Linear Multi View Reconstruction and Camera Recovery

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

Abstract

This paper presents a linear algorithm for the simultaneous computation of 3D points and camera positions from multiple perspective views, based on having four points on a reference plane visible in all views. The reconstruction and camera recovery is achieved, in a single step, by finding the null-space of a matrix using singular value decomposition. Unlike factorization algorithms, the presented algorithm does not require all points to be visible in all views. By simultaneously reconstructing points and views the numerically stabilizing effect of having wide spread cameras with large mutual baselines is exploited. Experimental results are presented for both finite and infinite reference planes. An especially interesting application of this method is the reconstruction of architectural scenes with the reference plane taken as the plane at infinity which is visible via three orthogonal vanishing points. This is demonstrated by reconstructing the outside and inside (courtyard) of a building on the basis of 35 *views in one single SVD.*

1 Introduction

The efficient computation of 3D structure and camera information from multiple views has been a subject of considerable interest in recent years [7]. The problem can be formulated, most generally, as a bilinear inverse problem for finding camera and 3D information from image data. Contrary to the case of parallel projection [25], no algorithm for direct factorization of camera parameters and 3D structure has been produced for perspective projection cameras. The perspective factorization algorithm suggested in [23] relies on the pre-computation of scale factors "projective depths" in order to cast the problem into the same form as in [25]. Approaches have been invented for efficient combination of groups of views [5, 13], or iterative methods exploiting all views [9]. The approach in [18] using "shape constraints" dual to epipolar constraints [1, 2, 27] can in principle be exploited for computing projective structure using arbitrary number of views [6, 22]. It is, however, limited to a restricted number of points at a time.

Ideally an algorithm for computing multiple view reconstruction and camera information should exploit all points and views simultaneously as in [25]. Inevitably points become occluded as the camera view changes. Therefore, a certain point is only visible in a certain set of views. An efficient algorithm should be able to deal with this problem. Note that this problem is not handled by any suggested general reconstruction algorithm so far, although it has been given some attention [12, 17, 19].

In this paper, we will show that by adding the simple assumption of having four points on a reference plane visible in all views [14], the problem of reconstruction and camera recovery can be formulated and solved very efficiently as a linear null-space problem. It is based on the fact that having a reference plane in *arbitrary* position in 3D, the problem is transformed into the equivalent problem of reconstructing a set of translating calibrated cameras. Variations of this have been observed and discussed [4, 7, 8, 16, 26], but it seems that its full potential for reconstruction and camera recovery has not yet been exploited. The advantage that the constraints, given by the 2, 3 or 4 view tensors, become linear if a plane is visible in all views has been exploited in [7, 8]. However, structure and camera motion cannot be recovered simultaneously.

The crucial observation exploited in this paper is that, with the reference plane, the projection relations between image points, scene points, and cameras become *linear* in 3D points and camera positions as opposed to being *bilinear* in the general case. We will show how this relation can be derived for general reference planes and also how it relates to general perspective projection. In particular, we will demonstrate the relation between general shape-viewpoint duality [1, 2, 27], and the dual structures that arise with a reference plane, [3, 10, 11]. This linear relationship enables us to *simultaneously* reconstruct points and camera positions even when not all points are visible in all views.

*This work was supported by the Swedish Foundation for Strategic Research in the VISIT program.

A potential problem for numerical calculations is the fact that the reference plane will be at infinity in the representation that linearizes the problem. We will demonstrate, however, that this problem can be dealt with both from a theoretical and practical point of view.

An especially interesting case is when the reference plane actually is at infinity. We will show that the knowledge of three orthogonal vanishing points, which span the plane at infinity, together with some simple natural assumptions about the camera parameters result in the same linear reconstruction problem as with four points on an arbitrary reference plane. As a practical demonstration we will simultaneously reconstruct architectural structures together with the camera positions based on multiple points which are viewed with only partial overlap.

2 Duality, symmetry and linearity of projection relations

General perspective projection to an image point with homogeneous coordinates p can be described as:

$$p \sim H\,(\,I \mid -\bar{Q})\,P \sim H\,(\bar{P} - \bar{Q}). \qquad (1)$$

Where P and $\bar{P}$ are the homogeneous and non-homogeneous cartesian coordinates of the 3D-points respectively and $\bar{Q}$ are cartesian coordinates of the camera centres. In a general projective representation the homography H will be factored out and we are left with relations between 3-D points and camera centres. Already from eqn. (1) we see that these quantities are symmetrically related. The symmetry relations in a complete projective description will be somewhat different depending on whether we exploit the presence of four points on a reference plane in 3D.

With five points $P_1 \dots P_5$ as basis, any point P and camera centre Q in 3D can be expressed using projective coordinates:

$$\begin{aligned} P &\sim XP_1^* + YP_2^* + ZP_3^* + WP_4^* \\ Q &\sim AP_1^* + BP_2^* + CP_3^* + DP_4^*. \end{aligned} \qquad (2)$$

Similarly, four image points can be used as a basis for expressing image coordinates

$$p \sim xp_1^* + yp_2^* + wp_3^*. \qquad (3)$$

The normalizations P^* and p^* are chosen so that points P_5 and p_4 have projective coordinates $(1,1,1,1)^T$ and $(1,1,1)^T$ respectively.

The mapping

$$M : (X,\, Y,\, Z,\, W)^T \longrightarrow (x,\, y,\, w)^T \qquad (4)$$

can be computed for the general case and for the case of having four points on a reference plane.

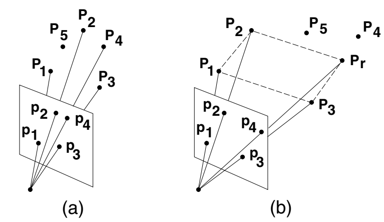


Figure 1. Projective 3D basis points $P_1 \dots P_5$ and image basis points $p_1 \dots p_4$ for general configurations (a) and reference plane configurations (b).

2.1 General point configurations

This case was treated in [1]. The image basis points p_1, p_2, p_3, p_4 are projections of the 3D basis points P_1, P_2, P_3, P_4 (see Fig. 1 (a)), which constrains the mapping M:

$$M : \begin{matrix} P_1 & P_2 & P_3 & P_4 & Q \\ - & - & - & - & - \\ 1 & 0 & 0 & 0 & A \\ 0 & 1 & 0 & 0 & B \\ 0 & 0 & 1 & 0 & C \\ 0 & 0 & 0 & 1 & D \end{matrix} \quad \longrightarrow \quad \begin{matrix} p_1 & p_2 & p_3 & p_4 & 0 \\ - & - & - & - & - \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 \end{matrix} \qquad (5)$$

This results in the following projection relations:

$$\begin{pmatrix} x \\ y \\ w \end{pmatrix} \sim \begin{pmatrix} \frac{X}{A} - \frac{W}{D} \\ \frac{Y}{B} - \frac{W}{D} \\ \frac{Z}{C} - \frac{W}{D} \end{pmatrix} \qquad (6)$$

which can be written as two constraint equations:

$$\begin{aligned} w\,\frac{X}{A} - x\,\frac{Z}{C} + (x - w)\,\frac{W}{D} &= 0 \\ w\,\frac{Y}{B} - y\,\frac{Z}{C} + (y - w)\,\frac{W}{D} &= 0. \end{aligned} \qquad (7)$$

These relations make explicit the duality of space points and camera centres in the sense that the *homogeneous* projective coordinates of a 3D point $(X, Y, Z, W)^T$ and inverse coordinates of a camera centre $(A^{-1}, B^{-1}, C^{-1}, D^{-1})^T$ are *bilinearly* related in a symmetric way.

2.2 Four points on a reference plane

If four points P_1, P_2, P_3, P_r on a reference plane are visible in all views, their images can be used as a four point basis for image coordinates (see Fig. 1 (b)). This has similar consequences leading to a similar but non-equivalent duality or symmetry between space points and camera centres. Since a projective basis in 3D cannot be formed from four coplanar points, only points P_1, P_2, P_3 can be used in the 3D basis. Points P_4 and P_5 of the 3D basis will be assumed to lie outside the reference plane. The situation is summarized as:

Projective 3D basis points: P_1, P_2, P_3, P_4, P_5

3D points on reference plane: P_1, P_2, P_3, P_r

Projective image basis points: p_1, p_2, p_3, p_4.

If $(X_r, Y_r, Z_r, 0)^T$ are the projective coordinates of the reference point P_r in the 3D basis, the constraints on the mapping M becomes:

$$M: \begin{array}{ccccc} P_1 & P_2 & P_3 & P_r & Q \\ - & - & - & - & - \\ 1 & 0 & 0 & X_r & A \\ 0 & 1 & 0 & Y_r & B \\ 0 & 0 & 1 & Z_r & C \\ 0 & 0 & 0 & 0 & D \end{array} \longrightarrow \begin{array}{ccccc} p_1 & p_2 & p_3 & p_4 & 0 \\ - & - & - & - & - \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 \end{array} \tag{8}$$

resulting in the projection relations:

$$\begin{pmatrix} x \\ y \\ w \end{pmatrix} \sim \begin{pmatrix} X_r^{-1}\left(\frac{X}{W} - \frac{A}{D}\right) \\ Y_r^{-1}\left(\frac{Y}{W} - \frac{B}{D}\right) \\ Z_r^{-1}\left(\frac{Z}{W} - \frac{C}{D}\right) \end{pmatrix}. \tag{9}$$

Comparing this to the general projection relations in eqn. (6) we see that the symmetry of points and cameras still remains. The symmetry now relates to the substitutions $(X, Y, Z, W) \leftrightarrow (A, B, C, D)$. The factors $X_r^{-1}, Y_r^{-1}, Z_r^{-1}$ are common to all points and views so they do not affect the symmetry. More importantly, the relations are *linear* in the *non-homogeneous* projective coordinates for points (X/W etc.) and cameras (A/D etc.). This means that the projection relations only apply to non-homogeneous points and cameras which are outside the plane at infinity (specified by $W = 0$). The fact that the reference plane is the plane at infinity, in this projective basis, was noted in [26] as a fundamental requirement for obtaining this simple structure, which is mathematically equivalent to purely translating calibrated cameras.

The projection relations can be rewritten as linear constraint equations:

$$\begin{aligned} x\left(\frac{Z'}{W} - \frac{C'}{D}\right) - w\left(\frac{X'}{W} - \frac{A'}{D}\right) &= 0 \\ y\left(\frac{Z'}{W} - \frac{C'}{D}\right) - w\left(\frac{Y'}{W} - \frac{B'}{D}\right) &= 0 \end{aligned} \tag{10}$$

using the notation: $X' = X_r^{-1}\, X$ etc. for points and $A' = X_r^{-1}\, A$ etc. for cameras.

The simple addition of a fourth coplanar point therefore implies that the general *bilinear* problem of reconstruction and camera recovery from multiple points and views is transformed into a *linear* problem, i.e. finding the null-space of a matrix with elements computed from the image coordinates in all available views. Arbitrary numbers of points and views can be used to build this matrix as long as all reference points are visible in all views. For n points in m views the linear system takes the form

$$\begin{pmatrix} S_{11} & 0 & 0 & \dots & 0 & 0 & -S_{11} & 0 & \dots & 0 \\ S_{12} & 0 & 0 & \dots & 0 & 0 & 0 & -S_{12} & \dots & 0 \\ \vdots & & & & & \vdots & & & & \\ S_{1m} & 0 & 0 & \dots & 0 & 0 & 0 & 0 & \dots & -S_{1m} \\ 0 & S_{21} & 0 & \dots & 0 & 0 & -S_{21} & 0 & \dots & 0 \\ 0 & S_{22} & 0 & \dots & 0 & 0 & 0 & -S_{22} & \dots & 0 \\ & \vdots & & & & \vdots & & & & \\ 0 & S_{2m} & 0 & \dots & 0 & 0 & 0 & 0 & \dots & -S_{2m} \\ & & & & \vdots & & & & & \\ 0 & 0 & 0 & \dots & 0 & S_{n1} & -S_{n1} & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 & S_{n2} & 0 & -S_{n2} & \dots & 0 \\ & & & & & \vdots & & & & \\ 0 & 0 & 0 & \dots & 0 & S_{nm} & 0 & 0 & \dots & -S_{nm} \end{pmatrix} \begin{pmatrix} \bar{X}'_1 \\ \bar{Y}'_1 \\ \bar{Z}'_1 \\ \vdots \\ \bar{X}'_n \\ \bar{Y}'_n \\ \bar{Z}'_n \\ \bar{A}'_1 \\ \bar{B}'_1 \\ \bar{C}'_1 \\ \vdots \\ \bar{A}'_m \\ \bar{B}'_m \\ \bar{C}'_m \end{pmatrix} = 0 \tag{11}$$

for non-homogeneous projective points coordinates $\bar{X}' = X'/W$ etc. and camera centres $\bar{A}' = A'/D$ etc. where

$$S_{ij} = \begin{pmatrix} -w_{i,j} & 0 & x_{i,j} \\ 0 & -w_{i,j} & y_{i,j} \\ -y_{i,j} & x_{i,j} & 0 \end{pmatrix} \tag{12}$$

are 3×3 matrices built up from image coordinates of point i visible in view j. In the case of missing data, e.g. point i not visible in view j, the three equations which correspond to S_{ij} are omitted in the system. Note, S_{ij} contains three linearly dependent equations since the two equations in eqn. (10) are insufficient for special cases ($w = 0$ and $\frac{Z'}{W} - \frac{C'}{D} = 0$).

Finding the 3D coordinates of points and camera centres directly from the null-space of the matrix eliminates the problem of computing fundamental matrices [23] in order to get depth scale factors or complete camera projection matrices [5] before doing reconstruction.

The next section describes how to obtain the position in the normalized projective basis of points and cameras which lie both on and not on the reference plane.

3 Finite versus Infinite Reference Plane

3.1 Finite Reference Plane

Let us complete the reconstruction method for 4 coplanar reference points. Adding an arbitrary common scaling and translation to all points and cameras leaves the equations in eqn. (10) unaltered. This is related to the fact that the points P_4 and P_5 of the normalized projective 3D basis were not used for the derivation of eqn. (10). Therefore, we fix one arbitrary point which does not lie on the reference plane, e.g. the first camera centre Q_1, as the basis point $P_4 = (0, 0, 0, 1)^T = (\bar{A}'_1, \bar{B}'_1, \bar{C}'_1, 1)^T$. The camera Q_1 can now be removed from the vector of unknowns in eqn. (11), which implies that the null-space of the matrix in eqn. (11) reduces to dimension one. In order to complete the projective basis, we choose e.g. $Q_2 = (\bar{A}'_2, \bar{B}'_2, \bar{C}'_2, 1)^T$ as the basis point P_5. This defines the reference point as: $P_r = (X_r, Y_r, Z_r, 0)^T = (1/\bar{A}'_2, 1/\bar{B}'_2, 1/\bar{C}'_2, 0)^T$. As the projective basis must not contain four coplanar points, P_4 and P_5 have to be selected carefully.

Finally, all points and cameras can be expressed in the normalized projective 3D basis as:

$$(X_r\bar{X}', Y_r\bar{Y}', Z_r\bar{Z}', 1)^T \text{ for points and}$$
$$(X_r\bar{A}', Y_r\bar{B}', Z_r\bar{C}', 1)^T \text{ for cameras.}$$

Since the reference plane is the plane at infinity in the projective basis representation, points on the reference plane cannot be reconstructed by the linear system in eqn. (11). Therefore, those points have to be treated separately. The projection relations (eqn. (9)) can be written

$$\begin{pmatrix} x \\ y \\ w \end{pmatrix} \sim \begin{pmatrix} X_r^{-1} & 0 & 0 & -X_r^{-1}\frac{A}{D} \\ 0 & Y_r^{-1} & 0 & -Y_r^{-1}\frac{B}{D} \\ 0 & 0 & Z_r^{-1} & -Z_r^{-1}\frac{C}{D} \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ W \end{pmatrix} \tag{13}$$

to obtain for a point $(X, Y, Z, 0)^T$ on the plane at infinity the projective relations

$$\begin{pmatrix} x \\ y \\ w \end{pmatrix} \sim \begin{pmatrix} X_r^{-1}X \\ Y_r^{-1}Y \\ Z_r^{-1}Z \end{pmatrix}. \tag{14}$$

This means that this point has the coordinates $(X_r x, Y_r y, Z_r z, 0)^T$ in the projective basis.

In practice, points on the reference plane can be dealt with separately. Consider the case of including such a point into the system in eqn. (11). It can be shown that this results in a two-dimensional null-space of the matrix, i.e. 2 singular values which are zero. Therefore, numerical instability of the system can be expected for points on or "close" to the reference plane. Firstly, due to errors in the image coordinates the singular vector which does *not* represent the correct solution for all points and cameras could have the smallest singular value. Secondly, points which are closer to the reference plane have larger values of coordinates in the projective basis. Due to normalization, this potentially increases the inaccuracy in all those points which are further away of the reference plane. Therefore, we have to separate points on and off the reference plane. The four reference points P_1, P_2, P_3 and P_r induce a planar homography between each pair of views. For a certain world point, which is visible in two views, the residual parallax vector, which contains the relative distance of this world point to the reference plane [10, 4], can be determined. On the basis of the magnitude of the parallax vector we can come to the decision whether this point lies on or off the reference plane.

Having 4 coplanar points visible in all views is equivalent to the assumption of a reference plane visible in all views (e.g. [4, 10]). A reference plane visible in all views induces planar homographies between each pair of views. Introduce four "virtual" basis points which lie on the reference plane. By fixing them in one view they can be determined via the homographies in all views. This shows the equivalence of both assumptions. However, inaccurate image coordinates might reduce the quality of the inter-view homographies which could introduce a substantial numerical instability in the reconstruction process.

3.2 Reference Plane at Infinity

If the points P_1, P_2, P_3, P_r on the reference plane are moved to infinity it can be easily shown that the projective 3D coordinates in the basis $P_1 \ldots P_5$ become affine coordinates in an affine system defined by the direction vectors $P_1 - P_4, P_2 - P_4$ and $P_3 - P_4$. If these directions are orthogonal, the affine 3D coordinates become Euclidean by proper choice of the normalizing point P_5. This is true for a *general* un-calibrated perspective camera. This can typically be achieved if the points P_1, P_2, P_3 are chosen as the orthogonal vanishing points in e.g. a city block architectural scene. However, determining the 4th reference point P_r on the plane at infinity which is visible in all views will substantially restrict the usefulness of this approach. By considering a *special case* of perspective cameras, however, we can make use of the reference plane at infinity given solely by the three orthogonal vanishing points.

For perspective cameras:

$$p \sim KR\,(I \mid -\bar{Q})\,P \tag{15}$$

where the calibration matrix K can be written as

$$K = \begin{pmatrix} \sigma & 0 & \bar{x}_0 \\ 0 & \sigma & \bar{y}_0 \\ 0 & 0 & 1 \end{pmatrix} \tag{16}$$

(zero skew and the same scale factor horizontally and vertically) it is possible to recover internal parameters K and camera rotations R from knowledge of three orthogonal vanishing points [15, 24, 20]. Normalizing image coordinates with this knowledge we can write the projection relation as:

$$p' \sim \bar{P} - \bar{Q} \tag{17}$$

which of course is the translating calibrated camera, derived as before using the pure projective representation in the previous section. The use of metric information, i.e. orthogonal directions, implies that metric reconstruction (3D structure up to similarity transformations) is possible.

For certain configurations of the three orthogonal vanishing points the special camera in eqn. (16) cannot be

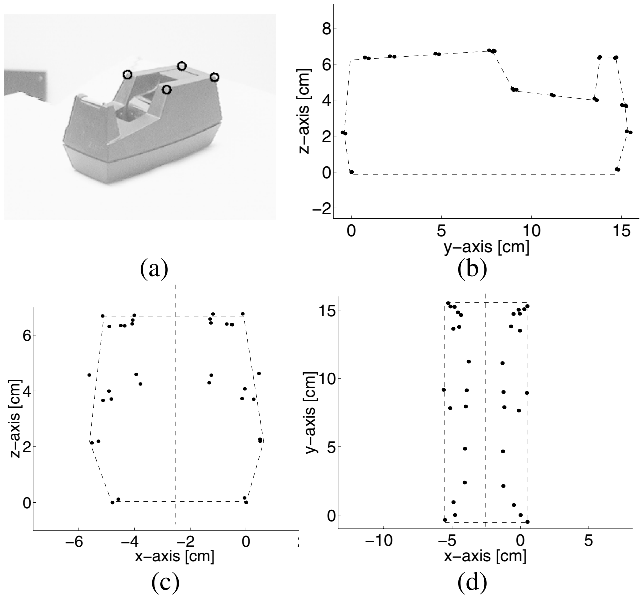


Figure 2. Original view (a), side (b), top (c) and front (d) view of the reconstruction. The dots represent the reconstructed model points and the dashed lines display the contour and the symmetry axis of the model.

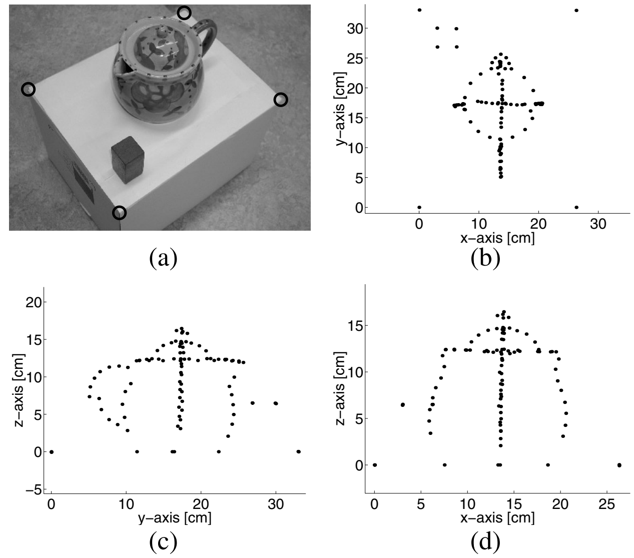


Figure 3. Original view (a), top (b), side (c) and front (d) view of the reconstruction. The dots represent the reconstructed model points.

fully calibrated (see [15, 20]). In practice, one can deal with this problem by assuming fixed internal camera parameters. Furthermore, the rotation matrix R is ambiguous (24 possible variations of R) as long as the correspondence of the vanishing points between different views is unknown. In the current version of the algorithm this correspondence problem is solved manually.

4 Experiments

4.1 Finite Reference Plane

In the first experiment a tape holder was reconstructed. Four images of the tape holder were taken from viewpoints with considerable wide mutual baselines (see Fig. 2 (a)). Since the tape holder itself contains a plane which is visible in all images this plane was used as the finite reference plane. The four points, which are marked with circles, define the reference plane. On the basis of this, the reconstruction of 30 model points was achieved in one SVD. The 6 model points which lie on or close the reference plane were detected automatically. In order to visualize the result we assumed knowledge of five Euclidean coordinates to rectify the projective structure (see Fig. 2 (b-d)). We see that the reconstruction matches with the approximate size of the object which is 6.0cm (x-direction), 15.8cm (y-direction) and 6.8cm (z-direction). Furthermore, the symmetry of the object is maintained in the reconstruction. Since the ratio between the second last singular value (0.766) and the last singular value (0.031) is substantial (24.7) this configuration can be considered as non-critical.

In the second experiment we reconstructed a teapot, which was positioned on a box (see Fig. 3 (a)). The four corner points of the box, which are marked with circles, specify the reference plane. For better visualization, only those model points which lie on the contour in the top, side or front view of the model were reconstructed. Fig. 3 (b-d) shows the reconstruction of 99 model points, where 4 model points which lie on or close the reference plane were detected automatically. The reconstructed model points include the corner points of the box and a cuboid, which was used to rectify the projective reconstruction. We see that the reconstruction matches with the approximate size of the teapot which is 14.5cm (x-direction), 19.7cm (y-direction) and 15.9cm (z-direction). The ratio between the second last singular value (0.0545) and the last singular value (0.0014) was 38.9.

4.2 Infinite Reference Plane

In the first experiment with an infinite reference plane we reconstructed three buildings of the campus of the Royal Institute of Technology (KTH) in Stockholm,Sweden. 26 images of size 1600×1200 pixel were taken with a handheld camera (Olympus 3030) (see Fig. 4 (a, b)). In order to establish a correspondence between the three buildings, we also utilized a postcard of the campus (see Fig. 4 (c)). Naturally, we had no calibration information for the camera

(a) (b) (c)

Figure 4. Two original views (a, b) and a postcard (c) of the campus. The corresponding camera positions are labeled in the top view (Fig. 8).

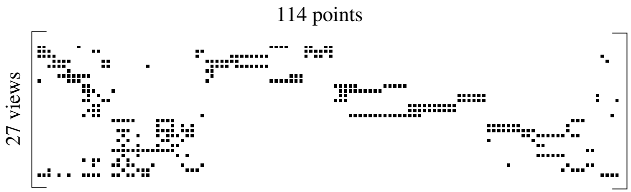


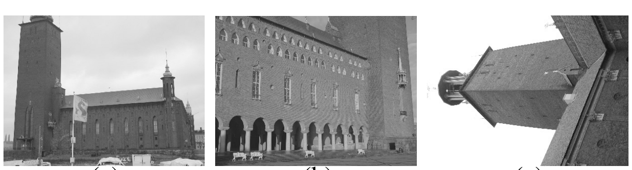
Figure 5. The visibility matrix of the campus with 27 images and 114 model points.

used for the postcard. On the basis of 114 manually selected points and orthogonal edges K and R were determined for each view and the campus was reconstructed in one single SVD (see Fig. 8). We stress that no further constraints, e.g. orthogonality, were imposed, which would presumably improve the reconstruction. By manually selecting model points which lie on same model planes a textured VRML model was created (see Fig. 10 (a-d)).

Let us consider the results. The last and second last singular value of the SVD were 12.55 and 143.5 respectively, which corresponds to a ratio of 11.44. The average error between selected image points and back-projected model points was 0.83 pixels. The respective maximum error was 35.2 pixels, which is 1.8% of the image diagonal. The accurate match between the top view of the reconstruction and the true map of the campus (see Fig. 8) demonstrates the high quality of the reconstruction.

The matrix in Fig. 5, denoted as the *visibility matrix* V, depicts the 114 points partly visible in 27 images. If the jth point is visible in the ith view, the corresponding element $V(i, j)$ is set (a black square). We see that the matrix is only sparsely filled, i.e. 10.4% of the entries are set.

In the second experiment with an infinity reference plane we reconstructed the outside and inside (courtyard) of the City Hall in Stockholm,Sweden on the basis of 35 images (see Fig. 6). Since some parts of the building can be seen from both the outside and inside, e.g. the tower (see Fig. 6 (a-c)), a correspondence between the outside and inside can be established. With the knowledge of the correspondences of 129 model points, the building was reconstructed in one single SVD (see the top view in Fig. 9 and VRML model in Fig. 10 (e-h)). As in the previous example, no further constraints were imposed in order to improve the reconstruction.



(a) (b) (c)

Figure 6. Three original views of the City Hall. The camera positions are labeled in the top view (Fig. 9).

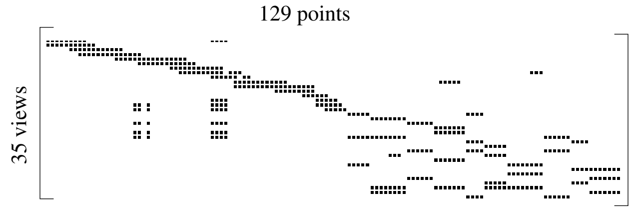


Figure 7. The visibility matrix of the City Hall with 35 images and 129 model points.

The ratio between the second last singular value (57.24) and the last singular value (12.75) was 4.49. The average error between selected image points and back-projected model points was 0.81 pixels. The respective maximum error was 97.31 pixels, which is 4.9% of the image diagonal. Let us consider the quality of the reconstruction (see Fig. 9). It is clear that the building was not designed as a perfect rectangular building. However, this fact does not considerably affect the quality of the reconstruction. The fact that the detected vanishing points are not perfectly mutually orthogonal influences the camera calibration as well as the estimation of the rotation matrix R. Since the accuracy of R directly affects the camera's position, we would expect a higher "positioning error" for cameras with less accurate R. This reasoning would explain the deviation between the reconstruction and the true map at the top, left corner of the building.

The visibility matrix in Fig. 7 depicts the 129 model points partly visible in the 35 images. The upper half of the matrix comprises images of the outside of the building. Most of these correspondences between points and images are close to the diagonal of the matrix. This reflects the fact that model points appear and disappear as the camera moves around the outside of the building. The lower half of the matrix which represents images of the inside of the building is less structured. This is due to the fact that the strategy of taking pictures was more complex.

5 Summary and Conclusions

We have demonstrated theoretically and experimentally that points and camera centres in a multi view situation can be simultaneously, linearly, projectively reconstructed by computing the null-space of a matrix built from image coordinates in an arbitrary number of views. The only specific requirement is to have four coplanar points visible in all views. This results in a substantial simplification relative to previous algorithms for multi view reconstruction and calibration that e.g. rely on systematic procedures for exploiting two or three views at a time [5]. The fact that, contrary to factorization algorithms [25, 23], we do not need to have all points visible in all views gives a very flexible algorithm for the reconstruction of e.g. architectural environments where the reference plane can be chosen as the plane at infinity using vanishing points as the reference points. Furthermore, we showed how to deal with the problem of choosing the reference plane as the plane at infinity in the specific projective representation.

Experimental results indicate that the use of an arbitrary number of cameras leads to numerically robust reconstructions which can be expected since large mutual baselines are exploited. We consider this as a major practical advantage over existing algorithms.

The linearity and specific symmetry relation between points and camera centres implies that any analysis of critical configurations, minimal cases and numerical stability will be easier (see [21] for preliminary results).

References

[1] Carlsson, S. 1995. Duality of reconstruction and positioning from projective views. In P. Anandan, editor, IEEE Workshop on Representation of Visual Scenes, Boston, USA.

[2] Carlsson, S. and Weinshall, D. 1998. Dual computation of projective shape and camera positions from multiple images. In Int. J. Computer Vision, 27(3):227-241.

[3] Criminisi, A., Reid, I. and Zisserman A. 1998. Duality, rigidity and planar parallax. In European Conf. Computer Vision, Springer-Verlag, Freiburg, Germany, pages 846-861.

[4] Cross, G., Fitzgibbon, A. W. and Zisserman A. 1999. Parallax geometry of smooth surfaces in multiple views. In Int. Conf. Computer Vision, Kerkyra, Greece, pages 323-329.

[5] Fitzgibbon, A. W. and Zisserman, A. 1998. Automatic camera recovery for closed or open image sequences. In European Conf. Computer Vision, Freiburg, Germany, pages 311-326.

[6] Hartley, R. and Debunne, G. 1998. Dualizing scene reconstruction algorithms. In R. Koch and L. Van Gool, editors, Workshop on 3D Structure from Multiple Images of Large-scale Environments (SMILE-98), LNCS 1506, Springer-Verlag, pages 14-31.

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

[8] Hartley, R., Dano, N. and Kaucic, R. 2001. Plane-based Projective Reconstruction. In Int. Conf. Computer Vision, Vancouver, Canada, (to appear).

[9] Heyden, A., Berthilsson, R. and Sparr, G. 1999. An iterative factorization method for projective structure and motion from image sequences. Image and Vision Computing, 17(13), pages 981-991.

[10] Irani, M. and Anandan, P. 1996. Parallax geometry of pairs of points for 3d scene analysis. In European Conf. Computer Vision, Springer-Verlag, Cambridge, UK, pages 17-30.

[11] Irani, M., Anandan, P. and Weinshall, D. 1998. From reference frames to reference planes: Multi-view parallax geometry and applications. In European Conf. Computer Vision, Springer-Verlag, Freiburg, Germany, pages 829-845.

[12] Jacobs, D. 1997. Linear Fitting with Missing Data for Structure-from-Motion. In IEEE Conf. Computer Vision and Pattern Recognition, San Juan, Puerto Rico, pages 206-212.

[13] Koch, R., Pollefeys, M. and VanGool, L. 1998. Multi viewpoint stereo from uncalibrated video sequences. In European Conf. Computer Vision, Springer-Verlag, Freiburg, Germany, pages 55-65.

[14] Kumar, R., Anandan, P. and Hanna, K. 1994. Direct recovery of shape from multiple views: a parallax based approach. Int. Conf. Pattern Recognition, Jerusalem, Israel, pages 685-688.

[15] Liebowitz, D. and Zisserman, A. 1999. Combining Scene and Auto-Calibration Constraints. In Int. Conf. Computer Vision, Kerkyra, Greece, pages 293-300.

[16] Oliensis, J. 1995. Multiframe Structure from Motion in Perspective. In Workshop on Representations of Visual Scenes, Boston, USA, pages 77-84.

[17] Qian, C. and Medioni, G. 1999. Efficient Iterative Solution to M-View Projective Reconstruction Problem. In IEEE Conf. Computer Vision and Pattern Recognition, Fort Collins, Colorado, pages 55-61.

[18] Quan, L. 1994. Invariants of 6 points from 3 uncalibrated images. In European Conf. Computer Vision, Springer-Verlag, Stockholm, Sweden, pages 459-470.

[19] Quan, L., Heyden A. and Kahl, F. 1999. Minimal Projective Reconstruction with Missing Data. In IEEE Conf. Computer Vision and Pattern Recognition, Fort Collins, Colorado, pages 210-216.

[20] Rother, C. 2000. A new Approach for Vanishing Point Detection in Architectural Environments. In 11th British Machine Vision Conference, Bristol, UK, pp. 382-391.

[21] Rother, C. and Carlsson S. 2001. Linear Multi View Reconstruction and Camera Recovery. In Tecnical Report CVAP 250, ISRN KTH/NA/P–01/08–SE.

[22] Schaffalitzky, F., Zisserman, A., Hartley, R. I. and Torr, P.H.S. 2000. A Six Point Solution for Structure and Motion. In European Conf. Computer Vision, Springer-Verlag, Dublin, Ireland, pages 632-648.

[23] Sturm, P. and Triggs, B. 1996. A factorization based algorithm for multi-image projective structure and motion. In European Conf. Computer Vision, Springer-Verlag, Cambridge, U.K., pages 709-719.

[24] Svedberg, D. and Carlsson, S. 1999. Calibration, Pose and Novel Views from Single Images of Constrained Scenes. In Scandanavian Conf. Image Analysis, Kangerlussuaq, Greenland, pages 111-118.

[25] Tomasi, C. and Kanade, T. 1992. Shape and motion from image streams under orthography: a factorization method. In Int. J. Computer Vision, 9(2):137-54.

[26] Triggs, B. 2000. Plane + Parallax, Tensors and Factorization. In European Conf. Computer Vision, Springer-Verlag, Dublin, Ireland, pages 522-538.

[27] Weinshall, D., Werman, M. and Shashua, A. 1995. Shape tensors for efficient and learnable indexing. In IEEE Workshop on Representation of Visual Scenes, June 1995, pages 58-65.

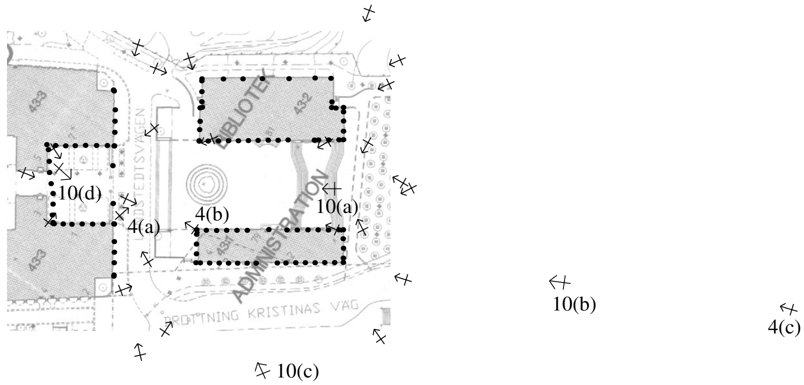


Figure 8. Top view of the reconstruction of the campus with 114 model points (dots) and 27 cameras (arrows). A map of the campus is superimposed. The labeled cameras correspond to images in the respective figures.

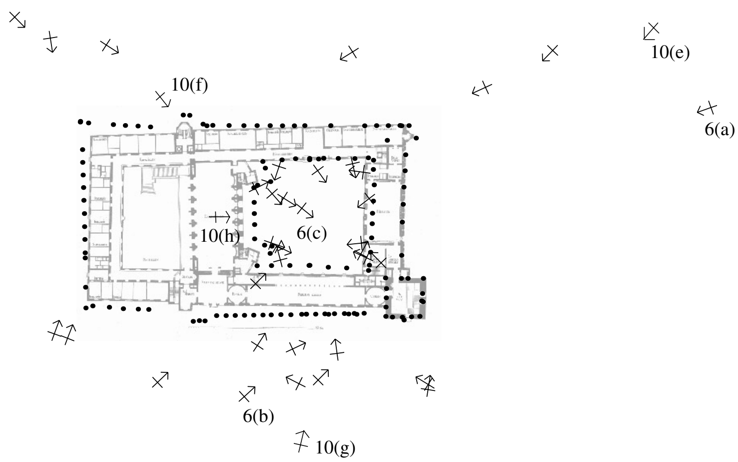


Figure 9. Top view of the reconstruction of the City Hall with 129 model points (dots) and 35 cameras (arrows). A map of the City Hall is superimposed. The labeled cameras correspond to images in the respective figures.

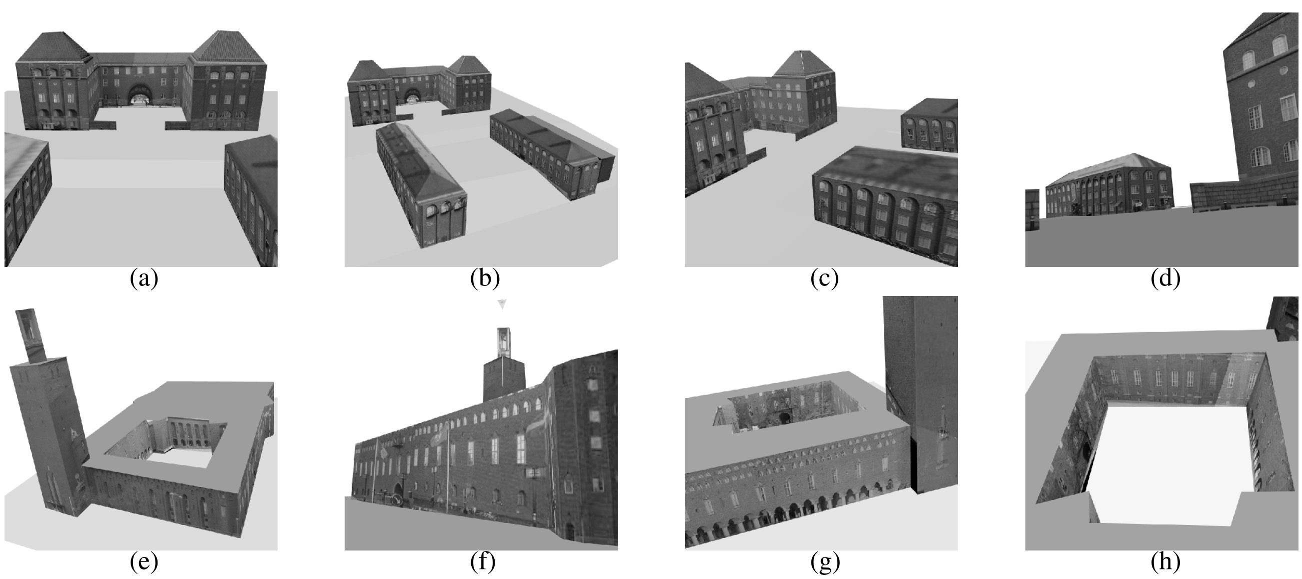


Figure 10. Novel views of the campus (a-d) and City Hall (e-h). The views are labeled in Fig. 8 (campus) and Fig. 9 (City Hall).