1} H_{ij}^k H_i^\infty - I$. From the SVD of the matrix $\hat{H}_{ij}^k = UDV^T$ we obtain

$$\begin{aligned} \bar{Q}_j - \bar{Q}_i &= \lambda_{ij}^k \, u_1 && (11) \\ v_k &= \frac{\sigma_1}{\lambda_{ij}^k} \, (v_k^T \bar{Q}_j + v_{k4})^{-1} \, v_1, && (12) \end{aligned}$$

where σ_1 is the first singular value of D and u_1, v_1 are the first columns of U and V respectively. The scalar λ_{ij}^k is undetermined in this case.

Ideally, all camera centers, i.e. $\bar{Q}_i, \bar{Q}_j$, all planes $\pi_k = (v_k, v_{k4})^T$ and all scalars λ_{ij}^k should be determined simultaneously. However, due to the non-linear nature of this problem, we suggest a two step method: a) derive all camera centers simultaneously b) determine the planes. From eqn. (11) we may derive the three linear relations:

$$\begin{aligned} u_y(X_j - X_i) - u_x(Y_j - Y_i) &= 0 \\ u_z(X_j - X_i) - u_x(Z_j - Z_i) &= 0 \qquad (13) \\ u_z(Y_j - Y_i) - u_y(Z_j - Z_i) &= 0\,, \end{aligned}$$

where $\bar{Q}_i = (X_i, Y_i, Z_i)^T$ and $u_1 = (u_x, u_y, u_z)^T$. Therefore, each homography $\hat{H}_{ij}^k$ provides three linear equations of the form (13) which can be put into a set of linear equations:

$$L \;\; (X_1, Y_1, Z_1 \ldots X_m, Y_m, Z_m)^T \; = \; 0\,. \qquad (14)$$

The null-space of L has dimension 1 since our specific projective space has 1 dof. Therefore, the last singular vector of the SVD of L provides a solution for all camera centers. The scalars λ_{ij}^k can now be derived from the camera centers (see eqn. (11)). With known λ_{ij}^k and $\bar{Q}_i$ each plane $\pi_k = (v_k, v_{k4})^T$ can be directly determined from eqn. (12).

Optionally, these two steps can be iterated. This means that the scalars λ_{ij}^k are recalculated from the plane π_k and used to redetermine all the camera centers simultaneously.

5 Experiments

5.1 Synthetic Data

A 8 frame synthetic sequence was generated based on the scene shown in fig. 1 (a) which consists of 9 planes forming a house. The homographies between views were computed from point matches, where each plane has on average 20 points. The size of the ground plane (reference plane) is 100×100 units and the height of the house is 80 units.

In a first experiment the assumption was made that all planes are visible in all views, i.e. all planes are transparent. Our two algorithms which assume either all planes to be visible in all views (*All planes*; sec. 3) or only a reference plane (*Ref. plane*; sec. 4) were applied to the synthetic sequence. For comparison, the projective factorization algorithm (*Factorization*) of Sturm-Triggs [10] was applied to the point matches. This method assumes that all points are visible in all views and is known to give nearly optimal results. The performance of the algorithms is evaluated in terms of a 3D error in units. This error represents the average point distance in a Euclidean frame, which means that the projective reconstruction was aligned with the ground truth Euclidean model. For our plane based algorithms, the 3D points were obtained by triangulation (see [3]).

Fig 1 (b) shows the average performance of 20 runs with respect to different levels of Gaussian noise: $\sigma = 0, 0.2, \ldots, 3.0$ (standard deviation) which was added to the image data, i.e. reprojected 3D points. The performance of the *Ref. plane* algorithm and the *Factorization* algorithm are virtually identical. The performance of the *All planes* algorithm is slightly worse for smaller noise, e.g. $\sigma < 1$, and significantly worse otherwise. A plausible explanation is that the algorithm utilizes only those homographies which include a certain reference view. In contrast to this, the *Ref. plane* algorithm exploits the knowledge of all homographies. The number of iterations for the *Ref. plane* algorithm was on average 7.

In a second experiment the influence of missing homographies was investigated. In contrast to the previous experiment, the house was assumed to be non-transparent. This means that each plane, except the ground plane, is visible in only 3 successive views. Fig. 1 (c) depicts the performance of the *Ref. plane* algorithm in this case (*true vis.*). Additionally, the performance of the previous experiment is shown (*compl. vis.*). We see that the *Ref. plane* algorithm performs very well even with missing homographies (note the different scale between Fig. 1 (b) and (c)).

5.2 Real Data

The reference plane algorithm was applied to the toy-house sequence (available at [15]) consisting of 8 images. Fig. 2 (a) shows the first frame of the sequence. The ground plane, which dealt as the reference plane, the roof and the front side of the house are visible in all views. The other 3 planes are only visible in the first four views. The homographies were determined on the basis of point matches. The metric rectified VRML-model is shown in fig. 2 (b,c). The intersection lines of the planes were identified manually.

6 Summary and Conclusions

This paper presented two novel methods for the projective recovery of planes and cameras from multiple views. The first method determines planes and cameras by factorizing a matrix containing all planar homographies between a reference view and all other views. This requires that all planes are visible in all views. The second method assumes only a reference plane to be visible in all views. The reconstruction is determined in two steps: a) all camera centers are recovered simultaneously b) all planes are reconstructed. Optionally, these two steps can be iterated.

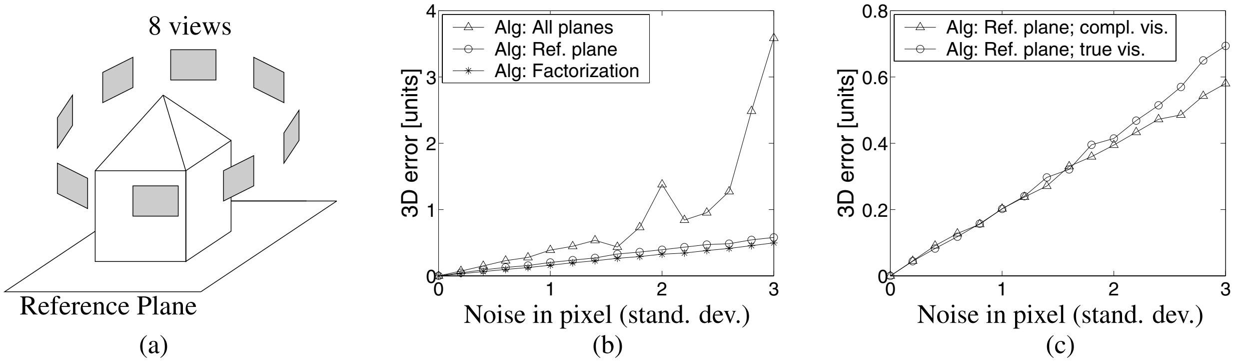

Figure 1. 8 **views of a synthetic house (a). The performance of the algorithms for the case of all planes visible in all views (b) and true visibility (c).**

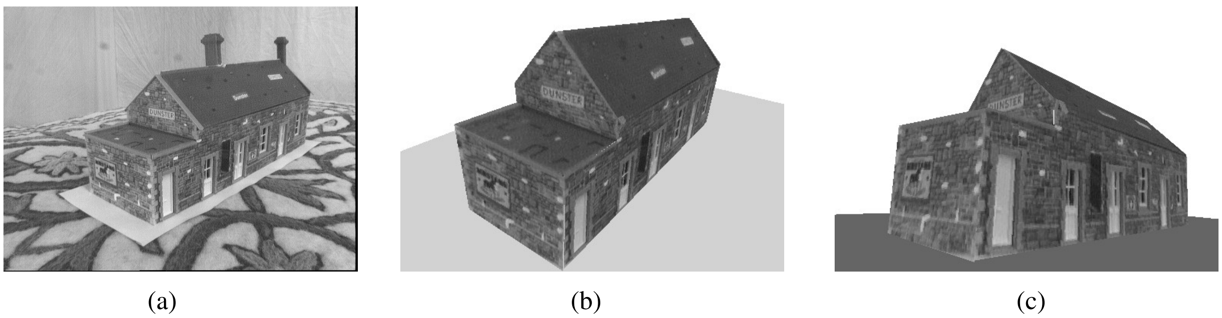

Figure 2. Real view (a) of the toyhouse sequence and synthetic views (b,c) of the VRML model.

The key idea of both methods is to specify one of the planes, which is visible in all views, as the plane at infinity. This implies that the unknown scale of each homography can be determined directly from an eigenvalue analyses.

Both methods were tested on real and synthetic data and performed well in comparison to the projective factorization algorithm for points of Sturm-Triggs [10], which requires all points to be visible in all views. Our method which handles missing planes performed slightly better than our method which assumes all planes to be visible. An explanation is that the latter method utilizes only those homographies which include a certain reference view.

References

[1] Bartoli, A., Sturm, P. and Horaud, R. 2001. Projective Structure and Motion from Two Views of a Piecewise Planar Scene. In ICCV, Vancouver, Canada, p. 593-598.

[2] Faugeras, O. and Luong, Q.-T. 2001. The Geometry of Multiple Images. The MIT Press.

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

[4] Irani, M. and Anandan, P. 1999. About Direct Methods. In IEEE Workshop on Vis. Algo., Kerkyra, Greece, p. 267-278.

[5] Johansson, B. 1999. View Synthesis and 3D Reconstruction of Piecewise Planar Scenes using Intersection Lines between the Planes. In ICCV, Kerkyra, Greece, p. 54-59.

[6] Morris, D. D. and Kanade, T. 1998. A unified factorization algorithm for points, line segments and planes with uncertainty models. In ICCV, Bombay, India, p. 696-702.

[7] Rother, C. and Carlsson, S. 2001. Linear Multi View Reconstruction and Camera Recovery. In ICCV, Vancouver, Canada, p. 42-51.

[8] Semple, J.G. and Kneebone, G.T. 1952. Algebraic Projective Geometry. Oxford University Press.

[9] Shashua, A. and Avidan, S. 1996. The Rank 4 Constraint in Multiple ($\geq$ 3) View Geometry. In ECCV, Cambridge, U.K., p. 196-206.

[10] Sturm, P. and Triggs, B. 1996. A factorization based algorithm for multi-image projective structure and motion. In ECCV, Cambridge, U.K., p. 709-719.

[11] Szeleski, R. and Torr, P. 1998. Geometrically Constrained Structure from Motion: Points and Planes. In Workshop on 3D Structure from Multiple Images of Large-Scale Environments, LNCS 1506, Springer-Verlag, p. 208-223.

[12] Triggs, B. 2000. Plane + Parallax, Tensors and Factorization. In ECCV, Dublin, Ireland, p. 522-538.

[13] Xu, G, Terai, J. and Shum, H. 2000. A linear algorithm for Camera Self-Calibration, Motion and Structure Recovery for Multi-Planar Scenes from Two Perspective Images. In CVPR, Hilton Head Island, South Carolina, p. 474-479.

[14] Zelnik-Manor, L. and Irani, M. 1999. Multi-View Subspace Constraints on Homographies. In ICCV, Kerkyra, Greece, p. 710-715.

[15] "Visual Geometry Group Oxford ", http://www.robots.ox.ac.uk/ vgg/data/.