International Journal of Computer Vision 49(2/3), 117–141, 2002


# Linear Multi View Reconstruction and Camera Recovery Using a Reference Plane

CARSTEN ROTHER AND STEFAN CARLSSON
*Computational Vision and Active Perception Laboratory (CVAP), Department of Numerical Analysis and Computer Science, KTH, SE-100 44 Stockholm, Sweden*
carstenr@nada.kth.se
stefanc@nada.kth.se

*Received March 14, 2001; Revised December 26, 2001; Accepted March 1, 2002*

**Abstract.** This paper presents a linear algorithm for simultaneous computation of 3D points and camera positions from multiple perspective views based on having 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 built from image data using Singular Value Decomposition. Contrary to factorization algorithms this approach does not need to have all points visible in all views. This paper investigates two reference plane configurations: Finite reference planes defined by four coplanar points and infinite reference planes defined by vanishing points. A further contribution of this paper is the study of critical configurations for configurations with four coplanar points. By simultaneously reconstructing points and views we can exploit the numerical stabilizing effect of having wide spread cameras with large mutual baselines. This is demonstrated by reconstructing the outside and inside (courtyard) of a building on the basis of 35 views in one single Singular Value Decomposition.

**Keywords:** structure from motion, projective reconstruction, multiple views, missing data, duality, critical configurations, reference plane, planar parallax

## 1. Introduction

The efficient computation of 3D structure and camera information from multiple views has been a subject of considerable interest in recent years (Hartley and Zisserman, 2000; Faugeras and Luong, 2001). The problem can be formulated most generally as a bi-linear inverse problem (including unknown scale factors) for finding camera and 3D information from image data. Contrary to the case of parallel projection (Tomasi and Kanade, 1992) no algorithm for direct factorization of camera parameters and 3D structure has been produced for perspective projection cameras. The perspective factorization algorithm suggested in Sturm and Triggs (1996) relies on the pre-computation of scale factors "projective depths" in order to cast the problem into the same form as in Tomasi and Kanade (1992). Approaches have been invented for efficient combination of groups of views (Fitzgibbon and Zisserman, 1998; Koch et al., 1998), or iterative methods exploiting all views (Heyden et al., 1999). The approach in Quan (1994) using "shape constraints" dual to epipolar constraints (Carlsson, 1995; Carlsson and Weinshall, 1998; Weinshall et al., 1995) can in principle be exploited for computing projective structure using arbitrary number of views (Hartley and Debunne, 1998; Schaffalitzky et al., 2000). It is however limited to a restricted number of points at a time.

Ideally an algorithm for reconstructing camera and scene information from multiple views should exploit all points and views simultaneously as in Tomasi and Kanade (1992). 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 (Jacobs, 1997; Qian and Medioni, 1999; Quan and Heyden, 1999).

In this paper we will show that by adding the simple assumption of having four points on a reference plane visible in all views (Kumar et al., 1994; Heyden and Åström, 1995, 1997; Triggs, 2000), 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 has been observed and discussed (Oliensis, 1995, 1999; Heyden and Åström, 1995; Oliensis and Genc, 1999; Triggs, 2000) but it seems that its full potential for reconstruction and camera recovery has not yet been exploited. A more detailed discussion and comparison of these methods to our approach will be given later. The advantage that the constraints, given by the 2, 3 or 4 view tensors, become linear if a plane is visible in all views (Heyden and Åström, 1995; Heyden, 1998) has been exploited in Hartley and Zisserman (2000) and Hartley et al. (2001). However, the number of geometrically corresponding views is limited to four and structure and motion cannot be simultaneously reconstructed.

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 (Carlsson, 1995; Carlsson and Weinshall, 1998; Weinshall et al., 1995), and the dual structures that arise with a reference plane (Criminisi et al., 1998; Irani and Anandan, 1996; Irani et al., 1998; Weinshall et al., 1998). Most of the past work which studied the geometry for scenes containing a reference plane, i.e. parallax geometry, focused on the reconstruction of scene points for two views (Kumar et al., 1994; Irani and Anandan, 1996; Criminisi et al., 1998) or multiple views (Irani et al., 1998). The reference plane formulation of this paper, which reveals the linear relationship between points and camera centers in multiple views, can therefore been seen as an extension and simplification of most planar parallax approaches.

A potential problem for numerical calculations is the fact that the reference plane will be at infinity in the representation that linearizes the problem. The consequence is that points which are on or close to the reference plane have to be reconstructed separately. 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. As a practical demonstration of this we consider the reference plane at infinity spanned by three mutual orthogonal vanishing points obtained from viewing typical architectural structures. Multiple points are viewed with only partial overlap. 3D positions of points and camera centers are reconstructed using a single Singular Value Decomposition based on all observed points in all views simultaneously.

The linearity and symmetry of space points and camera positions makes it especially easy and interesting to investigate problems of numerical stability and critical configurations of points and cameras in the scene. In this paper the problem of critical configurations for the special case of having four coplanar points will be discussed. A configuration of points and cameras is critical if the projected image points are insufficient to determine the points and cameras uniquely, up to a projective transformation. We will show that if all points are visible in all views, i.e. no missing data, all configuration (apart from trivial ones) where points and camera centers are non-coplanar are non-critical. If not all points are visible in all views, i.e. missing data, a method to construct non-critical configurations is proposed.

The content of this paper is based on Rother and Carlsson (2001). However, in contrast to Rother and Carlsson (2001) this paper presents novel results about critical configurations for the case of four coplanar scene points. Furthermore, the reconstruction algorithms and the underlying theory are presented in more detail.

## 2. Duality, Symmetry and Linearity of Projection Relations

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

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

where $P$ and $\bar{P}$ are the homogeneous and non-homogeneous cartesian coordinates of the 3D-points respectively. The $3 \times 3$ matrix $I$ is the identity matrix and $\bar{Q}$ are cartesian coordinates of the camera centers. The $3 \times 4$ matrix $H(I \mid -\bar{Q})$ represents the camera matrix. In a general projective representation the homography $H$ will be factored out and we are left with relations between 3-D points and camera centers. Already from Eq. (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 $P_1, \ldots, P_5$ as basis, any point $P$ and camera center $Q$ in 3D can be expressed using projective coordinates:

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

Similarly, four image points $p_1, \ldots, p_4$ can be used as a basis for expressing image coordinates

$$p \sim xp_1^* + yp_2^* + wp_3^*. \tag{3}$$

The normalizations $P^*$ and $p^*$ are chosen so that points $P_5$ and $p_4$ get projective coordinates $(1, 1, 1, 1)^T$ and $(1, 1, 1)^T$ respectively. Specifying 5 scene points and 4 image points fixes the 15 degrees of freedom of the projective space $P^3$ and the 8 degrees of freedom projective space $P^2$. Choosing a specific basis in an image implies that a projective transformation has to be applied to the observed image points which are in a camera specific basis.

The mapping of scene points to image points

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

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

### 2.1. General Point Configurations

This case was treated in e.g. Faugeras (1992), Quan (1994) and Carlsson (1995) and it means that we take the image basis points $p_1$, $p_2$, $p_3$, $p_4$ as projections of the 3D basis points $P_1$, $P_2$, $P_3$, $P_4$ (see Fig. 1(a)). We

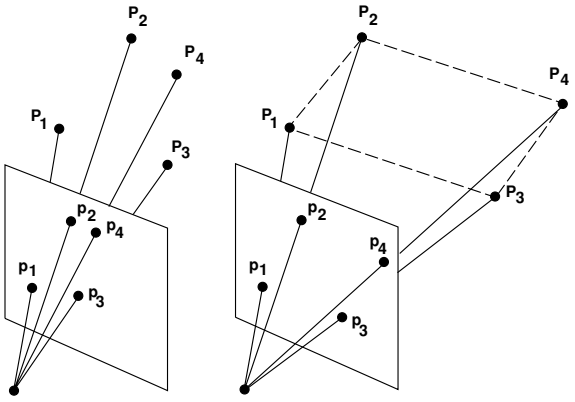



*Figure 1.* Projective basis points in 3D $P_1, \ldots, P_4$ and image $p_1, \ldots, p_4$ for general configurations (a) and reference plane configurations (b).

then get the constraints on the mapping $M$:

$$M: \begin{array}{ccccc} P_1 & P_2 & P_3 & P_4 & Q \\ - & - & - & - & - \\ 1 & 0 & 0 & 0 & A \\ 0 & 1 & 0 & 0 & B \\ 0 & 0 & 1 & 0 & C \\ 0 & 0 & 0 & 1 & D \end{array} \longrightarrow \begin{array}{ccccc} p_1 & p_2 & p_3 & p_4 & 0 \\ - & - & - & - & - \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 \end{array}. \tag{5}$$

This results in the following projection relations:

$$\begin{matrix} x \\ y \\ w \end{matrix} \sim \begin{matrix} \frac{X}{A} - \frac{W}{D} \\ \frac{Y}{B} - \frac{W}{D} \\ \frac{Z}{C} - \frac{W}{D} \end{matrix} \tag{6}$$

which can be written as two constrained equations:

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

These relations make explicit the duality of space points and camera centers in the sense that the *homo-geneous* projective coordinates of space points $X, Y, Z, W$ and inverse coordinates of camera centers $A^{-1}, B^{-1}, C^{-1}, D^{-1}$ are *bilinearly* related in a symmetric way.

For completeness, Eq. (1) can be explicitly written as:

$$\begin{pmatrix} x \\ y \\ w \end{pmatrix} \sim \begin{pmatrix} A^{-1} & 0 & 0 \\ 0 & B^{-1} & 0 \\ 0 & 0 & C^{-1} \end{pmatrix} \times \begin{pmatrix} 1 & 0 & 0 & -A/D \\ 0 & 1 & 0 & -B/D \\ 0 & 0 & 1 & -C/D \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ W \end{pmatrix} \quad (8)$$

where $\bar{Q} = (A/D, B/D, C/D)^T$ as the non-homogeneous camera center. This means that the homography $H$ in Eq. (1) depends on the camera centers. The choice of the fifth basis point $P_5$ has further consequences (see e.g., Carlsson, 1995) which are, however, not relevant in this context.

### 2.2. *Four Points on a Reference Plane*

If we have four points $P_1, P_2, P_3, P_4$ on a reference plane visible in all views we can use the images of these 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 centers. Note, in this case $P_4$ can not be used as a basis point for the projective basis, which has to consist of five non-coplanar points. How the remaining degrees of freedom of the projective space, i.e. $P_4$, $P_5$, of the projective space, are fixed will be discussed later. Let us choose the point $P_4$ in a canonical way as in e.g. Heyden and Åström (1995) and Triggs (2000). The constraints on the mapping $M$ become:

$$M: \begin{array}{ccccc} P_1 & P_2 & P_3 & P_4 & Q \\ - & - & - & - & - \\ 1 & 0 & 0 & 1 & A \\ 0 & 1 & 0 & 1 & B \\ 0 & 0 & 1 & 1 & C \\ 0 & 0 & 0 & 0 & D \end{array} \longrightarrow \begin{array}{ccccc} p_1 & p_2 & p_3 & p_4 & 0 \\ - & - & - & - & - \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 \end{array}. \quad (9)$$

We see that the four points define the plane at infinity, which is specified by $W = 0$. Using this we can compute the projection relations:

$$\begin{matrix} x \\ y \\ w \end{matrix} \sim \begin{matrix} \frac{X}{W} - \frac{A}{D} \\ \frac{Y}{W} - \frac{B}{D} \\ \frac{Z}{W} - \frac{C}{D} \end{matrix}. \quad (10)$$

If we compare this to the general projection relations in Eq. (6) we see that the relationship between points and cameras is different, however, still symmetric. The symmetry now relates to the substitutions $(X, Y, Z, W) \leftrightarrow (A, B, C, D)$. 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, i.e. $W \neq 0$. We can now rewrite the projection relations as linear constrained equations:

$$\begin{aligned} x\left(\frac{Z}{W} - \frac{C}{D}\right) - w\left(\frac{X}{W} - \frac{A}{D}\right) &= 0 \\ y\left(\frac{Z}{W} - \frac{C}{D}\right) - w\left(\frac{Y}{W} - \frac{B}{D}\right) &= 0 \\ x\left(\frac{Y}{W} - \frac{B}{D}\right) - y\left(\frac{X}{W} - \frac{A}{D}\right) &= 0. \end{aligned} \quad (11)$$

Obviously only two of the three projection relations are linearly independent. However, two relations are insufficient for special cases where e.g. $w = 0$ and $\frac{Z}{W} - \frac{C}{D} = 0$.

Arbitrary numbers of points and views can be used to build *one* matrix consisting of *all* projection relations in terms of image coordinates. For $n$ points in $m$ views the linear system takes the form:

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

for non-homogeneous projective point coordinates $\bar{X} = X/W$ etc. and camera centers $\bar{A} = A/D$ etc.

where

$$S_{i,j} = \begin{pmatrix} 0 & w_{i,j} & -y_{i,j} \\ -w_{i,j} & 0 & x_{i,j} \\ y_{i,j} & -x_{i,j} & 0 \end{pmatrix} \tag{13}$$

are $3 \times 3$ matrices built up from image coordinates of point $i$ visible in view $j$. In the following we denote the matrix which forms the linear system in Eq. (12) as the *S-matrix*.

Before the solution of points and cameras can be obtained from the linear system, the projective space defined by the four coplanar points have to be considered in more detail. With the four points $P_1$, $P_2$, $P_3$, $P_4$ on the plane at infinity (Eq. (1)) can be explicitly written as:

$$\begin{pmatrix} x \\ y \\ w \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & 0 & -A/D \\ 0 & 1 & 0 & -B/D \\ 0 & 0 & 1 & -C/D \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ W \end{pmatrix} \tag{14}$$

or more compact as (see as well Eq. (10)):

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

We see that the homography $H$ in Eq. (1) is now the identity matrix. This means that having four coplanar points is projectively equivalent to having purely translating calibrated cameras. This simple result was already stated in e.g. Heyden and Åström (1995) and Triggs (2000), however, it was differently exploited in contrast to this paper.

It is well known that a 3D projective space has 15 degrees of freedom. This can be expressed as a $4 \times 4$ homography which transforms a point $P$ to $P'$ as:

$$P' = \begin{pmatrix} A & t \\ b^T & \lambda \end{pmatrix} P, \tag{16}$$

where $A$ is a $3 \times 3$ matrix, $b^T, t$ are 3 dimensional vectors and $\lambda$ a scalar (see e.g. Hartley and Zisserman 2000; Faugeras and Luong, 2001). The choice of the points $P_1$, $P_2$, $P_3$, $P_4$ as in Eq. (9) implies that $A = \mu I$ and $b^T = (0, 0, 0)$. This means that 11 of the 15 degrees of freedom of the projective space are fixed. The remaining 4 degrees of freedom correspond to the arbitrary choice of $t$, $\mu$ and $\lambda$ (minus an overall scale).

Let us apply a Singular Value Decomposition (SVD) on the $S$-matrix (Eq. (12)) which gives the null-space of the $S$-matrix. Since the chosen projective space has 4 degrees of freedom, the null-space of $S$ is at least of dimension 4. However, three of the four singular vectors of the null-space have the trivial form: $\bar{P}_i = \bar{Q}_j = (1, 0, 0)^T$, $\bar{P}_i = \bar{Q}_j = (0, 1, 0)^T$ and $\bar{P}_i = \bar{Q}_j = (0, 0, 1)^T$. This reflects the fact that the translation $t$ in Eq. (16) can be chosen arbitrarily. Therefore, if $S$ has a four dimensional null-space, the summation of all four singular vectors of the null-space gives the non-trivial solution for all camera centers and points. However, certain configurations of points and cameras might give a null-space of dimension larger than 4. Such configurations are called *critical* and they will be the subject of Section 5.

To summarize, the simple addition of a fourth coplanar point 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. In contrast to this, reconstruction algorithms for general scenes are not that straight forward. Factorization based methods, e.g. Sturm and Triggs (1966), Sparr (1996), and Heyden et al. (1999) have to determine fundamental matrices (Sturm and Triggs, 1966) or iterate the solution (Sparr, 1996; Heyden et al., 1999) in order to obtain depth scale factors. Other methods, e.g. Fitzgibbons and Zisserman (1998) and Koch et al. (1998) have to determine complete camera projection matrices before doing reconstruction.

Utilizing the homography $H$ in Eq. (1) to linearize the reconstruction problem has also been exploited by Oliensis (1995, 1999) and Oliensis and Genc (1999). It is known (Hartley and Zisserman, 2000) that corresponding image points of purely rotating cameras define the homography: $H = K'RK^{-1}$, where $K$, $K'$ is the calibration matrix of the first and second camera and $R$ the rotation between them. The basic assumption in Oliensis work is a small movement of the camera between successive frames. This means that $H$ can be approximately determined. With the knowledge of $H$, the relationship between points and camera centers can be linearized by applying the inverse homography on the image points: $H^{-1}p \sim \bar{P} - \bar{Q}$ (compare Eq. (1)). This is used to initialization an iterative reconstruction algorithm (Oliensis, 1999). Furthermore, if the calibration is known, i.e. $K$ and $K'$, the rotation $R$ can be determined and a Euclidean reconstruction may be obtained (Oliensis and Genc, 1999).

## 3. Finite versus Infinite Reference Plane

The relative simplicity of the $S$-matrix hides a potentially disturbing fact for numerical calculations. The linearity is expressed in *non-homogeneous* projective coordinates $X/W$ etc. for points and $A/D$ etc. for cameras. Therefore, for points on the reference plane we have $W = 0$, i.e. they are moved to infinity in the projective basis representation. This was noted in Triggs (2000) as a fundamental requirement for obtaining this simple structure of purely translating calibrated cameras. In contrast to the reconstruction method in Triggs (2000), this has consequences for points which are on or close to the reference plane. In this section 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. However, having 4 points on the plane at infinity, which are visible in all views, in general constrains the camera positions in a multi view situation. When we consider the practically interesting case of architectural scenes, we see that they are often characterized by orthogonal directions, i.e three orthogonal vanishing points. 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.

### 3.1. Finite Reference Plane

We see that the reference plane is the plane at infinity, i.e. $W = 0$, in the projective basis representation. This means that only those points can be reconstructed by the linear system in (12) which do not lie on the reference plane. Therefore, points which are on the reference plane have to be reconstructed separately. However, points on the reference plane are particularly easy to determine. From Eq. (14) we see that a point at infinity $(X, Y, Z, 0)^T$ can be reconstructed directly as:

$$\begin{pmatrix} x \\ y \\ w \end{pmatrix} \sim \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}. \tag{17}$$

Let us consider the case if we put a point which lies on the reference plane, e.g. $P_1 = (X_1, Y_1, Z_1, 0)$, into the linear system (12). The projection relations (17) can be written as linear constrained equations:

$$\begin{aligned} xZ_1 - wX_1 &= 0 \\ yZ_1 - wY_1 &= 0 \cdot \\ xY_1 - yX_1 &= 0 \end{aligned} \tag{18}$$

In this case the submatrices $S_{1,j}$ contain these equations instead of the equations in Eq. (11). We see that, independent of all camera centers and all other points, the vector $(X_1, Y_1, Z_1, 0, \ldots, 0)$ represents an additional solution to the linear system (12). This means that we obtain a five-dimensional null-space for the $S$-matrix, i.e. 2 non-trivial solutions for the points and cameras.

Although in practice points are seldomly exactly on the reference plane, the linear system gets numerically instable if points which are "close" to the reference plane are included into the linear system (12). Firstly, due to errors in the image coordinates the singular vector which does *not* represent the complete solution for all points and cameras could have a smaller singular value than the singular vector which does represent the complete solution. Secondly, the coordinates of the points which are not close to the reference plane have very small values in contrast to the coordinates of the points which are close to the reference plane. This potentially increases the inaccuracy in all those points which are not close to the reference plane.[1]

Therefore, the points which are on or close to the reference plane have to be excluded from the linear system (12). This means that we have to decide for each point whether it lies on (or close to) the reference plane or not. However, this decision can be achieved with the knowledge of the image coordinates of the four reference points $P_1$, $P_2$, $P_3$ and $P_4$. Since these points lie on the reference plane and are visible in all views they introduce a planar homography between each pair of views. Therefore, we can determine for each pair of views, in which a certain world point is visible, a residual parallax vector (see Irani and Anandan, 1996; Hartley and Zisserman, 2000). It has been shown that in case of parallel projection the parallax vector depends only on the distance between the world point and the reference plane (see Irani and Anandan, 1996; Kumar et al., 1994). This result can be utilized for projective cameras as an approximation. Therefore, a ranking of all world points with respect to their distance to the reference plane can be obtained on the basis of the magnitude of their parallax vectors. Points on and off the plane can now be separated by defining a threshold. However, the choice of such a threshold depends

on the scene and camera motion, i.e. the distance between scene points and the camera. This choice can be circumvented by successively excluding points from the linear system (12) on the basis of this ranking, i.e. distance to the reference plane. This leads, however, to an iterative algorithm. An explicit formulation of the algorithm will be given in Section 4.

### 3.2. *Choosing a Reference Plane or Reference Points*

The plane plus points configurations, i.e. a reference plane and points which lie not on the reference plane, has received significant attention in the past e.g. Kumar et al. (1994), Irani and Anandan (1996), Criminisi et al. (1998), Irani et al. (1998), Cross et al. (1999), and Hartley et al. (2001). These approaches are characterized by the basic assumption of a reference plane visible in all views. This assumption seems to be more general in contrast to the assumption we do, i.e. four reference points visible in all views. However, we will see that those four reference points can be easily derived from the general assumption of a visible reference plane.

A reference plane visible in all views introduces planar homographies between each pair of views. Let us introduce four "virtual" basis points which lie on the reference plane. The position of these points can be fixed by choosing their image coordinates in one arbitrary view. With the use of the inter-view planar homographies, the image coordinates of these points can be established in all other views. Therefore, these four "virtual" basis points can be used as the points $P_1, P_2, P_3, P_4$ as in the previous section. This means that a reference plane visible in all views is sufficient for establishing a unique projective basis for all points and cameras.

However, depending on the image coordinates some inter-view homographies might be inaccurate. This could introduce a substantial numerical instability in the reconstruction process. In order to avoid this source of error, we concentrate our current interest on the special configuration of four points visible in all views.

### 3.3. *Reference Plane at Infinity*

If the points $P_1, P_2, P_3$ 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 $P_3 - P_4$. The point $P_4$ represents the origin of the affine system which is not at infinity. If these directions are specialized to being 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. The main advantage in these kinds of scenes is the use of images estimated for very different camera positions. However, the exploitation of the infinite reference plane needs additionally the image of a fourth coplanar reference point. Having a specific point at infinity visible in all views will substantially restrict the usefulness of the infinite reference plane for the general perspective camera. By considering a special case of perspective cameras, however, we can make use of the reference plane at infinity given by the three orthogonal vanishing points only.

For the special case of perspective camera:

$$p \sim K R(I \mid -\bar{Q})P \tag{19}$$

where $K$ contains the internal camera parameters:

$$K = \begin{pmatrix} \sigma & 0 & \bar{x}_0 \\ 0 & \sigma & \bar{y}_0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{20}$$

i.e. 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 (Caprile and Torre, 1990; Liebowitz and Zisserman, 1999; Svedberg and Carlsson, 1999). Normalizing image coordinates with this knowledge we can write the projection relation as:

$$p' \sim \bar{P} - \bar{Q}, \tag{21}$$

where $p' = R^T K^{-1} p$. This is the same situation of purely translating calibrated cameras just as derived using the pure projective representation in the previous section. The fact that we have knowledge of the projection of vanishing points at infinity together with their orthogonality implies that we can compute a metric reconstruction, i.e. 3D structure up to similarity transformations.

### 3.4. *Determination of K and R*

In order to identify the internal camera parameters $K$ and camera rotations $R$, mutual orthogonal directions

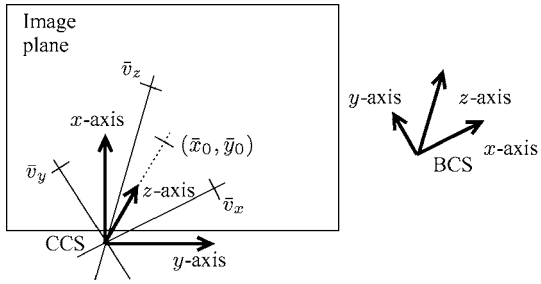



*Figure 2.* Relationship between the CCS and the BCS, which is defined by the orthogonal vanishing points $\bar{v}_x$, $\bar{v}_y$ and $\bar{v}_z$.

in the scene have to be detected. Although this task has recently raised great interest e.g. Rother (2000), it is manually performed in the current version of the algorithm. This has the reason that falsely detected vanishing points might significantly reduce the quality of the reconstruction.

Since the three mutual orthogonal vanishing points $\bar{v}_x$, $\bar{v}_y$ and $\bar{v}_z$ are visible in all views, they specify the directions of a 3D, cartesian basis coordinate system (BCS). Figure 2 shows the geometrical relation between the BCS and a camera coordinate system (CCS). The orthogonality relation of the three vanishing points can be algebraicly defined as: $\langle K^{-1}\bar{v}_x, K^{-1}\bar{v}_y\rangle = 0$, $\langle K^{-1}\bar{v}_x, K^{-1}\bar{v}_z\rangle = 0$ and $\langle K^{-1}\bar{v}_y, K^{-1}\bar{v}_z\rangle = 0$, with $\langle\cdot,\cdot\rangle$ as scalar product. From these equations the focal length $\sigma$ and the principal point $(\bar{x}_0, \bar{y}_0)$ of the specific camera model introduced in Eq. (19) can be derived. However, in case one or two of the vanishing points are at infinity (so-called degenerated cases) with respect to the image plane, i.e the image plane is parallel to the corresponding axis of the BCS, not all internal camera parameters are determinable (Liebowitz and Zisserman, 1999; Rother, 2000). In the current version of the algorithm, we assume fixed internal camera parameters for the process of acquiring images. This allows us to improve the camera calibration significantly, by averaging all those internal camera parameters which were derived from non-degenerated cases.

With the knowledge of $K$, the rotation matrix $R$ can be determined. For that the correspondence problem between the three vanishing points and the $x$-, $y$- and $z$-axis of the BCS has to be solved. Furthermore, directions given by each vanishing point have to be uniquely defined, i.e. the sign of $K^{-1}\bar{v}_{x,y,z}$ has to be determined. We define:

$$R = (\pm K^{-1}\bar{v}_x | \pm K^{-1}\bar{v}_y | \pm K^{-1}\bar{v}_z) \quad \text{with } \det(R) = 1.$$

Since the condition $\det(R) = 1$ has to be fulfilled, we obtain 24 possible $R$ in case of unknown correspondence and 4 possible $R$ otherwise. Note, for the determination of $R$ two of the three orthogonal vanishing points are sufficient, since $\bar{v}_z = \bar{v}_x \times \bar{v}_y$. In the current version of the algorithm, the ambiguity in $R$ is manually solved.

## 4. Outline of the Algorithm and Optimization

We have seen that with the use of a reference plane points and cameras can be simultaneously reconstructed in closed-form. However, the prize we have to pay is that an algebraic error function is minimized. Such an error is suboptimal in contrast to a geometric error, e.g. the Euclidean distance between image points and reprojected scene points. In this section this algebraic error function will be analysed and we will investigate how it can be optimized with the restriction of having a closed-form solution. The section concludes with an explicit description of the algorithm for finite and infinite reference planes.

### 4.1. *Optimization of the Error Function*

Let us reconsider Eq. (1) where the scene point $P_i$ is mapped by the camera $j$ on the image point $p_{ij}$:

$$\lambda_{ij}\, p_{ij} = H_j(\bar{Q}_j - \bar{P}_i). \tag{22}$$

The vector $p_{ij}$ represents now the observed image point before normalization (Eq. (3)) and $\lambda_{ij}$ is an unknown scale, which is denoted in Sturm and Triggs (1996) as projective depth. We have seen in Section 2.1 that the homography $H_j$ depends on the camera center $\bar{Q}_j$ for general scenes. In case of a reference plane visible in all views, $H_j$ is the identity matrix with respect to the normalized image points (see Eq. (15)). This means that $H_j^{-1}$ represents the projective transformation of image points $p_{ij}$ into the normalized image points $p'_{ij}$ defined in Eq. (3). Explicitly written:

$$p'_{ij} = H_j^{-1} p_{ij} \quad \text{with } H_j^{-1}\text{:}$$
$$(p_1\ p_2\ p_3\ p_4) \longrightarrow \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}, \tag{23}$$

where $p_1$, $p_2$, $p_3$, $p_4$ are the projections of the 4 coplanar points $P_1$, $P_2$, $P_3$, $P_4$. In order to eliminate the

unknown scale $\lambda_{ij}$, the ratios of $x$, $y$ and $w$-coordinates of the image points $p'_{ij}$ were considered (see Eq. (11)). These ratios are the subject of minimization in the linear system of Eq. (12). Minimizing such an algebraic error is, however, statistically suboptimal in contrast to geometric error functions used e.g. for bundle adjustment (see e.g. Triggs et al., 1999).

How can this algebraic error function be improved? We have seen that the key for obtaining a closed-form solution is the linear relationship between scene points and camera centers, i.e. the knowledge of $H_j$. This linear relationship is not affected by a change of the image basis, i.e. applying a homography $B$, and by an individually scaling $s_{ij}$ of the image points $p'_{ij}$:

$$p''_{ij} \sim s_{ij} B p'_{ij} \sim B\bar{P}_i - B\bar{Q}_j \sim \bar{P}'_i - \bar{Q}'_j. \quad (24)$$

If $B$ and $s_{ij}$ are chosen, $p''_{ij}$ can be derived and we obtain a linear system as in Eq. (12):

$$S'(\bar{X}'_1, \bar{Y}'_1, \bar{Z}'_1, \ldots, \bar{X}'_n, \bar{Y}'_n, \bar{Z}'_n, \bar{A}'_1, \bar{B}'_1, \bar{C}'_1, \ldots, \bar{A}'_n, \bar{B}'_n, \bar{C}'_n)^T = 0. \quad (25)$$

The matrix $S'$ consists now of the image points $p''_{ij}$ (see Eq. (13)).

Let us consider the choice of the homography $B$ first. It has been shown in Hartley (1997) that different normalizations of the image coordinates can dramatically influence the result of a computation based on image coordinates. Hartley (1997) suggested to choose the centroid of all image coordinates as the origin and to normalize the average distance of an image point to the origin to $\sqrt{2}$. If we consider Eq. (24), such a normalization would involve to determine for each view $j$ an individual matrix $B_j$, which represents the normalization. However, such a $B_j$ would destroy the linear relationship between points and camera centers. Therefore, the matrix $B$ has to be determined independently of a certain view $j$. We define:

$$B = \frac{1}{m} \sum_{j=1}^{m} B_j / \|B_j\|_2, \quad (26)$$

where $\|\cdot\|_2$ is the Frobenius norm of a matrix and $m$ is the number of views.

We have seen in Section 2.2 that a finite reference plane has to be chosen as the plane at infinity in order to obtain the simple situation of purely translating cameras. However, this suboptimal choice can be compensated by an appropriate selection of the scale factors $s_{ij}$. Let us consider a point $P_1$ which is closer to the reference plane than another point $P_2$. By choosing the reference plane as the plane at infinity, the coordinates of the reconstructed point $\bar{P}_1$ are larger than the ones of $\bar{P}_2$. This means that in the presence of noise, the point with larger coordinates is reconstructed more accurately. In order to eliminate this favoring of certain points we suggest to choose the scale factors in Eq. (24) as $s_{ij} = dis(P_i)$, where $dis(P_i) \in [0, 1]$ denotes the distance between $P_i$ and the reference plane (see Section 3.1). This scaling just inverses the effect of moving a finite plane to infinity.[2] This means that points which are closer to the reference plane are inhibited. The same applies to the equations in the linear system of Eqs. (12) and (25) of such a point.

### 4.2. *Outline of the Algorithm*

On the basis of the previous sections the algorithm for finite and infinite reference planes can be explicitly formulated. In case of a finite reference plane the algorithm is composed of the following steps:

1. Determine 4 coplanar points and other corresponding points
2. Normalize the image basis, i.e. $p'_{ij} = H_j^{-1} p_{ij}$ (Eq. (23))
3. Calculate the distance between scene points $P_i$ and the reference plane (Section 3.1)
4. Exclude iteratively points from the $S$-matrix (or choose a threshold) (Section 3.1)
5. Determine matrix $B$ (Eq. (26))
6. Determine scales $s_{ij}$ and image points $p''_{ij} = s_{ij} \|BH_j^{-1} p'_{ij}\|_2$ (Section 4.1)
7. Obtain $\bar{P}'_i$, $\bar{Q}'_j$ by SVD (Eq. (25)) and points $P_i$ on (or close to) the reference plane with Eq. (17)
8. Take the best result on the basis of RMS-error between image points and reprojected scene points
9. Undo the basis change: $\bar{P}_i = B^{-1}\bar{P}'_i$ and $\bar{Q}_j = B^{-1}\bar{Q}'_j$

The Euclidean norm is denoted by $\|\cdot\|_2$. The quality of the reconstruction is evaluated in terms of the Root-Means-Square (RMS) error between image points and reprojected scene points. However, other criteria could be used.

If three orthogonal vanishing points are detected in the scene, the algorithm has a simpler form since finite scene points do not lie on the reference plane. The algorithm can be explicitly written as:

1. Determine 3 orthogonal vanishing points and other corresponding points
2. Calculate $K_j$, $R_j$ for each camera $j$ (Section 3.4)
3. Normalize the image basis, i.e. $p'_{ij} = H_j^{-1} p_{ij}$ (Eq. (23))
4. Determine matrix $B$ and image points $p''_{ij} = \|BH_j^{-1} p'_{ij}\|_2$ (Section 4.1)
5. Obtain $\bar{P}'_i$, $\bar{Q}'_j$ by SVD (Eq. (25))
6. Undo the basis change: $\bar{P}_i = B^{-1}\bar{P}'_i$ and $\bar{Q}_j = B^{-1}\bar{Q}'_j$

## 5. Critical Configurations and Minimal Visibility

In the following we investigate the constraints that points and cameras have to satisfy in order to obtain a unique reconstruction for the special case of having four coplanar points.

In practice, not all $n$ points are visible in all $m$ views, i.e. we have to deal with missing data. In order to specify a certain overlap between points and views we introduce the *visibility matrix* $V$. An element $V(i, j)$ of the visibility matrix is set if the $j$th point is visible in the $i$th view. The following example shows a specific visibility matrix of $n = 5$ points partly visible in $m = 3$ views:

$$V = \begin{array}{c|ccccc} & \text{points} & & & & \\ & 1 & 2 & 3 & 4 & 5 \\ \text{views}\ 1 & \bullet & \bullet & & & \\ 2 & \bullet & \bullet & \bullet & \bullet & \bullet \\ 3 & & & \bullet & \bullet & \bullet \end{array} \tag{27}$$

Note, the visibility matrix does only specify which point is visible in which view. However, it does not specify the actual placement of points and camera centers in the scene. Therefore, we denote a specific placement of points and cameras in the scene as a *configuration*. Note, a configuration is only unique up to a certain transformation, which is in our case either a projective transformation (finite reference plane) or a similarity transformation (three orthogonal vanishing points). In this context a fundamental question is: *Is a certain visibility matrix sufficient, i.e. is there at least one configuration which represents a unique reconstruction?*

Sufficient visibility does not necessarily imply a unique reconstruction. Therefore, we denote a configuration as *critical* if the visibility matrix is sufficient but the projected image points are insufficient to determine a unique reconstruction (Hartley and Zisserman, 2000). This poses a second fundamental question: *What are the critical configurations of a sufficient visibility matrix?*

Let us consider these questions for the special case of four coplanar points. For $n$ points and $m$ views the total number of degrees of freedom (dofs) of the the linear system in (12) is: $\#dofs = 3(m + n) - 4$. Note, the reference points on the reference plane are not included in this counting. In the case of a finite reference plane we had additionally 4 reference points ($P_1$, $P_2$, $P_3$, $P_4$) and for an infinite reference plane 3 orthogonal vanishing points ($P_1$, $P_2$, $P_3$). Let us consider the rank of the $S$-matrix (Eq. (12)). This is at the most $\#dofs$. If the rank of the $S$-matrix is smaller than $\#dofs$, the dimensionality of the null-space is larger than four which means that the reconstruction is not unique. We can state: *A given visibility matrix is sufficient if the rank of the S-matrix is equal to the number of dofs, for a generic set of points and camera centers, i.e. points and camera centers in "general pose"*. Furthermore, we can state: *A given configuration is critical if the rank of the S-matrix is smaller than the number of dofs for this configuration.* The question of critical configurations is not only of theoretical interest, however, from a numerical point of view we should expect instabilities whenever the $S$-matrix comes close to rank deficiency, i.e. whenever a configuration is close to a critical one.

In the past these two questions have been investigated for two different cases: *no missing data and missing data.* This corresponds to two different types of visibility matrices: full and not full visibility matrix. For the special case of four coplanar points, these two types of visibility matrices will be discussed separately as well.

Before addressing these questions, let us recapitulate the constraints on points and cameras we have so far. If we reconstruct points which lie on the reference plane with the linear system in (12), we have seen, that we do not obtain a unique (up to 4 dofs) solution. However, points on the reference plane can be uniquely detected and reconstructed separately as we showed before. Therefore, such configurations are in general not critical configurations.

### 5.1. No Missing Data—Full Visibility Matrix

The problem of critical configurations for the general case of projective reconstruction has received considerable interest in computer vision over the years

(Maybank, 1992; Hartley and Debunne, 1998; Hartley, 2000). The classical case of 2 view critical configurations implies that all cameras and points are located on a ruled quadric (Krames, 1942). From the duality of camera centers and space points (Carlsson, 1995) follows that this applies also for 6 points and any number of cameras (Hartley and Debunne, 1998). The case of three cameras and an arbitrary number of points and its dual is somewhat more complex (Hartley, 2000). The non-linearity of the general case means that critical configurations generally imply a finite number of multiple solutions given projected image data. Having four points on a reference plane on the other hand, gives us a linear reconstruction problem and therefore either a unique solution or an infinite number. The case of an infinite number of solutions will occur when the $S$-matrix becomes rank deficient so that the dimensionality of the null-space increases.

We will prove that the only critical configurations for 2 points (excluding all points on the reference plane) visible in 2 views are if the camera centers and the points are coplanar. This is not a contradiction to the general case of projective reconstruction, since the placement of points and camera centers is not restricted in the general case. Additionally, we will prove that 2 points and 2 views provide minimal visibility. Furthermore, for the multi view case we will prove that a configuration is non-critical if (a) the points and the camera centers are non-coplanar and (b) all camera centers and one of the points are non-collinear and (c) all points and one of the camera centers are non-collinear.

***5.1.1. Two View Configurations.*** Let us consider the case of 2 points (excluding all points on the reference plane) visible in 2 views. From the projection relation (Eq. (11)) we obtain at the most 8 linearly independent constraints for the $S$-matrix. Note, only 2 of the 3 projection relations are linear independent. If all these 8 equations were linearly independent we would get a unique reconstruction, since the number of dofs is 8.

Therefore, we can make the conjecture that 6 points, where 4 of them are coplanar, are sufficient for a reconstruction from two un-calibrated views. In case of the reference plane at infinity defined by three orthogonal vanishing points we conjecture that 5 points are sufficient under the assumption of a camera model with zero skew and unit aspect ratio. We will now prove that this is indeed the case.

**Theorem 1.** *A configuration of* 2 *points* (*excluding all points on the reference plane*) *visible in* 2 *views is critical if and only if the points and the camera centers are coplanar.*

**Proof:** Since the $S$-matrix has a four dimensional null-space, we are free to choose either a space point or a camera center as the "origin", i.e. (0, 0, 0, 1), of the projective space. The $S$-matrix (Eq. (12)) then takes on either of the forms:

$$\begin{pmatrix} S_{21} & -S_{21} & 0 \\ S_{22} & 0 & -S_{22} \\ 0 & S_{11} & 0 \\ 0 & 0 & S_{12} \end{pmatrix} \begin{pmatrix} \bar{X}_2 \\ \bar{Y}_2 \\ \bar{Z}_2 \\ \bar{A}_1 \\ \bar{B}_1 \\ \bar{C}_1 \\ \bar{A}_2 \\ \bar{B}_2 \\ \bar{C}_2 \end{pmatrix} = 0$$

$$\begin{pmatrix} S_{12} & 0 & -S_{12} \\ 0 & S_{22} & -S_{22} \\ S_{11} & 0 & 0 \\ 0 & S_{21} & 0 \end{pmatrix} \begin{pmatrix} \bar{X}_1 \\ \bar{Y}_1 \\ \bar{Z}_1 \\ \bar{X}_2 \\ \bar{Y}_2 \\ \bar{Z}_2 \\ \bar{A}_2 \\ \bar{B}_2 \\ \bar{C}_2 \end{pmatrix} = 0 \tag{28}$$

where

$$S_{i,j} = \begin{pmatrix} 0 & w_{i,j} & -y_{i,j} \\ -w_{i,j} & 0 & x_{i,j} \\ y_{i,j} & -x_{i,j} & 0 \end{pmatrix} \tag{29}$$

are $3 \times 3$ matrices built up from image coordinates of point $i$ visible in view $j$.

In case of a non-critical configuration these matrices are of rank 8 which means that the null vector is unique up to scale. If the matrices were of rank 7 or less, the dimension of the null-spaces would be larger than one and the null vector no longer unique up to scale. Rank deficiency of a matrix is generally checked by computing the singular values. In our case, however, we are interested in the algebraic conditions on the elements of the matrix for it to be rank deficient. Rank

deficiency, i.e. a rank less than 7, of the $S$-matrix implies that the determinants of all the $8 \times 8$ submatrices of the $S$-matrix are zero.

These subdeterminants were computed using MAPLE and it was found that all subdeterminants that were not genericly zero have a simple common structure. By reordering rows and columns it can be shown that the two cases in Eq. (28) are completely equivalent by the choice of the origin. Therefore, all computations were made for the case of choosing the first camera as the origin, i.e. $\bar{A}_1 = \bar{B}_1 = \bar{C}_1 = 0$. By expressing the elements in the $S_{i,j}$ matrix in terms of coordinates of space points $\bar{P}_1, \bar{P}_2$ and coordinates of the second camera center $\bar{Q}_2$ it was found that all $8 \times 8$ subdeterminants could be factored into:

(A) The determinant:

$$det(\bar{P}_1\ \bar{P}_2\ \bar{Q}_2) \tag{30}$$

(B) A factor computed by selecting one coordinate element from five vectors in three different ways:

$$\begin{aligned} &1.\ (\bar{P}_2 - \bar{Q}_2)\,(\bar{P}_1 - \bar{Q}_2)\ \bar{P}_1\ \bar{P}_2\ \bar{Q}_2 \\ &2.\ (\bar{P}_2 - \bar{Q}_2)\,(\bar{P}_1 - \bar{Q}_2)\ \bar{P}_1\ \bar{P}_2\ \bar{P}_1 \\ &3.\ (\bar{P}_2 - \bar{Q}_2)\,(\bar{P}_1 - \bar{Q}_2)\ \bar{P}_1\ \bar{P}_2\ \bar{P}_2. \end{aligned} \tag{31}$$

This factor is then computed by multiplying these five elements together, e.g.:

$$(\bar{X}_2 - \bar{A}_2)\,(\bar{Y}_1 - \bar{B}_2)\ \bar{X}_1\ \bar{Z}_2\ \bar{A}_1. \tag{32}$$

Rank deficiency of the $S$-matrix, implying that all subdeterminants are zero, will occur if either the A factor or the B factor is zero for all combinatorial choices. Obviously rank deficiency will occur if:

$$det(\bar{P}_1\ \bar{P}_2\ \bar{Q}_2) = 0 \tag{33}$$

which means that points $P_1$, $P_2$ and $Q_2$ are coplanar with the origin, i.e. point $Q_1$.

We will now show that all rank deficient configurations are described by this coplanarity condition. Suppose this condition is not fulfilled, i.e.

$$det(\bar{P}_1\ \bar{P}_2\ \bar{Q}_2) \neq 0 \tag{34}$$

This means that the B factor for every determinant has to be zero. This in turn implies that at least one of the conditions:

$$\bar{P}_2 - \bar{Q}_2 = 0, \quad \bar{P}_1 - \bar{Q}_2 = 0, \quad \bar{P}_1 = 0, \quad \bar{P}_2 = 0 \tag{35}$$

has to be fulfilled. Let us assume that this is not the case. Let us consider the determinants which were constructed as in the second and third way. For such a determinant there is at least one element of each vector which is non-zero. If we select these very elements for the computation of the B factor we obtain a non-zero B factor after multiplying all those elements. Since the A factor was assumed to be non-zero we would obtain a subdeterminant which is non-zero and therefore an $S$-matrix which is not rank deficient. Therefore, at least one of the four conditions in Eq. (35) has to be fulfilled. Since these conditions imply coincidence of points and cameras they all imply coplanarity of the four points $\bar{P}_1, \bar{P}_2, \bar{Q}_2, \bar{Q}_1 = 0$, i.e. $det(\bar{P}_1\ \bar{P}_2\ \bar{Q}_2) = 0$. This concludes the proof that all rank deficient configurations are given by the coplanarity of the two points $\bar{P}_1, \bar{P}_2$ and camera centers $\bar{Q}_1, \bar{Q}_2$. □

We are now able to answer the question of minimal visibility.

**Theorem 2.** *The sufficient and minimal visibility matrix contains* 2 *points* (*excluding all points on the reference plane*) *visible in both* 2 *views.*

**Proof:** 2 points visible in both 2 views is obviously sufficient. All configurations where $\bar{P}_1$, $\bar{P}_2$, $\bar{Q}_1$ and $\bar{Q}_2$ are not coplanar give a unique reconstruction.

Furthermore, we have to prove that this visibility matrix is minimal. Let us assume that not all points are visible in all views. This means that we obtain: $\#equations < 8 = \#dofs$. If we assume that only one view is available, we obtain $\#equations = 2n < 3n - 1 = \#dofs$ for $n > 1$. However, one point visible in one camera can not be reconstructed. The case of one point is dual to the case of one view. This concludes the proof. □

***5.1.2. Multi View Configurations.*** For the case of $n$ points visible in all $m$ views the $S$-matrix has (at the most) $2mn$ linear independent equations and $3(n + m) - 4$ dofs. This means that the $S$ is over-constrained, if it is not rank deficient. Let us investigate the critical configurations for such a case.

**Theorem 3.** *A configuration of n points* (*excluding all points on the reference plane*) *visible in m views is non-critical if* (a) *the points and the camera centers are non-coplanar and* (b) *all camera centers and an arbitrary point are non-collinear and* (c) *all points and an arbitrary camera center are non-collinear.*

**Proof:** We will show that a configuration which does not fulfill the conditions (a), (b) and (c) is a non-critical configuration. This will be done by actually constructing such a unique reconstruction.

First of all we state, that a point and a camera center can never coincide, since such a point would not have a unique projection in such a camera. With the assumption that the condition (a) is not fulfilled we have at least two camera centers and two points which are not coplanar. W.l.o.g we denote the views as $\bar{Q}_1$ and $\bar{Q}_2$ and the points as $\bar{P}_1$ and $\bar{P}_2$. In the previous section we have proved that we obtain a unique reconstruction for such a configuration. We will now show that we can add an arbitrary view $\bar{Q}_i$ to the 2 view system and obtain a 3 view system with a unique reconstruction. Let us assume that the points $\bar{P}_1$, $\bar{P}_2$ and the camera center $\bar{Q}_i$ are not collinear. Figure 3(a) shows the geometric interpretation of such a configuration. Obviously the lines $l_1 = \bar{P}_1 - \bar{Q}_i$ and $l_2 = \bar{P}_2 - \bar{Q}_i$ uniquely define the camera center $\bar{Q}_i$.

In the other case, if $\bar{P}_1$, $\bar{P}_2$ and $\bar{Q}_i$ are collinear, the lines $l_1$ and $l_2$ coincide (see Fig. 3(b)). This means that the camera center $\bar{Q}_i$ has one dof, i.e. has to lie on the line $l_1$. Since we assume that the condition (c) is not fulfilled there is a point $\bar{P}_j$ which does not lie on the line $l_1$. Let us consider the epipolar plane $\Pi_{j1}$, which is defined by $\bar{P}_j$, $\bar{Q}_i$ and $\bar{Q}_1$, and the epipolar plane $\Pi_{j2}$, which is defined by $\bar{P}_j$, $\bar{Q}_i$ and $\bar{Q}_2$. The intersection of the epipolar plane $\Pi_{j1}$ and the line $l_1$ defines the camera center $\bar{Q}_i$ uniquely if $l_1$ and $\Pi_{j1}$ do not coincide. The

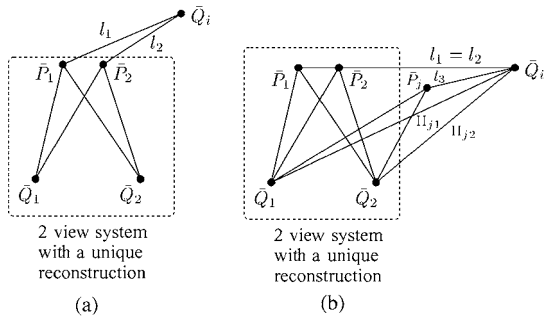



*Figure 3.* Geometric interpretations for the proof of Theorem 3.

same applies to the epipolar plane $\Pi_{j2}$. We will now show that either of these two cases is true. Let us assume that the two planes $\Pi_{j1}$ and $\Pi_{j2}$ are different. This implies that the two planes intersect uniquely in the line $l_3 = \bar{P}_j - \bar{Q}_i$. Since $\bar{P}_j$ does not lie on $l_1$, the two lines $l_1$ and $l_3$ are different. Therefore, either the plane $\Pi_{j1}$ or the plane $\Pi_{j2}$ does specify the camera center $\bar{Q}_i$ uniquely. We are left with the case that $\Pi_{j1}$ and $\Pi_{j2}$ are identical. This implies that the plane $\Pi_{j1}$ contains the camera centers $\bar{Q}_1$ and $\bar{Q}_2$. However, if $l_1$ coincided with the plane $\Pi_{j1}$, the condition (a) would be violated, i.e. $\bar{P}_1$, $\bar{P}_2$, $\bar{Q}_1$ and $\bar{Q}_2$ would be coplanar. Therefore, $l_1$ can not coincide with $\Pi_{j1}$ and the camera center $\bar{Q}_i$ is uniquely defined by $l_1$ and $\Pi_{j1}$.

Furthermore, if $\bar{P}_j$ lay on the baseline between $\bar{Q}_1$ and $\bar{Q}_i$ or on the baseline between $\bar{Q}_2$ and $\bar{Q}_i$, the point $\bar{P}_j$ would specify this very baseline which means that the camera center $\bar{Q}_i$ is uniquely defined as well.

In this way all the views can be added to the 2 view system. Therefore, we obtain a unique reconstruction with $m$ views and the two points $\bar{P}_1$ and $\bar{P}_2$.

With the assumption that the condition (b) is not fulfilled we can finally reconstruct all points. This means that for every configuration which does not satisfy the conditions (a), (b) and (c) we obtain a unique reconstruction. This concludes the proof. □

Let us consider which of the configurations (a–c) are actually critical. A configurations of type (b) where all camera centers and one of the points are collinear is obviously critical. Such a point lies on the baselines of all pairs of cameras and can not be reconstructed. Therefore, configurations of type (c) are critical as well, since they are dual to the configurations of type (b). However, a configuration of type (a) where all camera centers and points are coplanar is not necessarily critical. Note, the fact that all pairs of possible 2-view cases are critical (as proved in Theorem 1) does not imply that the configuration is critical. The investigation of configurations of type (a) for $n$ points and $m$ views would be of theoretical interest.

Let us consider the question of sufficient visibility for $n$ points and $m$ views.

**Theorem 4.** *Every visibility matrix which contains* 2 *or more points* (*excluding all points on the reference plane*) *and* 2 *or more views is sufficient if all points are visible in all views.*

**Proof:** We choose a configuration which does not fulfill the conditions (a), (b) or (c). Obviously, this can be

done for an arbitrary (more than 2) amount of views and points. Such a configuration has a unique reconstruction as proved in Theorem 3. □

With Theorems 2 and 4 we can conclude that the basic condition that $\#equations \geq \#dofs$ is a sufficient check for sufficient visibility in the case of no missing data. With a full visibility matrix we obtain: $\#equations = 2mn$ and $\#dofs = 3(m+n) - 4$.

### 5.2. *Missing Data—Not Full Visibility Matrix*

Compared to the previous case of no missing data, the problem of minimal visibility and critical configurations for the general case of multi view projective reconstruction with missing data has received less attention in the past. In Quan and Heyden (1999) all reconstructions for sufficient visibility matrices with 3 and 4 images are cataloged.

We will now address the problem of minimal visibility and critical configurations for the case of missing data and with the assumption of having four coplanar points. Furthermore, we will introduce a constructive method of choosing points and cameras which provide sufficient visibility and non-critical configurations.

***5.2.1. Minimal Visibility and Critical Configurations.*** Let us first consider the question of minimal visibility in the case of missing data. The basic condition that $\#equations \geq \#dofs$ is insufficient to answer the question of sufficient visibility. For a non-full visibility matrix we obtain for the maximum number of linearly independent equations: $\#equations = 2\#(V(i, j) = set)$ and for the number of dofs: $\#dofs = 3(m+n) - 4$ (the reference points excluded). However, if these equations include linear dependences, the number of linearly independent equations reduces. In order to give a complete answer for a given visibility matrix, the rank (or the subdeterminants) of the corresponding $S$-matrix has to be investigated for a generic set of points and cameras. Such an investigation can be carried out with MAPLE.

Let us consider the specific visibility matrix in (27). Although the number of equations is equal the number of dofs, i.e. $\#equations = 20 = \#dofs$, the corresponding $S$-matrix has rank 19, i.e. is rank deficient, for a generic set of points. In this case the linear dependence of equations can be seen if we consider the views 1 and 2 and the views 2 and 3 as separated 2-view cases. The second 2-view case includes a linear dependence since $\#equations = 12 > 11 = \#dofs$. Excluding e.g. point 5 results in linear independent equations for the second 2-view case, since $\#equations = 8 = \#dofs$. However, in this case the resulting $S$-matrix for the 3-view case is under-constrained since $\#equations = 16 < 17 = \#dofs$.

The general problem of critical configurations in the case of missing data is very complex. Basically ever specific visibility matrix might give a different set of critical configurations. Therefore, the rank (or the subdeterminants) of the $S$-matrix for a specific configuration has to be investigated in the same manner as we did in the 2-view case with no missing data.

***5.2.2. A Constructive Method.*** So far we have considered the questions of sufficient visibility and critical configurations for a given visibility matrix. However, in practice the placement of cameras and the number of visible points can be chosen freely to a certain extent. Therefore, it is of particular interest of having a method of choosing points and cameras which provide sufficient visibility and non-critical configurations.

We will now introduce and prove such a method for the multi view case. This will be done in an iterative way in terms of the number of cameras. Let us assume that we have a unique reconstruction of $n$ points and $m$ views and we want to add a new view $\bar{Q}_{m+1}$ to this $m$ view system. In order to obtain a unique reconstruction with the additional view, we have to specify the 3 dofs of the new camera center $\bar{Q}_{m+1}$. There are various ways of doing this. Let us assume that a point $\bar{P}_i$ which is already reconstructed is visible in the view $\bar{Q}_{m+1}$. Furthermore, a new point $\bar{P}_{n+1}$ is visible in $\bar{Q}_{m+1}$ and in $\bar{Q}_j$, which belongs to the $m$ view system. Figure 4 shows the geometric interpretation of such a case. The point $\bar{P}_i$ gives at least 2 more constraints. The point $\bar{P}_{n+1}$ adds 3 dofs to the new $m+1$ view system,

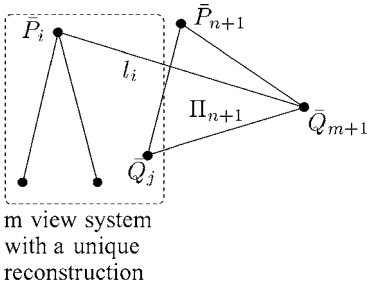



*Figure 4.* Adding a new view $\bar{Q}_{m+1}$ to a system with $n$ points and $m$ views which has a unique reconstruction.

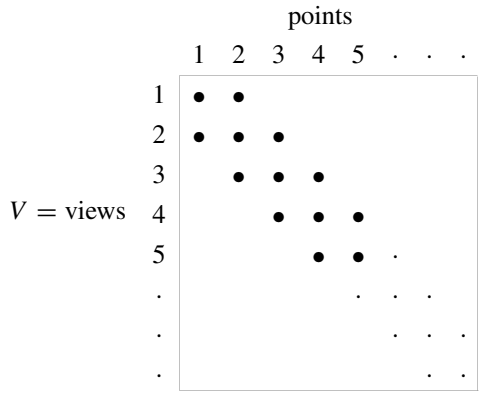
however, it supplies 4 more constraints on the system as well. This is sufficient for specifying the 3 dofs of the new camera center $\bar{Q}_{m+1}$. Therefore, such a visibility is sufficient for obtaining a unique reconstruction for the $m+1$ views case.

The remaining question is: which are the critical configurations of such a multi view system?

Figure 4 shows the geometric relationship between points and cameras. The point $\bar{P}_i$ introduces a line $l_i = \bar{P}_i - \bar{Q}_{m+1}$, where the camera center $\bar{Q}_{m+1}$ has to lie on. Furthermore, the point $\bar{P}_{n+1}$ introduces the epipolar plane $\Pi_{n+1}$ which contains the camera centers $\bar{Q}_j$ and $\bar{Q}_{m+1}$. Let us assume that the point $\bar{P}_{n+1}$ does not lie on the baseline between $\bar{Q}_j$ and $\bar{Q}_{m+1}$. In this case, the camera center is uniquely defined if $l_i$ does not coincide with the plane $\Pi_{n+1}$. This is true if $\bar{P}_i$, $\bar{P}_{n+1}$, $\bar{Q}_j$ and $\bar{Q}_{m+1}$ are not coplanar. We are left with the case that $\bar{P}_{n+1}$, $\bar{Q}_j$ and $\bar{Q}_{m+1}$ are collinear. The point $\bar{P}_{n+1}$, which is assumed to be visible only in $\bar{Q}_j$ and $\bar{Q}_{m+1}$, cannot be reconstructed, since it lies on the baseline between $\bar{Q}_j$ and $\bar{Q}_{m+1}$. Let us summarize: *a configuration of such a multi view system is critical if and only if (a) $\bar{P}_i$, $\bar{P}_{n+1}$, $\bar{Q}_j$ and $\bar{Q}_{m+1}$ are coplanar or (b) $\bar{Q}_j$, $\bar{P}_{n+1}$ and $\bar{Q}_{m+1}$ are collinear.*

Obviously, adding more points to this system does not affect a non-critical configuration as long as the following condition is satisfied, such a point is not collinear with those camera centers from which the point is visible.

A possible visibility matrix for such a multi view system is:




Such a band-structured matrix typically appears for reconstructing large scale scenes, e.g. architectural environments, as we will see in the experimental section. It reflects the fact that model points appear and disappear in the sight of view while the camera moves around an object, e.g. a building.

## 6. Experiments

### 6.1. Synthetic Data

The synthetic experiments were conducted for the case of a finite reference plane. However, some of the conclusions drawn from the synthetic experiments can also be applied to infinite reference planes since the algorithm for infinite reference planes is part of the algorithm for finite reference planes (see Section 4.2). In order to investigate the performance of our algorithm, it was applied to two different synthetic configurations (see Fig. 5). The synthetic scene consists of a cube with 26 points floating above a reference plane. The reference plane is a square where the four corners depict the reference points. In the first configuration (Fig. 5(a)) a camera circled around the cube with a radius of 10 units and shot 8 images (CIR-configuration). In the second configuration (Fig. 5(b)) the camera moved translationally towards the scene (TRA-configuration). The dimensions of the configurations are as in Fig. 5. The internal calibration matrix of the camera was set to diag(1000,1000,1).

In a first experiment, the influence of noise on the image data was investigated. Therefore different levels of Gaussian noise: $\sigma = 0, 0.2, \ldots, 3.0$ (standard deviation) were added to the image data, i.e. reprojected 3D points. Additionally, our algorithm was applied to a special situation: Gaussian noise was added to all image points *except for the reference points* (denoted in the following as perfect basis). The computed reconstructions were evaluated in terms of the Root-Mean-Square (RMS) error between reprojected 3D points and 2D image data (potentially corrupted by noise). Figure 6(a) shows the results for the CIR-configuration and Fig. 6(b) for the TRA-configuration. Additionally, the performance of the projective factorization algorithm of Sturm-Triggs (Sturm and Triggs, 1996) is shown, which assumes that all points are visible in all views. The "projective depths" used in this method were initialised to one and reestimated by reprojection. This is a simplification of the original approach by Sturm-Triggs, however, it has been demonstrated by Heyden et al. (1999), Hartley and Zisserman, (2000), and Hartley et al. (2001) to produce good results as well. The first observation is that the different algorithms performed approximately the same for both

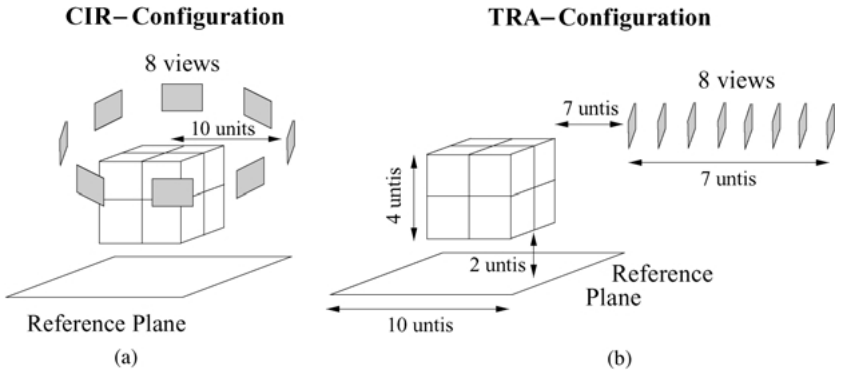



*Figure 5*. Two synthetic configurations with circular motion (radius 10 units) of the camera (a) and translational movement of the camera towards the scene (b).

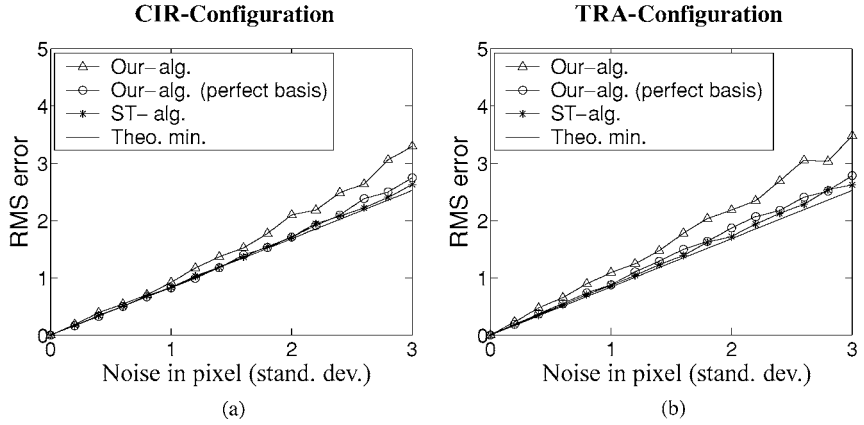



*Figure 6*. Performance of our algorithm (Our-alg.) and Sturm-Triggs algorithm (ST-alg.) with respect to noise on the image points for the CIR-configuration (a) and the TRA-configuration (b).

configurations. If the reference points were not corrupted by noise (perfect basis), our method and projective factorization performed nearly identical. The performance of both algorithms are close to the theoretical minimum, i.e. Cramer-Rao lower bound. If noise was added on the reference points, the performance of our algorithm is slightly worse. This leads to the conclusion that, independent of the configuration, the *noise on the reference points is crucial* for the performance of our algorithm. Further experiments, including the case of an infinite reference plane, confirmed this conclusion.

The second experiment investigated the problem of separating points on (or close to) and off a finite reference plane. Therefore the distance between the cube and the reference plane was varied between 0 and 2 units (see Fig. 5(b)). If the distance is 0, 9 out of 26 points of the cube lied on the reference plane. Two different variations of our algorithm were utilized: always all points are used for the $S$-matrix (*without threshold*) and points are iteratively excluded form the $S$-matrix (*with threshold*). The iterative exclusion of reference points means that the threshold for separating points is automatically detected (see Section 4.2). Figure 7(a) shows the performance of the algorithms in terms of RMS-error for the CIR-configuration. The performance is very similar above a certain distance, i.e. about 0.5 units. However, if the cube is closer to the reference plane, the performance of the algorithm "without threshold" is worse and eventually fails. The algorithm which reconstruct points close to the reference plane separately, i.e. "with threshold", has a constant performance for all distances. The ratio between the fifth and fourth last singular value is depicted in Fig. 7(b). The curves are as expected. The solution gets less stable if the cube moves closer to the reference plane. If the cube is closer than 0.5 units to the reference

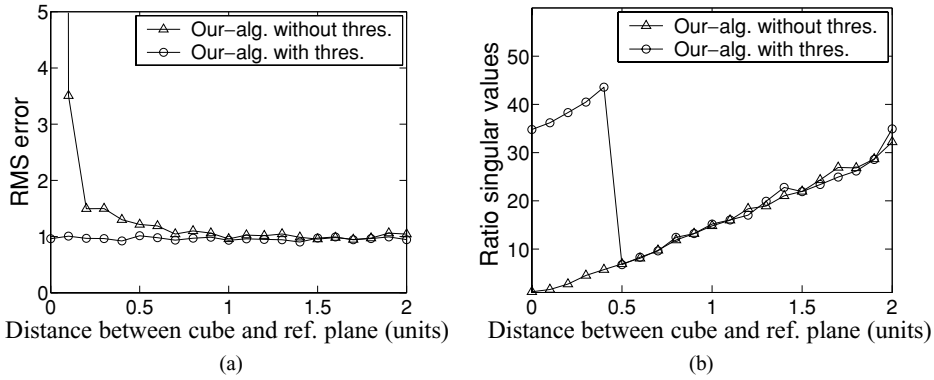



*Figure 7.* The performance of two variations of our algorithm: always all points are used for the $S$-matrix (*without threshold*) and points are iteratively excluded form the $S$-matrix (*with threshold*). The performance is analysed in terms of RMS-error (a) and the ratio between the fifth and fourth last singular value (b).

plane the algorithm "with threshold" performed much more stable than the one "without threshold". This is due to the fact that in this case 9 out of 26 points of the cube were reconstructed separately. We may draw the conclusion that the problem of separating points on (or close to) and off the reference plane can be handled by our algorithm. However, an algorithm which does not take care of this problem eventually fails if points are on or close to the reference plane.

### 6.2. *Real Data—Finite Reference Plane*

In a first experiment a tape holder was reconstructed. Four images of the tape holder were taken from viewpoints with considerably wide mutual baselines (see Fig. 8(a)–(c)). 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 24 model points was achieved in one SVD. The 6 model points which lie on or close to the reference plane were automatically detected and separately reconstructed. In order to visualize the result we assumed knowledge of five Euclidean coordinates to rectify the projective structure. Figure 8(d)–(f) shows the results of the reconstruction from three different views. We see that the reconstruction matches with the approximate size of the object which is 6.0 cm ($x$-direction), 15.8 cm ($y$-direction) and 6.8 cm ($z$-direction). Furthermore, the symmetry of the object is maintained in the reconstruction. Since the ratio between the fifth last singular value (0.766) and the fourth last singular value (0.031) is substantially high, i.e. 24.7, this configuration can be considered as non-critical. By manually selecting points which lie on same model planes we created a VRML model, which consists solely of planes. Figure 8 (g)–(i) depicts 3 novel views of the VRML model.

In a second experiment we reconstructed a teapot. In order to achieve this, the teapot was posed on a box (see Fig. 9(a)–(c)). With the aid of the box both methods, with finite and infinite reference plane, can be applied. The four corner points of the box, which are marked with circles, specify the finite reference plane. The mutual orthogonal edges of the box were used to determine $K$ and $R$, i.e. the plane at infinity. For a better visualization, only those model points were reconstructed which lie on the contour in the top, side or front view of the model. Figure 9 shows the reconstruction of 99 model points, which were determined with the finite (d)–(f) and infinite (g)–(i) reference plane approach. The reconstructed model points include the corner points of the box and a cuboid, which was placed beside the teapot. The Euclidean coordinates of the cuboid were used to rectify the projective reconstruction, which we obtained with the finite reference plane approach. Let us consider the metric reconstruction which was derived with the infinite reference plane approach. The average error between selected image points and back-projected model points was 0.65 pixel, whereas the size of the image is $1600 \times 1200$ pixel. The respective maximum error was 5.2 pixel. The ratio between the fifth last singular value (360.36) and the fourth last singular value (4.65) was 77.5. For the approach with a finite reference plane, the ratio between the fifth last singular value (0.0545) and the fourth last singular value (0.0014) was 38.9. Let us compare both reconstructions

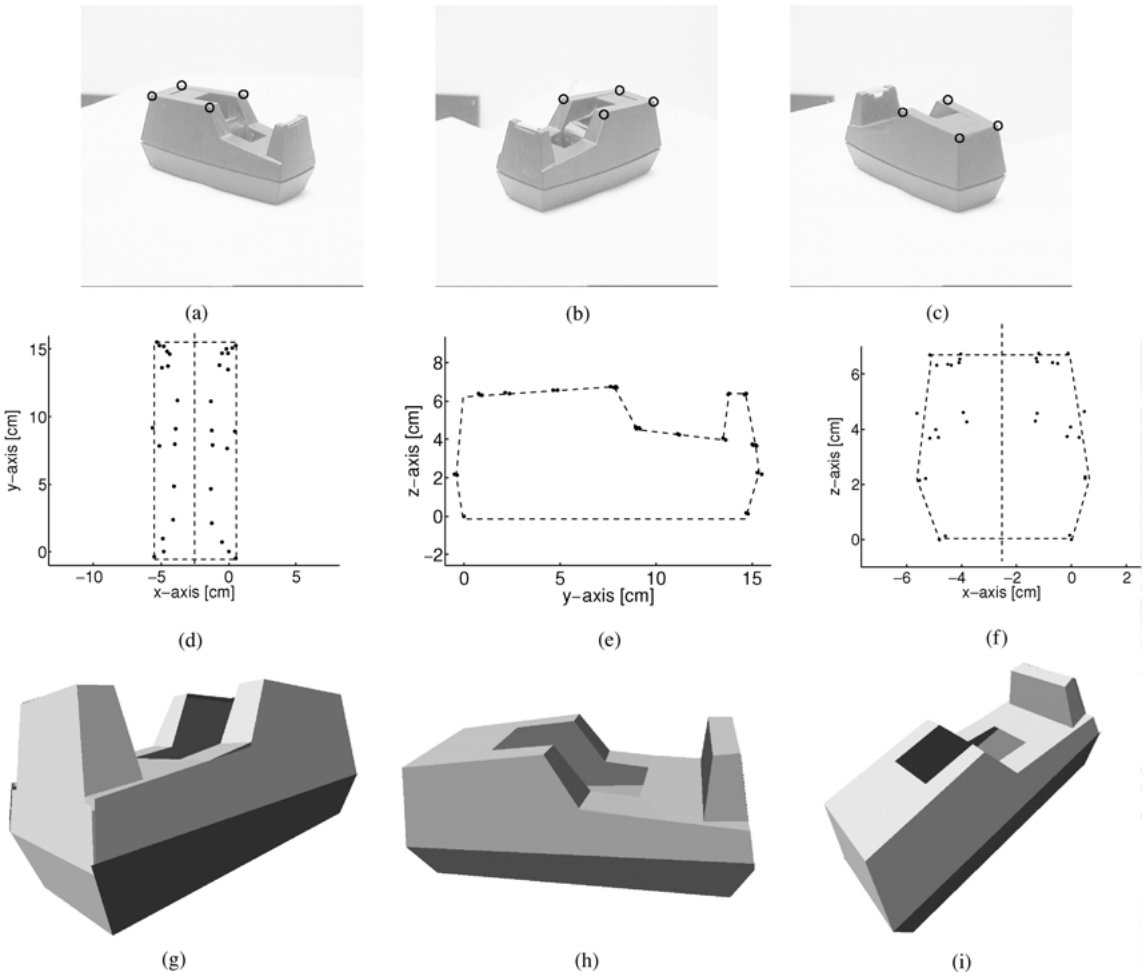



*Figure 8.* Three original views of the tape holder (a)–(c). The top (d), side (e) and front (f ) view of the reconstruction, where the dots represent the reconstructed model points and the dashed lines display the contour and the symmetry axis of the model. Three novel views of the tape holder (g)–(i).

of the teapot, which has the approximate dimensions of 14.5 cm ($x$-direction), 19.7 cm ($y$-direction) and 15.9 cm ($z$-direction). The average difference is 0.76 cm and the maximal difference is 1.31 cm.

### *6.3. Real Data—Infinite Reference Plane*

In a first experiment of a large scale environment we reconstructed three buildings of the campus of the Royal Institute of Technology in Stockholm/Sweden. 26 images of size $1600 \times 1200$ pixel were taken with a handheld camera (Olympus 3030) (see Fig. 10(a) and (b)). The internal camera parameters remained fix while the pictures were taken. In order to establish a correspondence between the three buildings, we utilized additionally a postcard of the campus (see Fig. 10(c)). Naturally, we had no calibration information, e.g. the focal length, of the postcard available. For this application the plane at infinity was used as reference plane. Therefore, we manually selected mutual orthogonal edges in each image, which were used to determine $K$ and $R$ for each view. The camera's calibration was improved by assuming fixed internal camera parameters. In case of the postcard, one of the vanishing points is close to infinity (horizontal lines). However, the focal length can still be determined for this degenerate configuration with the additional assumption that the principal point is close to the middle of the image. Furthermore, the correspondences of 114 model points were manually achieved. On the basis of this the campus was reconstructed in one single SVD. Figure 11 shows the

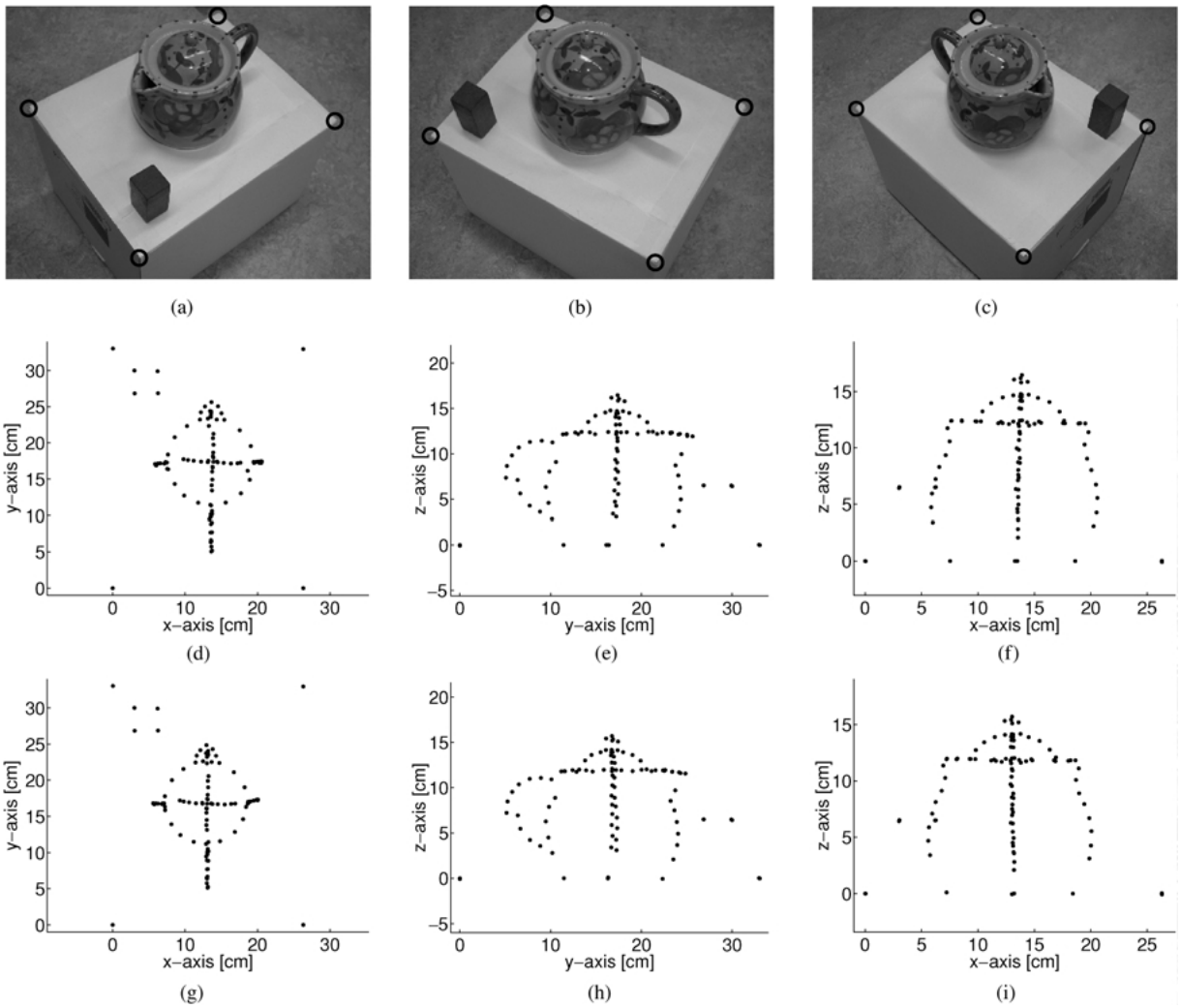



*Figure 9.* Three original views of the teapot (a)–(c). The top, side and front views of the reconstruction with a finite (d)–(f) and infinite (g)–(i) reference plane.

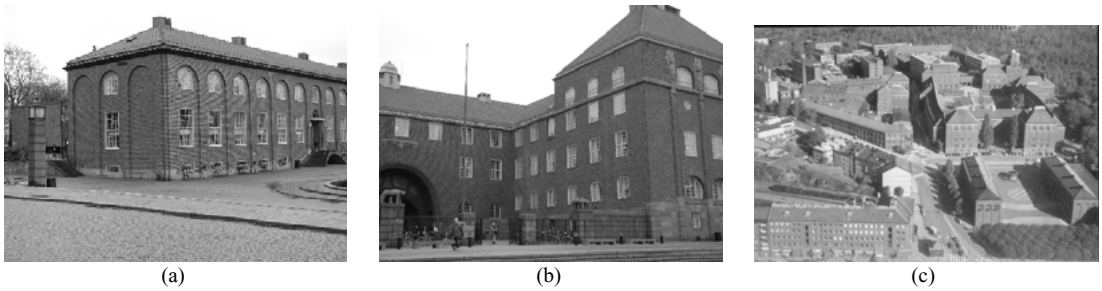



*Figure 10.* Two original views (a) and (b) and a postcard (c) of the campus. The corresponding camera positions are labeled in the top view (Fig. 11).

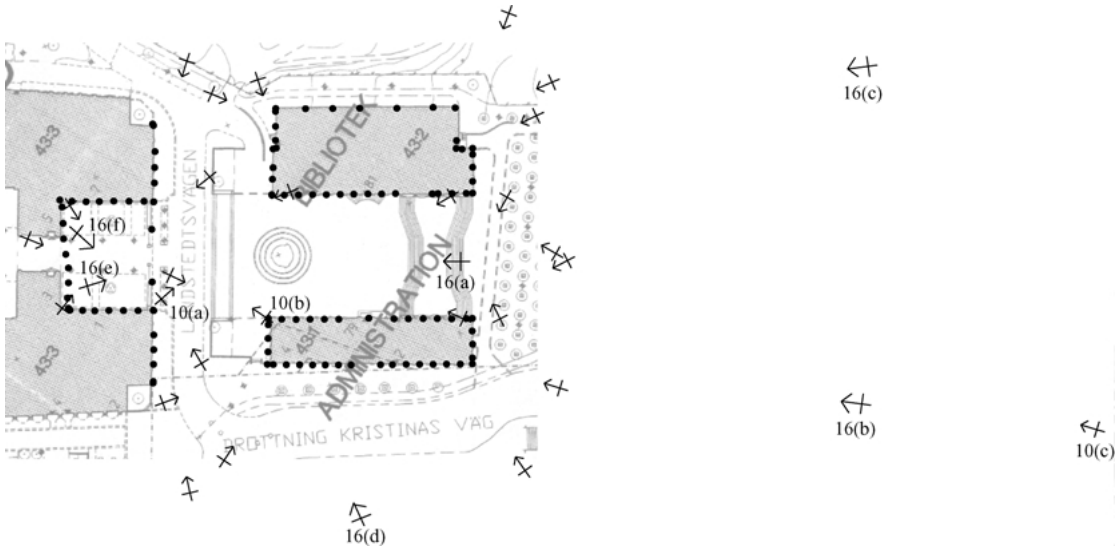



*Figure 11*. Top view of the reconstruction of the campus with 114 model points (dots) and 27 cameras (arrows). A map of the campus is superimposed. The labeled cameras correspond to images in the respective figures.

top view of the reconstruction, whereas the dots represent reconstructed points, arrows depict cameras and the grey structure represents the superimposed map of the campus. The labeled cameras correspond to images in the respective figures. We stress that no further constraints, e.g. orthogonality, were imposed, which would presumably improve the reconstruction. As in the previous experiment, we obtain a VRML model by manually selecting model points which lie on same model planes. By projecting image texture onto the planes we acquire the final VRML model of the campus. Figure 12 shows 6 novel views of the VRML model.

Let us consider the results. The fourth and the fifth last singular values 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 pixel. The respective maximum error was 35.2 pixel, 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. 11) demonstrates the high quality of the reconstruction.

The visibility matrix in Fig. 13 shows the 114 points partly visible in 27 images. We see that the matrix is only sparsely filled, i.e. 10.4% of the entries are set. The first 24 images can be divided into three groups of 8 images, whereas each group represents images of a certain building. Therefore, most of the model points are visible exclusively in one group. Those model points

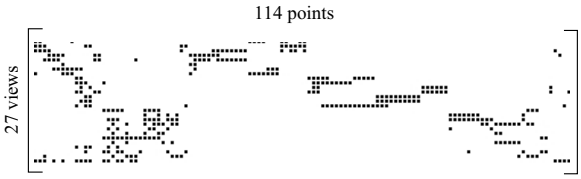



*Figure 13*. The visibility matrix of the campus with 27 images and 114 model points. If the $j$th point is visible in the $i$th view, the corresponding element $V(i, j)$ is set (a black square).

which are visible in more than one group, have to be visible in images which display more than one building, e.g. Fig. 10(c).

In a second experiment we reconstructed the outside and inside (courtyard) of the City Hall in Stockholm/Sweden. 35 images of size $1600 \times 1200$ pixel were taken, whereas the internal camera parameters remained fix (see Fig. 14). As in the previous experiment, the plane at infinity was used as the reference plane. Since some parts of the building can be seen from both the outside and inside, e.g. the tower (see Fig. 14(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. 6 novel views of the textured VRML model of the building are displayed in Fig. 15. Since part of the roof can not be seen from a ground plane position, the roof was not reconstructed.

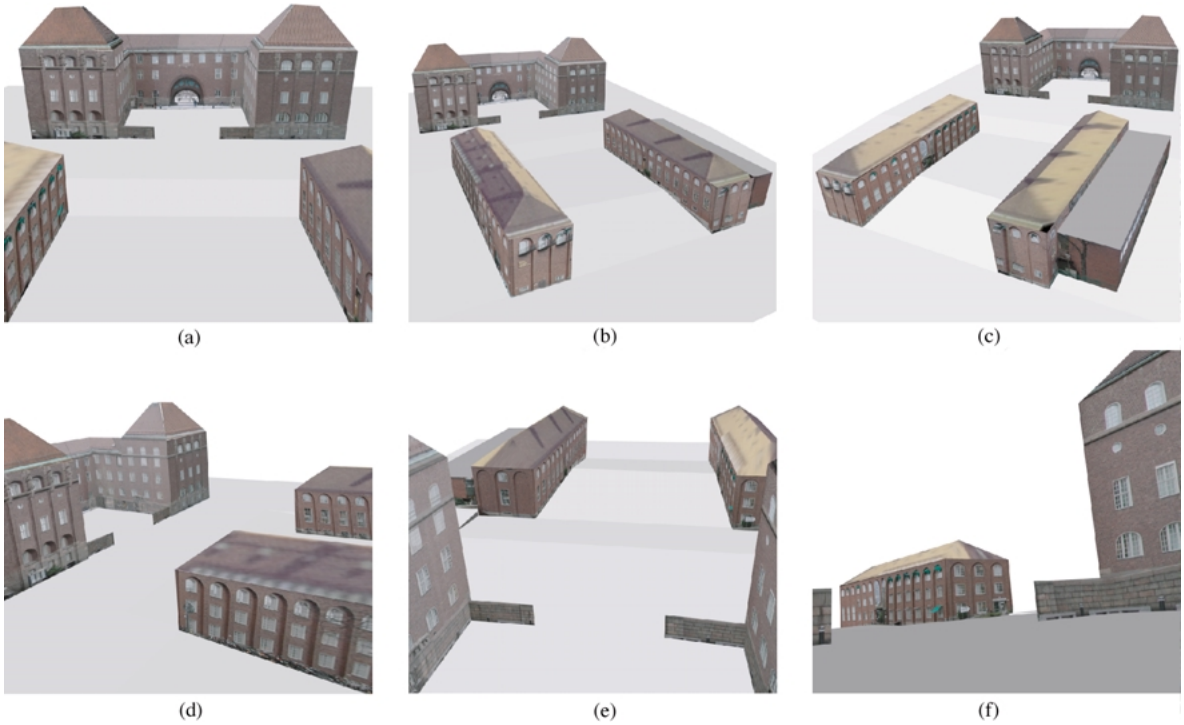



*Figure 12*. Six novel views of the campus. The corresponding camera positions are labeled in the top view (Fig. 11).

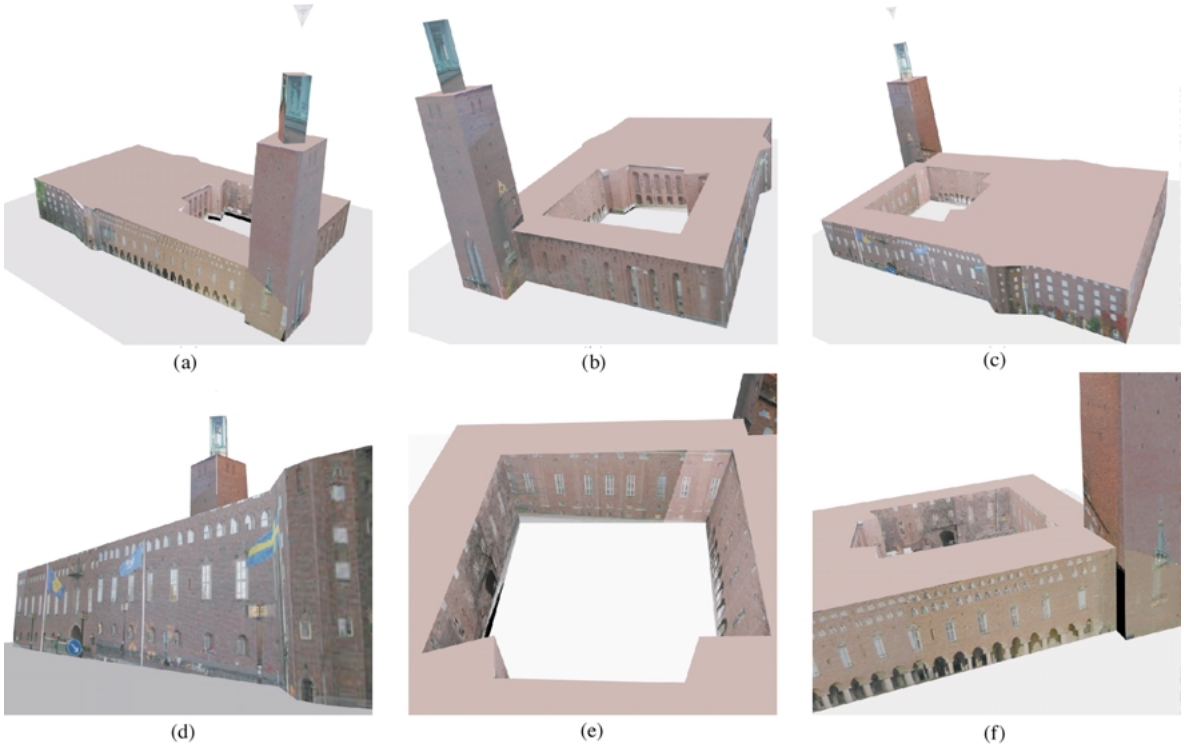



*Figure 15*. Six novel views of the City Hall. The corresponding camera positions are labeled in the top view (Fig. 16).

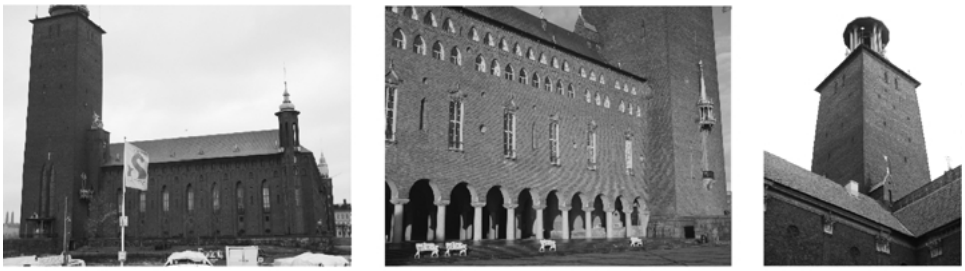



*Figure 14.* Three original views of the City Hall. The corresponding camera positions are labeled in the top view (Fig. 16).

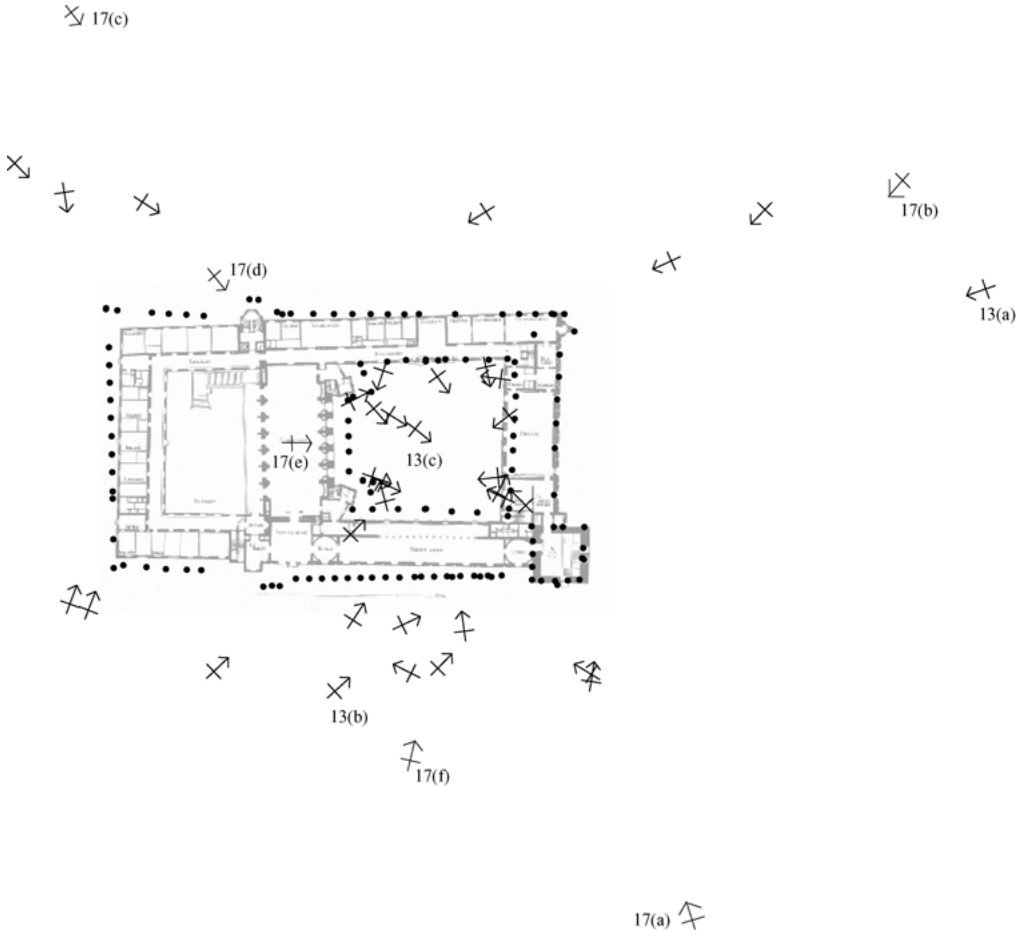



*Figure 16.* Top view of the reconstruction of the City Hall with 129 model points (dots) and 35 cameras (arrows). A map of the City Hall is superimposed. The labeled cameras correspond to images in the respective figures.

The top view of the reconstruction with a superimposed map of the City Hall is shown in Fig. 16. As in the previous example, no further constraints were imposed in order to improve the reconstruction.

The ratio between the fifth last singular value (57.24) and the fourth last singular value (12.75) was 4.49. The average error between selected image points and back-projected model points was 0.81 pixel. The respective

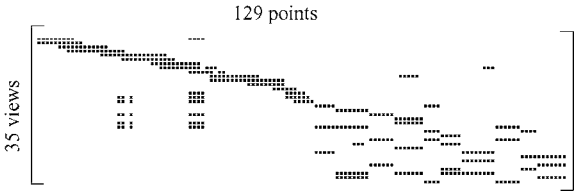



*Figure 17.* The visibility matrix of the City Hall with 35 images and 129 model points. If the $j$th point is visible in the $i$th view, the corresponding element $V(i, j)$ is set (a black square).

maximum error was 97.31 pixel, which is 4.9% of the image diagonal. Let us consider the quality of the reconstruction (see Fig. 16). It stands out that the building was not designed as a perfect rectangular building. However, this fact does not considerably affect the good 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. 17 depicts the 129 model points partly visible in the 35 images. As in the previous experiment, the matrix is only sparsely filled, i.e. 9.7% of the entries are set. 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 in the sight of view while the camera moves around the outside of the building. The lower half of the matrix which represents images of the inside of the building seems less structured. This is due to the fact that the strategy of taking pictures is more complex in this case. The strategy was to maximize both the baseline of the cameras and of the model points (see Fig. 16). The remaining correspondences which do not belong to one of the regions discussed above represent model points which are visible from the outside and the inside of the building, e.g. part of the tower (see Fig. 14(a)–(c)).

## 7. Summary and Conclusions

We have demonstrated theoretically and experimentally that points and camera centers in a multi view situation can be simultaneously, 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 (or a reference plane) 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 (Fitzgibbon and Zisserman, 1998). Contrary to factorization algorithms for affine reconstruction (Tomasi and Kanade, 1992) or projective reconstruction (Sturm and Triggs, 1996), we do not need to have all points visible in all views. This gives a very flexible algorithm for the reconstruction of e.g. large scale scenes such as architectural environments where the reference plane can be chosen as the plane at infinity using vanishing points as the reference points.

Most of the planar parallax approaches utilized a reference plane to either reconstruct points in two views (Kumar et al., 1994; Irani and Anandan, 1996; Criminisi et al., 1998) or multiple views (Irani et al., 1998) or to recover the camera's motion (Hartley et al., 2001). In contrast to this, the approach by Triggs (2000) and our method reconstruct structure and motion simultaneously. Both methods utilize the assumption of having four coplanar points (or a reference plane) to formalize the projective reconstruction problem in terms of purely translating calibrated cameras. In contrast to our approach, Triggs method needs all points to be visible in all views. This is a major restriction for practical applicability. However, this assumption makes it possible to determine points and cameras from a rank-1 factorization of a matrix containing all image points and projective depths. Furthermore, these projective depths have to be determined in a pre-processing step which implies that multi-image tensors, e.g. fundamental matrices, have to be additionally known. The main advantage of Triggs method is that all points, which includes points on and close to the reference plane, are used in the factorization step. In case of a finite reference plane those points, which are on or close to the reference plane, have to be reconstructed separately with our approach. This is suboptimal, however, it has been shown in experiments on synthetic and real data that this problem is not critical. The size of the matrix, which contains all image data, is in case of full visibility ($n$ points and $m$ views) $3mn \times 3(m + n)$ for our method and $3m \times n$ for Triggs method. Since both methods apply a SVD on this matrix, our method is eventually slower. Furthermore, we have seen that the

price we have to pay in order to obtain a closed-form solution for all points and cameras is that an algebraic error function is minimized. However, we have demonstrated experimentally that despite this suboptimal cost function the performance of our algorithm is nearly optimal, i.e. very close to the theoretical minimum.

Experimental results on real data indicate that the use of 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 reconstructions can potentially be further improved by applying the non-linear, iterative bundle adjustment method (Triggs et al., 1999).

The linearity and specific symmetry relation between points and camera centers implies that any analysis of critical configurations, numerical stability etc. is greatly simplified. The questions of critical configurations was discussed under the assumption of having four points on a reference plane. We have proved that if all points are visible in all views, i.e. no missing data, all configuration (apart from trivial ones) where points and camera centers are non-coplanar are non-critical. If not all points are visible in all views, i.e. missing data, a method to construct non-critical configurations was proposed.

## Notes

1. This issue will be reconsidered in the next section.
2. This can be seen by the mapping $(0, 1)^T \rightarrow (1, 0)^T$ and $(1, 1)^T \rightarrow (1, 1)^T$ in the projective space $P^1$.

## References


Caprile, B. and Torre, V. 1990. Using vanishing points for camera calibration. *Int. J. Computer Vision*, 4:127–139.

Carlsson, S. 1995. Duality of reconstruction and positioning from projective views. In *IEEE Workshop on Representation of Visual Scenes*, P. Anandan (Ed.), Boston, USA.

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

Criminisi, A., Reid, I., and Zisserman, A. 1998. Duality, rigidity and planar parallax. In *European Conf. Computer Vision*, Springer-Verlag, Freiburg, Germany, pp. 846–861.

Cross, G., Fitzgibbon, A.W., and Zisserman, A. 1999. Parallax geometry of smooth surfaces in multiple views. In *Int. Conf. Computer Vision*, Kerkyra, Greece, pp. 323–329.

Faugeras, O.D. 1992. What can be seen in three dimensions with an uncalibrated stereo rig? In *European Conf. Computer Vision*, D. Sandini (Ed.), Santa Margherita Ligure, Italy, Springer-Verlag, pp. 563–578.

Faugeras, O.D. and Luong, Q.-T. 2001. *The Geometry of Multiple Images*. The MIT Press: Cambridge, MA.

Fitzgibbon, A.W. and Zisserman, A. 1998. Automatic camera recovery for closed or open image sequences. In *European Conf. Computer Vision*, Freiburg, Germany, pp. 311–326.

Hartley, R. 1997. In defence of the 8-point algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(6):580–593.

Hartley, R. 2000. Ambiguous configurations for 3-view projective reconstruction. In *European Conf. Computer Vision*, Dublin, Ireland, pp. 922–935.

Hartley, R. and Debunne, G. 1998. Dualizing scene reconstruction algorithms. In *Workshop on 3D Structure from Multiple Images of Large-scale Environments (SMILE-98)*, R. Koch and L. Van Gool, (Eds.), LNCS, vol. 1506, Springer-Verlag: Berlin, pp. 14–31.

Hartley, R. and Zisserman, A. 2000. *Multiple View Geometry in Computer Vision*. Cambridge University Press: Cambridge.

Hartley, R., Dano, N., and Kaucic, R. 2001. Plane-based projective reconstruction. In *Int. Conf. Computer Vision*, Vancouver, Canada, pp. 420–427.

Heyden, A. 1998. Reduced multilinear constraints—theory and experiments. *Int. J. Computer Vision*, 30(1):5–26.

Heyden, A. and Åström, K. 1995. A canonical framework for sequences of images. In *IEEE Workshop on Representation of Visual Scenes*, P. Anandan (Ed.), Boston, USA.

Heyden, A. and Åström, K. 1997. Simplifications of multilinear forms for sequences of images. *Image and Vision Computing*, 15(10):749–757.

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):981–991.

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

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*, Freiburg, Germany, Springer-Verlag, pp. 829–845.

Jacobs, D. 1997. Linear fitting with missing data for structure-from-motion. In *IEEE Conf. Computer Vision and Pattern Recognition*, San Juan, Puerto Rico, pp. 206–212.

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

Krames, J. 1942. Über die bei der Hauptaufgabe der Luftphotogrammetrie auftretenden "gefährlichen" Flächen. Bildmessung und Luftbildwesen, 17, Heft 1/2:1–18.

Kumar, R., Anandan, P., and Hanna, K. 1994. Direct recovery of shape from multiple views: A parallax based approach. In *Int. Conf. Pattern Recognition*, Jerusalem, Israel, pp. 685–688.

Liebowitz, D. and Zisserman, A. 1999. Combining scene and auto-calibration constraints. In *Int. Conf. Computer Vision*, Kerkyra, Greece, pp. 293–300.

Maybank, S.J. 1992. *Theory of Reconstruction from Image Motion*. Springer-Verlag: Berlin.

Oliensis, J. 1995. Multiframe structure from motion in perspective. In *Workshop on Representations of Visual Scenes*, Boston, USA, pp. 77–84.

Oliensis, J. 1999. A multi-frame structure-from-motion algorithm under perspective projection. *Int. J. Computer Vision*, 34(2/3):163–192.

Oliensis J. and Genc, Y. 1999. Fast algorithms for projective multi-frame structure from motion. In *Int. Conf. Comp. Vision*, Kerkyra, Greece, pp. 536–542.

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, pp. 55–61.

Quan, L. 1994. Invariants of 6 points from 3 uncalibrated images. In *European Conf. Computer Vision*, Stockholm, Sweden, Springer-Verlag, pp. 459–470.

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, pp. 210–216.

Rother, C. 2000. A new approach for vanishing point detection in architectural environments. In *11th British Machine Vision Conference*, Bristol, UK, pp. 382–391.

Rother, C. and Carlsson, S. 2001. Linear multi view reconstruction and camera recovery. In *Int. Conf. Computer Vision*, Vancouver, Canada, pp. 42–51.

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*, Dublin, Ireland, Springer-Verlag, pp. 632–648.

Sparr, G. 1996. Simultaneous reconstruction of scene structure and camera locations from uncalibrated image sequences. In *IEEE Conf. Computer Vision and Pattern Recognition*, Vienna, Austria, pp. 328–333.

Sturm, P. and Triggs, B. 1996. A factorization based algorithm for multi-image projective structure and motion. In *European Conf. Computer Vision*, Cambridge, UK, Springer-Verlag, pp. 709–719.

Svedberg, D. and Carlsson, S. 1999. Calibration, pose and novel views from single images of constrained scenes. In *Scandanavian Conf. Image Analysis*, Kangerlussuaq, Greenland, pp. 111–118.

Tomasi, C. and Kanade, T. 1992. Shape and motion from image streams under orthography: A factorization method. *Int. J. Computer Vision*, 9(2):137–154.

Triggs, B. 2000. Plane + parallax, tensors and factorization. In *European Conf. Computer Vision*, Dublin, Ireland, Springer-Verlag, pp. 522–538.

Triggs, B., McLauchlan, P., Hartley, R., and Fitzgibbon, A. 1999. Bundle adjustment—A modern synthesis. In *IEEE Workshop on Vision Algorithms*, Kerkyra, Greece, pp. 298–376.

Weinshall, D., Anandan, P., and Irani, M. 1998. From ordinal to euclidean reconstruction with partial scene calibration. In *Workshop on 3D Structure from Multiple Images of Large-scale Environments (SMILE-98)*, R. Koch and L. Van Gool (Eds.), LNCS, vol. 1506, Springer-Verlag: Berlin, pp. 208–223.

Weinshall, D., Werman, M., and Shashua, A. 1995. Shape tensors for efficient and learnable indexing. In *IEEE Workshop on Representation of Visual Scenes*, pp. 58–65.