Linear Multi-View Reconstruction of Points, Lines, Planes and Cameras using a Reference Plane

Carsten Rother

Computational Vision and Active Perception Laboratory (CVAP)
Royal Institute of Technology (KTH), Stockholm, Sweden
carstenr@nada.kth.se   http://www.nada.kth.se/∼carstenr

Abstract

This paper presents a new linear method for reconstructing simultaneously 3D features (points, lines and planes) and cameras from many perspective views by solving a single linear system. It assumes that a real or virtual reference plane is visible in all views. We call it the Direct Reference Plane (DRP) method. It is well known that the projection relationship between uncalibrated cameras and 3D features is non-linear in the absence of a reference plane. With a known reference plane, points and cameras have a linear relationship, as shown in [16]. The main contribution of this paper is that lines and cameras, as well as, planes and cameras also have a linear relationship. Consequently, all 3D features and all cameras can be reconstructed simultaneously from a single linear system, which handles missing image measurements naturally. A further contribution is an extensive experimental comparison, using real data, of different reference plane and non-reference plane reconstruction methods. For difficult reference plane scenarios, with point or line features, the DRP method is superior to all compared methods. Finally, an extensive list of reference plane scenarios is presented, which shows the wide applicability of the DRP method.

1   Introduction

Reconstruction from multiple perspective views of a static scene has been the subject of research for more than a decade in the field of computer vision (see [5, 3]). Many different techniques for features like points, lines and planes have been suggested. Most existing methods are based on the so-called camera or matching constraints, which involve cameras and image measurements [1, 4, 9, 12]. These methods reconstruct the scene incrementally by considering a limited number of 2, 3 or 4 views at a time. A less popular approach is to reconstruct the scene using the dual structure constraints [20]. The class of factorization methods compute simultaneously the scene and motion parameters from a measurement matrix [24, 22, 13]. This approach has the major drawback that missing image measurements have to be hallucinated [10]. We will show experimentally that for certain, difficult scenarios all examined camera constraint and factorization methods fail. Our aim is to reconstruct

large scale scenes from a small set of wide baseline images. Naturally, the amount of missing image data is very high.

In order to reconstruct such scenarios, we exploit an additional source of information, a *reference plane*. The camera and structure constraints for $2, 3$ and 4 views with a known reference plane are well understood [7, 8, 2, 5, 25, 6, 19]. Hartley et al. [6] presented a linear reconstruction method for points. It exploits the fact that the camera constraints are linear in the unknown camera parameters. An extension of this approach to line features is examined in this paper. Camera constraint methods for planes were discussed in [23, 18]. All the camera constraint approaches have in common that the 3D structure and cameras are reconstructed separately. A reference plane factorization method for points, lines and planes was suggested by Triggs [25]. The drawbacks of this method is that it is only applicable to finite reference planes, 3D points on a line have to be known and missing data has to be hallucinated. In [16], a different approach for point features was presented. It is based on the observation that 3D points and cameras have a linear relationship. Consequently, all points and all cameras may be reconstructed simultaneously from a single linear system, using a minimum amount of image measurements. We will demonstrate experimentally, that for difficult scenarios, with the plane at infinity as reference plane, this approach is superior to previously suggested reference and non-reference plane methods [6, 18, 24, 1, 4, 13].

In this paper we will extend the linear approach in [16] to line and plane features. The main contribution of this paper is that 3D lines and cameras, as well as, 3D planes and cameras have a linear relationship, if a reference plane is known. In general, this relationship is non-linear. We exploit the fact that the direction of a 3D line and 3D plane, i.e. its normal, can be computed directly from the image data. Consequently, a 3D line can be represented by only 2 parameters (4(6) in general) and a 3D plane by only one parameter (3(4) in general). Due to the linear relationship, all 3D features (points, lines and planes) and all cameras can be reconstructed simultaneously from a single linear system. We call it the Direct Reference Plane (DRP) method. Note that for line features, it is *not* assumed that 3D points on a line are known. An approach, similar to the DRP method,
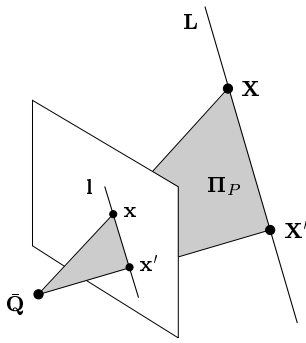
Figure 1. The projection of a 3D line L, represented by the two points X and X'.
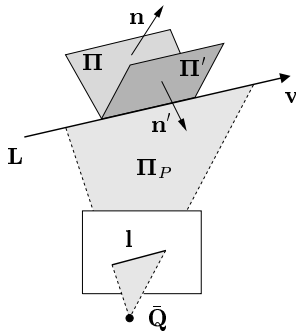

Figure 2. The 3 planes Π, Π', Π_P must intersect in the 3D line L with direction v.

was suggested in [21] where, however, the 3D structure and cameras are not determined simultaneously. This paper is based on the recently published PhD thesis [19].

2 Reference plane geometry: Linear relationship of 3D features and cameras

In the following, we investigate the projection relation of a 3D feature (point, line and plane) and an uncalibrated, perspective camera with the additional assumption of a known reference plane. We show that this relationship, which is in general non-linear, becomes linear in an affine space where the reference plane is the plane at infinity.

2.1 Points

Let us briefly review the point case, discussed in [16]. The projection of a 3D point X onto the 2D image point x is in homogeneous coordinates

$$\mathbf{x} \sim H\,(\,I\,|\,-\bar{\mathbf{Q}})\,\mathbf{X} \sim H\,(\bar{\mathbf{X}} - \bar{\mathbf{Q}}), \qquad (1)$$

where $P = H(I|-\bar{\mathbf{Q}})$ represents the projection matrix of the camera with center $\bar{\mathbf{Q}}$. Non-homogeneous and homogeneous coordinates are distinguished by a bar and "$\sim$" means equality up to scale. Consider the homography H. A point $\mathbf{X} = (X, Y, Z, 0)^T$, which lies on the plane at infinity π_∞, is mapped by eqn. (1) onto the image plane as

$$\mathbf{x} \sim H\,(X, Y, Z)^T. \qquad (2)$$

Therefore, H can be considered as the *infinite homography*, in the following denoted H^∞, between the plane at infinity π_∞ and the image plane.

In a projective setting, any (reference) plane may serve as the plane at infinity. Knowing a reference plane, real or virtual, is therefore equivalent to knowing the infinite homography. For simplicity, we stabilize the reference plane in the image, assuming that H^∞ is non-singular:

$$\mathbf{x}_{stable} \sim H^{\infty\,-1}\mathbf{x} \sim \bar{\mathbf{X}} - \bar{\mathbf{Q}}. \qquad (3)$$

In the following, $\mathbf{x}_{stable}$ is denoted $\mathbf{x}$. The stabilized image corresponds to a calibrated, translating camera, i.e. $P = (I|-\bar{\mathbf{Q}})$. The unknown scale in eqn. (3) may be eliminated by taking ratios, which gives

$$
\begin{aligned}
-y\,\bar{X} + x\,\bar{Y} + y\,\bar{A} - x\,\bar{B} &= 0 \\
-w\,\bar{X} + x\,\bar{Z} + w\,\bar{A} - x\,\bar{C} &= 0 \\
-w\,\bar{Y} + y\,\bar{Z} + w\,\bar{B} - y\,\bar{C} &= 0\,,
\end{aligned}
\qquad (4)
$$

where $\mathbf{x} = (x, y, w)^T$, $\bar{\mathbf{X}} = (\bar{X}, \bar{Y}, \bar{Z})^T$ and $\bar{\mathbf{Q}} = (\bar{A}, \bar{B}, \bar{C})^T$. These are 3 *linear* constraints for the non-homogeneous point $\bar{\mathbf{X}}$ and camera center $\bar{\mathbf{Q}}$. Note that only 2 constraints are linearly independent. This was exploited in [16] to reconstruct multiple points and cameras simultaneously from a single linear system.

2.2 Lines

In the following, two different linear constraints are derived, depending on the representation of a 3D line.

First, a 3D line L is represented by two arbitrary, but different, 3D points X, X' (fig. 1). Assume that the image is stabilized, i.e. the image line l is $l \sim H^{\infty\,T}l$. The condition that the image points x, x' lie on l may be written as

$$\mathbf{x}^T\,\mathbf{l} = 0 \ \text{ and } \ \mathbf{x}'^T\,\mathbf{l} = 0\,. \qquad (5)$$

Inserting eqn. (3) into eqn. (5) gives

$$\bar{\mathbf{X}}^T\mathbf{l} - \bar{\mathbf{Q}}^T\mathbf{l} = 0 \ \text{ and } \ \bar{\mathbf{X}}'^T\mathbf{l} - \bar{\mathbf{Q}}^T\mathbf{l} = 0\,. \qquad (6)$$

These are two *linear* constraints for the unknown line parameters $(\bar{\mathbf{X}}, \bar{\mathbf{X}}')$ and camera parameters $(\bar{\mathbf{Q}})$. Since two 3D points (6 degrees of freedom) over-determine a 3D line (4 degrees of freedom), two additional constraints are needed to specify the two points $\bar{\mathbf{X}}, \bar{\mathbf{X}}'$ uniquely. To reconstruct a 3D line from multiple views, the additional point constraints in eqn. (4) may be used, derived from the two endpoints of the line segment in one view.

We derive now linear constraints for a *minimal* line representation. The line L is uniquely defined by the two planes $\mathbf{\Pi} = (\mathbf{n}, l)^T$, $\mathbf{\Pi}' = (\mathbf{n}', l')^T$ (fig. 2). We first show that the plane normals $\mathbf{n}, \mathbf{n}'$ can be derived, given the plane at infinity (reference plane). The direction of a 3D line may be determined from the stabilized image lines l_i of multiple
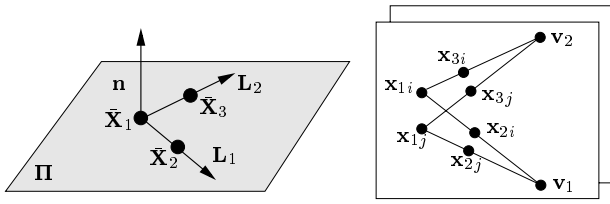
Figure 3. The points $\bar{\mathbf{X}}_1, \bar{\mathbf{X}}_2, \bar{\mathbf{X}}_3$ define uniquely the plane Π. Two stabilized images i, j are shown on the right.

views (see [5] page 330). All $\mathbf{l}_i$ have to intersect in one point $\mathbf{v}$, the vanishing point $(\mathbf{v}, 0)^T$ of $\mathbf{L}$. Therefore, the direction $\mathbf{v}$ is the 1-dimensional null-space of the linear system

$$(\mathbf{l}_1^T \cdots \mathbf{l}_n^T)^T \, \mathbf{v} = 0 \; . \tag{7}$$

Given $\mathbf{v}$, the plane normals $\mathbf{n}, \mathbf{n}'$ can be derived from the 2-dimensional null-space of $\mathbf{v}^T \mathbf{n} = 0$. Consequently, the remaining unknowns of the planes Π, Π', i.e. (l, l'), are a minimal representation of the line $\mathbf{L}$. Consider the projection constraints for $\mathbf{L}$. The condition that $\mathbf{L}$ projects onto $\mathbf{l}$ means that the 3 planes Π_P, Π and Π' intersect in $\mathbf{L}$ (see fig.2). The algebraic equivalent is that the 4×3 matrix $M = [\Pi_P \, \Pi \, \Pi']$ must have rank 2. The plane Π_P is $P^T \mathbf{l} = (\mathbf{l}, -\bar{\mathbf{Q}}^T \mathbf{l})^T$, using $P = (I|-\bar{\mathbf{Q}})$. Since M has rank 2, the 4 sub-determinants of size 3×3 have to vanish. Those sub-determinants which involve unknown parameters are

$$\begin{vmatrix} \mathbf{l}_x & \mathbf{n}_x & \mathbf{n}'_x \\ \mathbf{l}_y & \mathbf{n}_y & \mathbf{n}'_y \\ -\bar{\mathbf{Q}}^T\mathbf{l} & l & l' \end{vmatrix} = 0 \; , \quad \begin{vmatrix} \mathbf{l}_x & \mathbf{n}_x & \mathbf{n}'_x \\ \mathbf{l}_z & \mathbf{n}_z & \mathbf{n}'_z \\ -\bar{\mathbf{Q}}^T\mathbf{l} & l & l' \end{vmatrix} = 0 \; ,$$

$$\begin{vmatrix} \mathbf{l}_y & \mathbf{n}_y & \mathbf{n}'_y \\ \mathbf{l}_z & \mathbf{n}_z & \mathbf{n}'_z \\ -\bar{\mathbf{Q}}^T\mathbf{l} & l & l' \end{vmatrix} = 0 \; , \tag{8}$$

where $\mathbf{l}^{(')} = (\mathbf{l}_x^{(')}, \mathbf{l}_y^{(')}, \mathbf{l}_z^{(')})^T$ and $\mathbf{n}^{(')} = (\mathbf{n}_x^{(')}, \mathbf{n}_y^{(')}, \mathbf{n}_z^{(')})^T$. As above, the 3 constraints (2 linearly independent) are *linear* in the unknown camera parameters $\bar{\mathbf{Q}}$ and line parameters (l, l'). An alternative derivation using Plücker line coordinates gives equivalent constraints (see [19]).

2.3 Planes

Consider a 3D point $\mathbf{X}$ which lies on the plane $\Pi = (\mathbf{n}, d)^T$ and is visible in two views i, j as $\mathbf{x}_i, \mathbf{x}_j$. The image points are related by the homography H_{ij} as $\mathbf{x}_j \sim H_{ij} \mathbf{x}_i$ (see [5]). For stabilized images, the corresponding homography is $H_{ij} = H_j^{\infty -1} H_{ij} H_i^{\infty}$. Given the homography, there are in general two options to reconstruct the planar scene. Either use the homography directly or "hallucinate" image features, e.g. points, on the plane (see [23]). Our method belongs to the first class and is based on linear constraints involving plane and camera parameters. For stabilized im-

ages, the following relationship holds (see [23])

$$H_{ij} = \lambda \left(I + (\mathbf{n}^T \bar{\mathbf{Q}}_i + d)^{-1} \left(\bar{\mathbf{Q}}_j - \bar{\mathbf{Q}}_i \right) \mathbf{n}^T \right) \; . \tag{9}$$

The unknown scale λ may be determined directly from the homography H_{ij}, as we will explain. Eqn. (9) may be rewritten as

$$\hat{H}_{ij} = \lambda^{-1} H_{ij} - I = (\mathbf{n}^T \bar{\mathbf{Q}}_i + d)^{-1} \left(\bar{\mathbf{Q}}_j - \bar{\mathbf{Q}}_i \right) \mathbf{n}^T . \tag{10}$$

Since the matrix $\hat{H}_{ij}$ has rank 1, H_{ij} has the double eigenvalue λ. This was previously shown in e.g. [11] and relates to the fact that H_{ij} is a planar homology (see [2]).

As with 3D lines, the orientation of the plane, i.e. its normal $\mathbf{n}$, can be derived given the plane at infinity. Fig. 3(left) shows 3 finite points $\bar{\mathbf{X}}_{1-3}$ defining the plane Π. The vanishing points $(0, \mathbf{v}_1)^T, (0, \mathbf{v}_2)^T$ of the lines $\mathbf{L}_1 = (\mathbf{X}_1, \mathbf{X}_2), \mathbf{L}_2 = (\mathbf{X}_1, \mathbf{X}_3)$ are given as (see fig. 3(right))

$$\begin{aligned} \mathbf{v}_1 &= (\mathbf{x}_{1i} \times \mathbf{x}_{2i}) \times (\mathbf{x}_{1j} \times \mathbf{x}_{2j}) \\ \mathbf{v}_2 &= (\mathbf{x}_{1i} \times \mathbf{x}_{3i}) \times (\mathbf{x}_{1j} \times \mathbf{x}_{3j}) \; . \end{aligned} \tag{11}$$

Consequently, the plane's normal is $\mathbf{n} = \mathbf{v}_1 \times \mathbf{v}_2$. In practice, we hallucinate many points to compute $\mathbf{n}$ stably.

Given $\mathbf{n}$ and λ, the remaining unknown plane parameter is d. Eqn. (10) may be rewritten as the following 9 constraints

$$\bar{A}_i(\hat{h}_{1l} n_x - n_l) + \bar{B}_i(\hat{h}_{1l} n_y) + \bar{C}_i(\hat{h}_{1l} n_z) + \bar{A}_j n_l + d\hat{h}_{1l} = 0$$
$$\bar{A}_i(\hat{h}_{2l} n_x) + \bar{B}_i(\hat{h}_{2l} n_y - n_l) + \bar{C}_i(\hat{h}_{2l} n_z) + \bar{B}_j n_l + d\hat{h}_{2l} = 0$$
$$\bar{A}_i(\hat{h}_{3l} n_x) + \bar{B}_i(\hat{h}_{3l} n_y) + \bar{C}_i(\hat{h}_{3l} n_z - n_l) + \bar{C}_j n_l + d\hat{h}_{3l} = 0$$
$$\tag{12}$$

for $l = 1, 2, 3$, $\hat{H}_{ij} = (\hat{h})_{ml}$ and $\bar{\mathbf{Q}}_i = (\bar{A}_i, \bar{B}_i, \bar{C}_i)^T$. These constraints are *linear* in the unknown plane parameter (d) and camera parameters $(\bar{\mathbf{Q}}_i, \bar{\mathbf{Q}}_j)$. It can be shown that only 3 of the 9 equations are linearly independent.

3 Outline of the DRP Method

The Direct Reference Plane (DRP) method for reconstructing multiple 3D features and cameras consists of the following 3 steps:

1. Determine a reference plane, i.e. homography H^{∞}
2. Compute the directions (orientations) of lines and planes
3. Recover all features and cameras from one linear system

For multiple points $\bar{\mathbf{X}}_i$, lines (l_i, l_i') or $(\bar{\mathbf{X}}_i, \bar{\mathbf{X}}_i')$, planes d_i and camera centers $\bar{\mathbf{Q}}_i$, the linear system takes the form

$$\begin{pmatrix} \text{formed from eqns. (4)} \\ \vdots \\ \text{formed from eqns. (8)} \\ \text{or eqns. (6) and eqns. (4)} \\ \vdots \\ \text{formed from eqns. (12)} \end{pmatrix} \begin{pmatrix} \bar{\mathbf{X}}_i \\ \vdots \\ l_i \text{ or } \bar{\mathbf{X}}_i \\ l_i' \text{ or } \bar{\mathbf{X}}_i' \\ \vdots \\ d_i \\ \vdots \\ \bar{\mathbf{Q}}_i \end{pmatrix} = 0 \; . \tag{13}$$

Real or Virtual Reference Plane Scenarios	Reconst.
Real scene plane	projective
3 mutually orthogonal scene directions	metric
Additional orthographic "over"view (map)	metric
Calibrated cameras with known/constant rotation	metric
Translating camera with constant calibration	affine
Affine cameras + 3 points in all views	affine
Known epipolar geometry + 3 points in all views	projective
Small image baseline	projective

Table 1. Various reference plane scenarios.

The null-space of the linear system is 4-dimensional (see [17]). Therefore, the sum of the 4 singular vectors of the null-space provide the complete reconstruction. The main advantages of the DRP method, in contrast to iterative methods, like [1, 4], and factorization methods, like [24, 22, 13], is that *all* 3D features and cameras are reconstructed simultaneously using a *minimum* amount of image measurements. We denote the DRP method for single feature types Point-DRP, Line-DRP and Plane-DRP method respectively. The DRP method which uses a minimal line representation, i.e. (l, l'), is called Line-DRP(min).

It is well known that a reference plane, i.e. the infinite homography, may be derived for scenes without a real plane. Table 1 lists many known reference plane scenarios (see [19] for details). The list shows the wide applicability of the DRP method. We point out that for some scenarios iterative (sub-optimal) methods were suggested in the past, e.g. for translating cameras with constant calibration [26]. The case of an additional map was recently presented in [15]. The assumption of affine cameras or known epipolar geometry was discussed in [17]. Oliensis [14] exploited a small baseline to simplify the reconstruction problem.

Finally, we mention an important practical issue. The linear eqns. (4), (6), (8) and (12) are only valid for 3D features and camera centers outside the reference plane. This is not an issue if the reference plane is at infinity. Indeed, all scenarios in table 1 which give an affine or metric reconstruction have this property[1]. For finite reference planes, features on and off the plane have to be separated (see [17]). Further, minor issues, omitted here (see [19]) are normalization of image features, weighting of the linear system in eqn. (13) and a dealing with a singular infinite homography.

4 Experiments

We are interested in reconstructing complex, large scale objects, e.g. buildings, from a small set of of images with a wide baseline. Such scenarios are very different to reconstructing objects from a continuous image sequences, since the amount of missing image measurements is high (up to 90%). In the following, we present only real data experiments. Further experiments using real and synthetic data,

---

[1]An exception are affine cameras and 3 common points (see [19]).

infinite and finite reference planes are in [19]. For point and line features, the real plane at infinity served as the reference plane. It is derived from 3 mutually orthogonal scene directions and the assumption of a square pixel camera (see table 1 and [16]). To compare the DRP method with other methods, the different feature types are considered separately.

4.1 Points

For point features, our method (**Point-DRP**) is compared to the following 5 methods. The **Point-Fmat** method in [4] merges subsets of views in an optimal fashion. A bundle adjustment is carried out after each merging step. The algorithm is initialized by computing all possible subsets of 2 views, which is feasible for wide-baseline images. The **Point-ResInt** method in [1] is based on the "intersection-resection" scheme, where bundle adjustment is carried out after each resection step. The recently published projective factorization method (**Point-ProjFac**) in [13] extends the method in [22] for missing data using [10]. The same missing data algorithm [10] is applied to the affine factorization method (**Point-AffFac**) in [24]. The **Point-Cam** method of Hartley et al. [6] reconstructs all cameras simultaneously using a reference plane. In our implementation, all camera constraints derived from 2, 3 and 4 views are used simultaneously. Note that Triggs reference plane method [25] is not applicable here since the reference plane is at infinity.

In the first experiment (campus sequence) 191 3D points belonging to 3 buildings were reconstructed from 27 views (see fig. 4 (a,b)). The visibility matrix in fig. 5(b) shows that only 10% of the image data is available (black dots). The accurate reconstruction of the Point-DRP method with a superimposed map is depicted in fig. 5(a). Points are dots and cameras as arrows. Fig. 6(a) depicts the RMS reprojection error after applying the different methods. Only the two reference plane methods (Point-DRP, Point-Cam) could reconstruct this scenario. For 3 methods (ResInt, ProjFac, AffFac) the amount of image data was too little. For further comparison, we "synthesized" the campus sequence. This means that the reconstruction in fig. 5(a) was taken as "ground truth". Fig. 6(b) shows the 3D error (approx. in meter) depending on Gaussian noise on the image points (before bundle adjustment). The 3D error is measured after aligning the obtained reconstruction with the ground truth in an optimal way. The Point-Fmat method could only reconstruct the scene in the case of no noise. Furthermore, the Point-DRP method is superior to the Point-Cam method.

In the next experiment (house sequence), a house is reconstructed from 9 views (fig. 4 (c, d)) and 451 automatically matched image points. The visibility matrix in fig. 7(a) has 64% missing data. Fig. 7(b) illustrates the 3D reconstruction using the Point-DRP method. Fig. 8(a) shows that only the reference plane methods and the Point-ResInt method achieved a reconstruction with a small RMS error. The metrically rectified 3D reconstruction using the Point-

Figure 4. Two views of the campus sequence (a,b) and the house sequence (c,d).

ResInt method is depicted in fig. 7(b). The result is bad, i.e. the non-linear optimisation got stuck in a local minimum, far from the global minimum. The difficulty of this scenario is that certain subsets of views observe a dominant scene plane, which is a critical configuration. That is the reason for the failure of the iterative Point-Fmat method. In [12], strategies to overcome this problem were suggested. Note that the set of critical configurations is very small in the reference plane case (see [19]). Fig. 8(b) shows the performance on the synthetic house sequence. For a noise level of more than 0.1, all non-reference plane methods failed. Both reference plane methods performed equally good.

We may conclude that for difficult scenarios, non-reference plane methods might fail due to (a) too few image measurements are available, (b) error accumulation due to noisy image measurements or (c) critical configurations (like a dominant scene plane).

4.2 Lines

For line features, we compare the following 3 methods. In sec. 3 we introduced the **Line-DRP** method, which represents a 3D line by two 3D points, and the **Line-DRP(min)** method, which uses a minimal line representation (l, l'). Furthermore, we extended the Point-Cam method in [6] for line features (**Line-Cam**). This is possible since the camera constraints for lines in 3 and 4 views are linear if a reference plane is known (see [19]).

We reconstructed the house (fig. 4(c,d)) from 70 manually selected line segments. The amount of missing line features is 54%. Fig. 9 (a,b) shows two views of the reconstruction using the Line-DRP method. For comparison, fig. 9 (c,d) depicts the result of the Point-DRP method using the endpoints of the line segments. This result is qualitatively better. This can be expected since 3D points provide more geometric constraints than 3D lines. Fig. 10(a) shows the RMS reprojection error using various methods. The RMS error is based on the endpoints of the line segments, determined by intersection. Although the line-based methods performed worse, the errors are identical for all methods after bundle adjustment. Fig. 10(b) depicts the performance on the synthetic house sequence, i.e. the 3D reconstruction in fig. 9 (c,d) was taken as ground truth. As in the real data case (fig. 10(a)), the Line-DRP method is superior to other line-based methods. A potential drawback of the Line-DRP(min) method is that the line directions are computed separately in a pre-processing step.

We may conclude that point-based methods are significantly superior to line-based methods. Furthermore, the Line-DRP method was the best line-based method for this scenario (with an infinite reference plane).

4.3 Planes

As discussed in sec. 2.3, planes may be reconstructed directly, given the homographies, or by "hallucinating" image features, e.g. points. In the following, we compare our **Plane-DRP** method with the **Plane-Cam** method suggested in [18]. Both methods use the homographies directly. The Plane-Cam method reconstructs all cameras simultaneously from a linear system. Furthermore, we will hallucinate point features and reconstruct them with several point-based methods introduced in sec. 4.1. As pointed out in [23], it is important to hallucinate point features inside the image area of the homographies.

We reconstructed 6 planes of a toyhouse from 8 images (see fig. 11(a,b)). The homographies were determined manually from point matches. The ground plane served as the reference plane. Fig. 11(c) shows the reconstruction using the Plane-DRP method. The 3D points on the planes were determined by intersection. The RMS reprojection error of the reconstructed 3D points is depicted in fig. 12(a). Apparently, the "direct homography" methods (Plane-DRP, Plane-Cam) performed worse than methods using hallucinated points. In this case, the iterative Point-Fmat method performed best. This might be due to the small baseline and the non-critical configuration (no dominant plane). Note that this method performs bundle adjustment after each merging step. As with lines, the RMS errors of all methods are identical after bundle adjustment. The result using the synthetic toyhouse sequence (fig. 12(b)) confirms the conclusions of the previous, real data experiment.

We may conclude that for this scene, methods which use hallucinated point features are superior to methods which reconstruct the planes directly, given the homographies.

5 Summary and Conclusions

This paper presented a novel linear method for reconstructing simultaneously 3D features (points, lines and planes) and cameras from multiple perspective views. It assumes that a reference plane is visible in all views. We call it the Direct Reference Plane (DRP) method. It extends the point-based technique in [16] to line and plane features. The main

theoretical contribution of this paper is that lines and cameras, as well as, planes and cameras have a linear projection relationship if a reference plane is known. This relationship is in general non-linear. Consequently, all 3D features and cameras can be reconstructed from a single linear system. Since the linear projection relations are used directly, the minimum amount of image measurements is sufficient.

The main practical contribution of this paper is a comparison of the DRP method with several reference plane [6, 18] and non-reference plane methods [24, 1, 4, 13] for difficult scenarios. We choose to reconstruct large scale objects, i.e. buildings, from a small set of images with the reference plane as the plane at infinity. Consequently, the amount of missing image measurements is high. For point and line features, only the DRP method and Hartley's et al. [6] reference plane method performed successfully. The failure of the non-reference plane methods had the following reasons: (a) too few image measurements, (b) error accumulation or (c) critical configurations (dominant scene plane). The reference plane methods circumvent all these problems. The DRP method was superior to Hartley's et al. method before bundle adjustment. The experiments with planes showed that reconstructing hallucinated point features, with the Point-DRP method, gives better results than reconstructing the planes directly from the homographies.

Finally, we presented an extensive list of reference planes scenarios (table 1) for which the DRP method is applicable. We pointed out that for some scenarios iterative (sub-optimal) methods were suggested in the past, e.g. translating cameras with constant calibration [26].

As described in this paper, the DRP method can reconstruct all features and cameras simultaneously. In the presented experiments the three feature types were considered separately. Experiments with combination of features are planed. A weighting of the linear constraints, corresponding to the uncertainty of the given image measurements might be adventurous. Furthermore, all scene constraints which are linear in the 3D features may be integrated in the linear system. These constraints are incidence relationships, e.g. a point lies on a plane, and constraints of known features (see [19]). The integration and test of such a method is the subject of future research.

References

[1] P. Beardsley, P. H. S. Torr and A. Zisserman (1996). 3D Model Acquisition from Extended Image Sequences, ECCV, Cambridge, UK, p. 683-695.

[2] A. Criminisi, I. Reid and A. Zisserman (1998), Duality, rigidity and planar parallax, ECCV, Germany, p. 846-861.

[3] O. D. Faugeras and Q.-T. Luong (2001). *The Geometry of Multiple Images*, The MIT Press.

[4] A. W. Fitzgibbon and A. Zisserman (1998). Automatic camera recovery for closed or open image sequences, ECCV, Freiburg, Germany, p. 311-326.

[5] R. I. Hartley and A. Zisserman (2000). *Multiple View Geometry in Computer Vision*, Cambridge University Press.

[6] R. I. Hartley, N. Dano and R. Kaucic (2001). Plane-based Projective Reconstruction, ICCV, Canada, p. 420-427.

[7] M. Irani and P. Anandan (1996). Parallax geometry of pairs of points for 3d scene analysis. ECCV, UK, p. 17-30.

[8] M. Irani, P. Anandan and D. Weinshall (1998). From reference frames to reference planes: Multi-view parallax geometry and applications. ECCV, Germany, p. 829-845.

[9] R. Koch, M. Pollefeys and L. VanGool (1998). Multi viewpoint stereo from uncalibrated video sequences. ECCV, Freiburg, Germany, p. 55-65.

[10] D. Jacobs (1997). Linear Fitting with Missing Data for Structure-from-Motion, CVPR, Puerto Rico, p. 206-212.

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

[12] M. Pollefeys, F. Verbiest and L. Van Gool (2002). Surviving Dominant Planes in Uncalibrated Structure and Motion Recovery, ECCV, Copenhagen, Denmark, p. 837-851.

[13] D. Martinec and T. Pajdla (2002). Structure from Many Perspective Images with Occlusions, ECCV, Copenhagen, Denmark, p. 355-369.

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

[15] D. P. Robertson and R. Cipolla (2002). Building Architectural Models from Many Views Using Map Constraints, ECCV, Copenhagen, Denmark, p. 155-169.

[16] C. Rother and S. Carlsson (2001). Linear Multi View Reconstruction and Camera Recovery, ICCV, Canada, p. 42-51.

[17] C. Rother and S. Carlsson (2002). Linear Multi View Reconstruction with Missing Data, ECCV, Denmark, p. II 309-324.

[18] C. Rother, S. Carlsson and D. Tell (2002). Projective Factorization of Planes and Cameras in Multiple Views, ICPR, Quebec, Canada, p. 737-740.

[19] C. Rother (2003). Multi-View Reconstruction and Camera Recovery using a Real or Virtual Reference Plane, PhD thesis, KTH, Stockholm, Sweden. ISBN 91-7283-422-6.

[20] F. Schaffalitzky, A. Zisserman, A., R. I. Hartley and P. H. S. Torr (2000). A Six Point Solution for Structure and Motion. ECCV, Dublin, Ireland, p. 632-648.

[21] H.-Y. Shum, M. Han and R. Szeliski (1998). Interactive construction of 3D models from panoramic mosaics. CVPR, Santa Barbara, WA, p. 427-433.

[22] P. Sturm and B. Triggs (1996). A factorization based algorithm for multi-image projective structure and motion, ECCV, Cambridge, UK, p. 709-719.

[23] R. Szeliski and P. H. S. Torr (1998). Geometrically Constrained Structure from Motion: Points and Planes, Workshop SMILE-98, LNCS 1506, Springer-Verlag, p. 171-186.

[24] C. Tomasi and T. Kanade (1992). Shape and motion from image streams under orthography: a factorization method, Int. J. Computer Vision, 9(2):137-154.

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

[26] L. Van Gool, T. Moons, M. Proesmans and M. Van Diest (1994). Affine Reconstruction from Perspective Image Pairs Obtained by a Translating Camera, ICPR, Israel, p. 290-294.
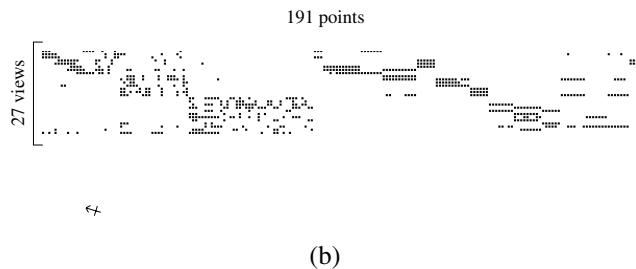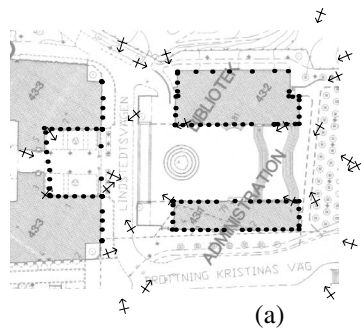
IEEE
COMPUTER
SOCIETY

(a)   (b)

Figure 5.Reconstruction of the campus with the Point-DRP method (a) and the visibility matrix (b).

(a)   (b)
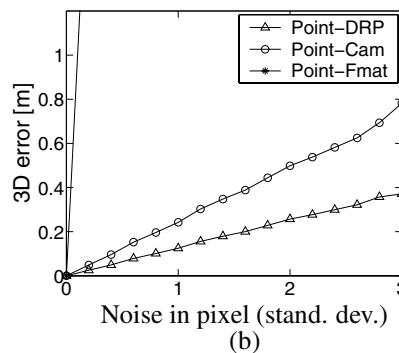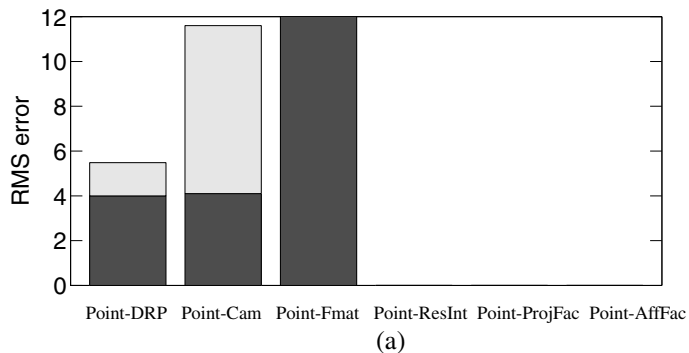
Figure 6. Results of various point-based algorithms on the *real* campus sequence (a) and the *synthetic* campus sequence (b), before bundle adjustment (grey) and after bundle adjustment (black).
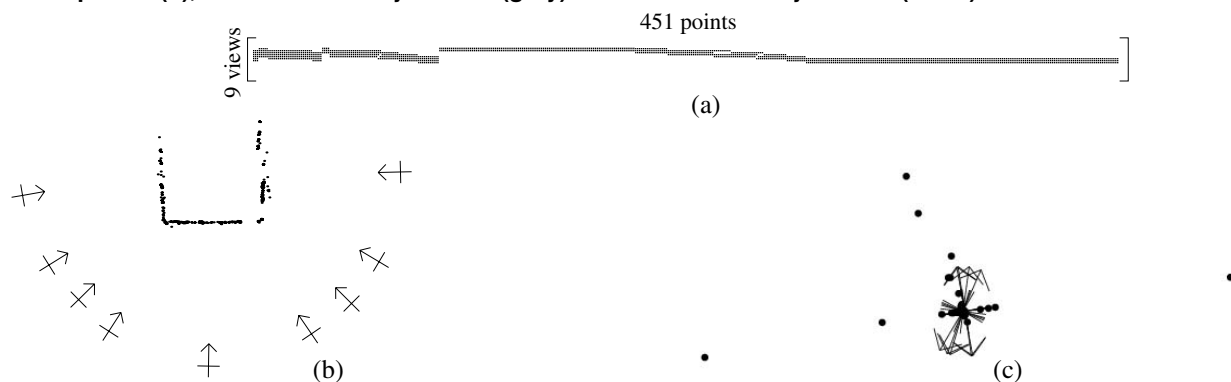

(a)

(b)   (c)

Figure 7. The visibility matrix of the house sequence (a). Top view of the house reconstruction with the Point-DRP method (b) and Point-ResInt method (c).

(a)   (b)
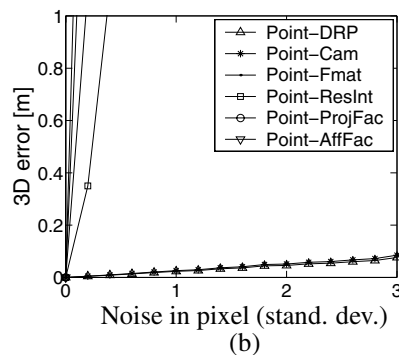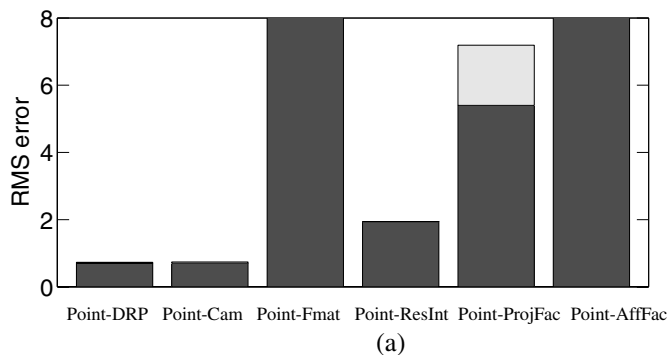
Figure 8. Results of various point-based algorithms on the *real* house sequence (a) and the *synthetic* house sequence (b), before bundle adjustment (grey) and after bundle adjustment (black).

(a)   (b)   (c)   (d)
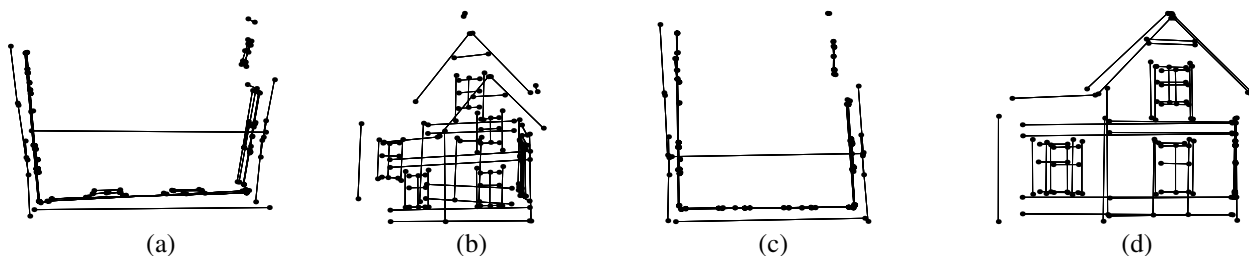
Figure 9. Two views of the reconstructed house using the Line-DRP algorithm (a,b) and the Point-DRP algorithm (endpoints of the line segments) (c,d). The results are before bundle adjustment.
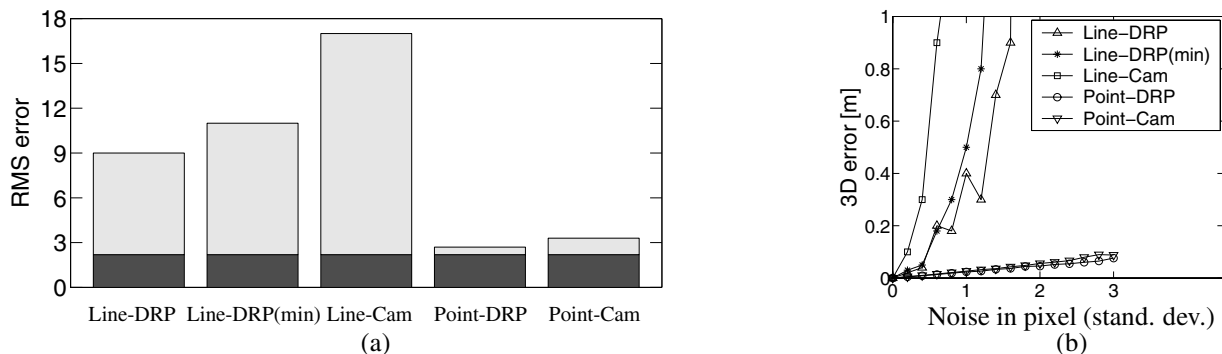

(a)   (b)

Figure 10. Results of various line-based algorithms on the *real* house sequence (a) and the *synthetic* house sequence (b), before bundle adjustment (grey) and after bundle adjustment (black).

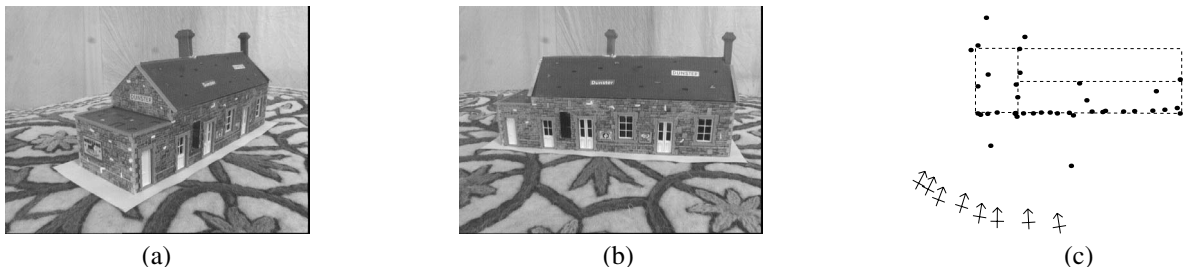(a)   (b)   (c)

Figure 11. Two images of the toyhouse sequence (a,b). Top view of the reconstruction of the toyhouse using the Plane-DRP method (c). The dashed lines mark the toyhouse.

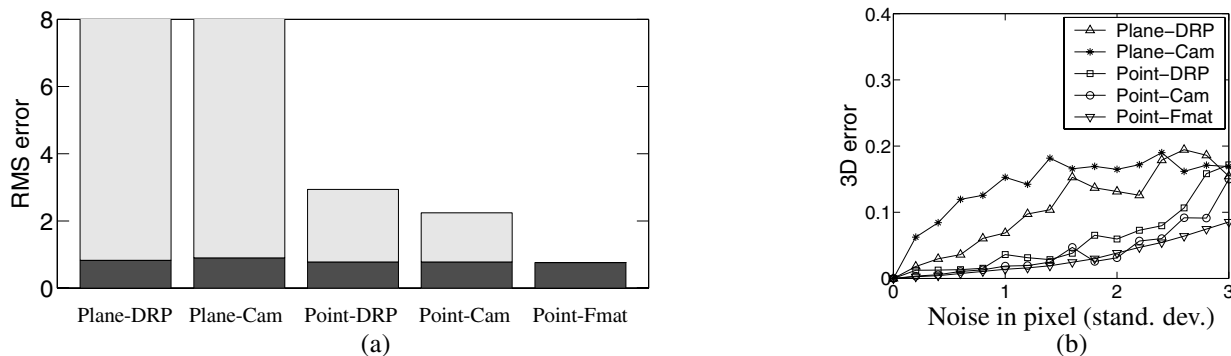(a)   (b)

Figure 12. Results of various plane-based algorithms on the *real* toyhouse sequence (a) and the *synthetic* toyhouse sequence (b), before bundle adjustment (grey) and after bundle adjustment (black).

IEEE
COMPUTER
SOCIETY