Ground-Plane Based Projective Reconstruction for Surveillance Camera Networks

David N. R. M^{c}Kinnon[1,2], Ruan Lakemond[1], Clinton Fookes[1] and Sridha Sridharan[1]

(1)Image & Video Research Laboratory, Queensland University of Technology, GPO Box 2434, 2 George St, Brisbane, Queensland 4001

(2)Australasian CRC for Interaction Design, PO Box 2008, Kelvin Grove, Brisbane, Queensland 4059

{*d.mckinnon, r.lakemond, c.fookes, s.sridharan*}@qut.edu.au

***Abstract*—This paper examines the task of calibrating a network of security cameras of arbitrary topology utilizing a common ground-plane to simplify the geometry. The method of chaining ground-plane homographies is employed to overcome the necessity that a common region is visible in all images of the ground-plane. We show that this method of recovering a projective reconstruction is ideal for the case of surveillance camera networks due to the ability to linearly initialize the cameras and structure when making use of the common ground-plane assumption.**

I. Introduction

In order to perform tracking and recognition tasks from networks of security cameras it is first necessary to determine the calibration properties of the cameras. The calibration of a camera can be described in terms of it's internal and external properties [1].

The internal properties (such as the focal length) enable metric measurements and angles to be properly recovered from the images. The external properties (such as rotation and translation) enable exact correspondences to be determined between the images via a projective coordinate system. Both the internal and external properties must be recovered in order for a full Euclidean reconstruction to be possible.

For a large range of tracking applications a projective reconstruction of the cameras and scene is adequate for the essential task of track hand-over between the images. In the case that the cameras form a large reconfigurable network, with no specific knowledge of the networks topology or fiducial markings with which to calibrate the cameras, it is desirable to have a reliable means to determine a projective reconstruction.

In this paper we present a projective reconstruction scheme utilizing the constraint that all the cameras in the network view the same ground-plane. We overcome the necessity that all the cameras need to view the same section of ground-plane by chaining correspondences from non-overlapping views back to a base view in the camera network.

A number of authors in the multiple-view geometry literature have addressed the topic of projective reconstruction utilizing common plane correspondences [2], [3], [4], [5], [6]. The geometry of a scene relative to a plane is commonly referred to as the plane + parallax model [7]. In this model the projective structure of each camera is defined by 3 parallax parameters in addition to 4 parameters specifying a homography between images of a common plane in the scene.

The general observation is that the calculation of the external properties of the cameras becomes simplified in the case that a common plane is readily identifiable in the images. In [4] the process of determining a projective reconstruction by standard sequential means [8] was compared with a plane-based scheme indicating improved stability and results for the plane-based approach.

The advantages of the plane-based approach are that correspondences between common scene points can be applied simultaneously across many images in the solution of the camera's parallax (only factorization [7] is able to achieve this), and far fewer correspondences are required as a minimum (3) to solve for the camera's parallax (assuming that the plane homography can be found).

The second of these points is critical for the purpose of surveillance imaging where the images are often of poor quality with very limited overlap capable of generating image correspondences. This means that an initial projective reconstruction can be obtained in one linear computation – overcoming the main difficulty in alternative structure from motion approaches – and used as a starting point for a non-linear minimization.

This paper proceeds by applying some of the plane-based approaches proposed in the literature with several adaptations to the task of determining a projective reconstruction of a network of surveillance cameras. The major novelties are the adaptation of the direct method into a sparse least-squares problem (Section II-A) and the proposal of the plane-based bundle adjustment (Section II-C). The concept of chaining homographies has been used in other plane-based approaches [6] but here is used in a different context to chain together arbitrary networks recursively.

A. Plane-Based Camera Geometry

The correspondence of a set of points $(\bar{\mathbf{x}}_i)$ and $(\bar{\mathbf{x}}_j)$ from images i and j lying on a plane (π) in the scene is expressed in terms of the homography $\mathbf{H}^{\pi}_{i,j}$ as follows,

$$\bar{\mathbf{x}}_i \simeq \mathbf{H}^{\pi}_{i,j}\bar{\mathbf{x}}_j. \tag{1}$$

The plane (π) is assumed to be the ground-plane exclusively in this paper. Furthermore the expression for a camera (i) observing points in general position (i.e. off-plane points) can be given in terms of the planar-homography $\mathbf{H}^{\pi}_{0,i}$ (from the first image labeled 0) and a further vector $\mathbf{d}_{0,i}$ specifying the parallax relative to the plane,

$$\mathbf{x}_i \simeq \left[\ \mathbf{H}^{\pi}_{0,i} \mid \mathbf{d}_i\ \right] \mathbf{X}, \tag{2}$$

where $\mathbf{x}_i$ are the projections of the off-plane points $\mathbf{X}$ to image i.

Assuming that the homography $\mathbf{H}^{\pi}_{0,i}$ can be reliably determined from correspondences relative to a common plane, the task of determining a projective reconstruction for a set of n cameras viewing a scene can be reduced to one of determining the $3(n-1)$ **DOF** associated with vectors $\mathbf{d}_{0,i}$ for $i = 1, \ldots, n-1$ describing the parallax off the plane.

B. Chaining Homographies

The assumption when solving for plane-based homographies $\mathbf{H}^{\pi}_{0,i}$ (for $i = 1, \ldots, n-1$) is that correspondences ($\mathbf{x}^k_0 \leftrightarrow \mathbf{x}^k_i$) exist between the set of points $\mathbf{x}^k_0$ (for $k = 1, \ldots, m$) in image 0 and $\mathbf{x}^k_i$ in image i. The minimum number of point (or line) based correspondences necessary to fully determine the homography is 4 and it's solution can be obtained via well-known least-squares methods [3].

It is common for the cameras in surveillance networks not to share a common section of ground-plane visible from all images. If this is the case then it is not possible to calculate the homographies $\mathbf{H}^{\pi}_{0,i}$ for all the cameras. Instead relative homographies must be determined $\mathbf{H}^{\pi}_{j,i}$ between sets of images that do share common visible sections of ground-plane.

In order to utilize the information contained in the homography $\mathbf{H}^{\pi}_{j,i}$ in either of the reconstruction schemes (outlined in Section II) it is necessary to be able to recover $\mathbf{H}^{\pi}_{0,i}$ from it by means of chaining it to other homographies in the set in the following fashion (see [6]),

$$\mathbf{H}^{\pi}_{0,i} = \mathbf{H}^{\pi}_{j,i}\mathbf{H}^{\pi}_{0,j}. \tag{3}$$

The key point here is that it is always possible to recursively calculate the homography $\mathbf{H}^{\pi}_{0,j}$ from (3) for networks of cameras where there is a continuous string of correspondences (of sufficient number) to determine $\mathbf{H}^{\pi}_{j-1,j}$ (for $j = 1, \ldots, i-1$) in a sequential manner or otherwise.

II. Plane-Based Reconstruction Schemes

In this Section we review two of the linear plane-based reconstruction schemes presented in the literature [3], [4], as well as presenting a novel formulation of the non-linear bundle-adjustment of the cameras and scene points assuming a plane-based camera model.

A. Direct Reconstruction

The Direct Reconstruction method was proposed in [3] and seeks to find a solution to the parallax vectors $\mathbf{d}_{0,i}$ (for $i = 1, \ldots, n-1$) and the affine part of the (off-plane) scene points $\mathbf{X}^j = [\hat{\mathbf{X}}^{j\top}, 1]^\top$ (for $j = 1, \ldots, m$). Firstly equation (2) is restated as

$$\mathbf{x}^j_i \simeq \left[\ \mathbf{H}^{\pi}_{0,i} \mid \mathbf{d}_{0,i}\ \right] \begin{pmatrix} \hat{\mathbf{X}}^j \\ 1 \end{pmatrix}, \tag{4}$$

which can then be rewritten in the form,

$$[\mathbf{x}^j_i]_\times \left[\ \mathbf{H}^{\pi}_{0,i} \mid \mathbf{d}_{0,i}\ \right] \begin{pmatrix} \hat{\mathbf{X}}^j \\ 1 \end{pmatrix} = \mathbf{0}_3. \tag{5}$$

This can be rearranged once again to present a linear system in the coefficients of the unknown parameters as follows:

$$\left[\begin{array}{ccc|c} -1 & 0 & x & x\mathbf{h}^3 - \mathbf{h}^1 \\ 0 & 1 & -y & -y\mathbf{h}^3 + \mathbf{h}^2 \end{array}\right] \begin{pmatrix} \mathbf{d}_{0,i} \\ \hat{\mathbf{X}}^j \end{pmatrix} = \mathbf{0}_2, \tag{6}$$

where $\mathbf{x}^j_i = [x, y, 1]^\top$ and $\mathbf{h}^k$ is the k^{th} row of the matrix $\mathbf{H}^{\pi}_{0,i}$. The left and right hand block of this constraint can be rewritten more succinctly as $[\ \mathbf{A}^j_i \mid \mathbf{B}^j_i\ ]$ and the unknown parameters as $\mathbf{d} = [\mathbf{d}^\top_{0,1}, \ldots, \mathbf{d}^\top_{0,n-1}]^\top$ and $\mathbf{X} = [\hat{\mathbf{X}}^{1\top}, \ldots, \hat{\mathbf{X}}^{m\top}]^\top$. Each constraint of the form (6) provides 2 **DOF** toward the total of $3(n-1) + 3m$ **DOF** in the unknown parameters ($\mathbf{d}$ and $\mathbf{X}$). The combined constraints for each observation of a scene point j in image i forms a sparse linear system,

$$\mathbf{D} \begin{bmatrix} \mathbf{d} \\ \mathbf{X} \end{bmatrix} = \mathbf{0}, \tag{7}$$

with the structure of $\mathbf{D}$ (assuming all the scene points are visible in all the images) conforming to,

$$\mathbf{D} = \left[\begin{array}{ccc|ccc} & & & \mathbf{B}^1_0 & & \\ & & & & \ddots & \\ & & & & & \mathbf{B}^m_0 \\ \mathbf{A}^1_1 & & & \mathbf{B}^1_1 & & \\ \ddots & & & & \ddots & \\ \mathbf{A}^m_1 & & & & & \mathbf{B}^m_1 \\ & \ddots & & & \vdots & \\ & & \mathbf{A}^1_{n-1} & \mathbf{B}^1_{n-1} & & \\ & & \ddots & & \ddots & \\ & & \mathbf{A}^m_{n-1} & & & \mathbf{B}^m_{n-1} \end{array}\right]. \tag{8}$$

Seeing as we seek a projective reconstruction the first camera is assumed to be fixed ($\mathbf{P}_0 = [\ \mathbf{I} \mid \mathbf{0}\ ]$) and is consequently omitted (since $\mathbf{d}_{0,0} = [0, 0, 0]^\top$). A solution to (7) is possible using least-squares ideally taking into account the sparsity of $\mathbf{D}$.

B. Matching Constraint Reconstruction

This approach was first proposed in [3] and extended to the three and four view cases in [4]. The method seeks to determine a solution to just the parallax vectors $\mathbf{d}_{0,i}$ (for $i = 1, \ldots, n-1$) by utilizing matching constraints arising from correspondences between the image points ($\mathbf{x}_i^j$) in different images.

The key observation is that under the parameterization (2) the coefficients of the multi-view tensors become linear in the coefficients of $\mathbf{d}$. We introduce the following notation for the cameras $\mathbf{P}_i = [\mathbf{A}|\mathbf{a}]$, $\mathbf{P}_j = [\mathbf{B}|\mathbf{b}]$. The equation of the fundamental matrix in terms of the camera matrices $\mathbf{P}_i$ and $\mathbf{P}_j$ is,

$$\mathbf{F}_{i,j} = (i+j)^{-1} \begin{vmatrix} \sim \mathbf{A}^i & \sim \mathbf{a}^i \\ \sim \mathbf{B}^j & \sim \mathbf{b}^j \end{vmatrix}, \tag{9}$$

where $\sim$ omits the row corresponding to the superscript from the vector / matrix, for example,

$$\mathbf{F}_{1,1} = \begin{vmatrix} \mathbf{A}^2 & \mathbf{a}^2 \\ \mathbf{A}^3 & \mathbf{a}^3 \\ \mathbf{B}^2 & \mathbf{b}^2 \\ \mathbf{B}^3 & \mathbf{b}^3 \end{vmatrix}, \tag{10}$$

which can be rewritten as follows,

$$\mathbf{F}_{1,1} = -a^2 \begin{vmatrix} \mathbf{A}^3 \\ \mathbf{B}^2 \\ \mathbf{B}^3 \end{vmatrix} + a^3 \begin{vmatrix} \mathbf{A}^2 \\ \mathbf{B}^2 \\ \mathbf{B}^3 \end{vmatrix} - b^2 \begin{vmatrix} \mathbf{A}^2 \\ \mathbf{A}^3 \\ \mathbf{B}^3 \end{vmatrix} + b^3 \begin{vmatrix} \mathbf{A}^2 \\ \mathbf{A}^3 \\ \mathbf{B}^2 \end{vmatrix}. \tag{11}$$

.

Converting the coefficients of $\mathbf{F}_{i,j}$ into a vector $\mathbf{f}$, the usual approach for solving for $\mathbf{f}$ [3] is by formulating a linear system $\mathbf{S}$ composed from correspondences $\mathbf{x}_i^k \leftrightarrow \mathbf{x}_j^k$,

$$\mathbf{S}\mathbf{f} = \mathbf{0}. \tag{12}$$

Now a substitution can be made for $\mathbf{f} = \mathbf{Q}\mathbf{d}$ where $\mathbf{d} = [\mathbf{a}^\top, \mathbf{b}^\top]^\top$ and $\mathbf{Q} = [\mathbf{Q}_i|\mathbf{Q}_j]$ is a matrix composed according to the linear relationship between the coefficients of $\mathbf{f}$ and $\mathbf{d}$ of the form (11) – partitioned in order to separate the components of $\mathbf{d}$ – resulting in,

$$\mathbf{S}\mathbf{Q}\mathbf{d} = \mathbf{0}. \tag{13}$$

Assuming that the matrices $\mathbf{A}$ and $\mathbf{B}$ correspond with the known planar homographies $\mathbf{H}_{0,i}^\pi$ and $\mathbf{H}_{0,j}^\pi$, equation (13) allows for a direct solution to the parallax component of the camera matrices $\mathbf{d} = [\mathbf{d}_{0,i}^\top, \mathbf{d}_{0,j}^\top]^\top$.

By obtaining correspondences between common off-plane points over the entire set of images a sparse matrix $\mathbf{M}$ can be constructed to solve for the vector $\mathbf{d} = [\mathbf{d}_{0,1}^\top, \ldots, \mathbf{d}_{0,n-1}^\top]^\top$ as follows,

$$\mathbf{M}\mathbf{d} = \mathbf{0}. \tag{14}$$

The structure of the matrix $\mathbf{M}$ when using only sequential correspondences is,

$$\begin{bmatrix} \mathbf{SQ}_1 & & & & \\ \mathbf{SQ}_1 & \mathbf{SQ}_2 & & & \\ & \mathbf{SQ}_1 & \mathbf{SQ}_2 & & \\ & & \ddots & \ddots & \\ & & & \mathbf{SQ}_{n-2} & \mathbf{SQ}_{n-1} \end{bmatrix}, \tag{15}$$

as noted in [3] this configuration does not provide sufficient $\mathbf{DOF}$ toward resolving the parameters of $\mathbf{d}$ if $n > 4$.

In this case it is necessary to add additional constraints derived from correspondences over three or more views to obtain a feasible solution. If such correspondences are available it is also advisable [4] to utilize the trifocal tensor version of the linearized matching constraint for improved performance at lower noise levels. We currently have not implemented this additional constraint.

C. Bundle-Adjustment

It is possible to reformulate the task of projective bundle-adjustment to provide a non-linear solution for the system of cameras ($\mathbf{P}_i$) and points $(\mathbf{X}_j)$ with respect to the parameterization of the camera given in equation (2). Considering as a template the sparse partitioned Levenberg-Marquardt minimization scheme given in [3], we set out to minimize the following objective function for n cameras and m points,

$$\min_{\mathbf{d}_{0,i}, \mathbf{X}^j} \|\mathbf{x}_i^j - \left[\, \mathbf{H}_{0,i}^\pi \mid \mathbf{d}_{0,i} \,\right] \mathbf{X}^j\|^2, \tag{16}$$

where $\hat{\mathbf{x}}_i^j = \left[\, \mathbf{H}_{0,i}^\pi \mid \mathbf{d}_{0,i} \,\right] \mathbf{X}^j$ are the estimated locations of the reprojected image points. The first camera is once again assumed to be fixed ($\mathbf{P}_0 = [\, \mathbf{I} \mid \mathbf{0} \,]$). Consequently the number of parameters involved in the minimization are $3(n-1)+3m$ and the parameter vector $\mathbf{p} = \left[\mathbf{p}_a^\top, \mathbf{p}_b^\top\right]^\top$ is partitioned as $\mathbf{p}_a = [\mathbf{d}_{0,1}^\top, \ldots, \mathbf{d}_{0,n}^\top]^\top$ and $\mathbf{p}_b = [\mathbf{X}^{0\top}, \ldots, \mathbf{X}^{m\top}]^\top$. This results in a simple form for the Jacobian of the camera component of the parameter vector,

$$\frac{(\partial \hat{\mathbf{x}}_i^j)_k}{\mathbf{p}_a} = \begin{bmatrix} \frac{1}{\mathbf{m}_2} \\ \frac{1}{\mathbf{m}_2} \\ -\frac{\mathbf{m}_g}{\mathbf{m}_2^2} \end{bmatrix}, \tag{17}$$

for entry $k = 0, 1$ of the image vector $\hat{\mathbf{x}}_i^j = [\mathbf{m}_0, \mathbf{m}_1, \mathbf{m}_2]^\top$ where $g \neq k$. This results in an efficient non-linear solution to the problem at the cost of assuming that the plane-based homographies are not contributing to the overall error associated with the reconstruction.

III. Experimental Evaluation

In order to assess the performance of the plane-based reconstruction schemes we have conducted a series of experiments on synthetic data in addition to two examples with real images with known ground truth data. The process used to test the algorithms is outlined in Algorithm 1.

The different reconstruction schemes referred to in this Section are the **Direct**, Matching Constraint (**MC**) and Plane-Based Bundle-Adjustment (**PBA**) algorithms discussed in Section II. In addition to these schemes a full projective bundle-adjustment is applied (**FBA**) [3] as a final stage in this process. Using a full bundle-adjustment no longer strictly enforces the plane-based geometry but does absorb some of the error associated with misalignments of the ground-plane homographies.

Also note that preconditioning is applied to the points in each image independently to ensure well conditioned linear systems for the **Direct** and **MC** schemes, this should not be considered optional. The resulting calculations of the cameras must then be deconditioned accordingly.

The assumption when using the plane-based approach is that the ground-plane can be reliably and accurately determined. Consequently less error was added to the ground-plane correspondences in the synthetic experiments and care was taken in the real image experiments to ensure that the ground-plane homographies were calculated with relative accuracy. This also allows for a more meaningful analysis of the impact of noise in the off-plane correspondence data.

Algorithm 1: Plane-Based Reconstruction

Input : Set of images ($\mathbf{I}_0, \ldots, \mathbf{I}_{n-1}$), point sets ($\mathbf{x}_0, \ldots, \mathbf{x}_{n-1}$) and correspondences $\mathbf{x}_i \leftrightarrow \mathbf{x}_{i+1}$ on and off the ground-plane.

Output: Projective cameras $\mathbf{P}_i$, scene points $\mathbf{X}^k$ and ground-plane homographies $\mathbf{H}^{\pi}_{i,j}$

begin
- **for** $\forall\{i, j\}$ *where* $i < j \leq n-1$ **do**
  - **if** *at least* 4 *correspondences* $\mathbf{x}_i \leftrightarrow \mathbf{x}_j$ *exist* **then**
    - (i) Calculate $\mathbf{H}^{\pi}_{i,j}$ using least-squares [3]
    - (ii) Refine $\mathbf{H}^{\pi}_{i,j}$ non-linearly [3]
- I. Condition the image points [9]
- II. Use the **Direct** or **MC** scheme to solve linearly for the cameras ($\mathbf{P}_i$) (the scene points $\mathbf{X}^k$ must be triangulated after reconstruction for **MC**)
- III. Decondition the cameras ($\mathbf{P}_i$)
- IV. Refine the cameras ($\mathbf{P}_i$) and points ($\mathbf{X}^k$) using **PBA**
- V. Refine the cameras ($\mathbf{P}_i$) and points ($\mathbf{X}^k$) using **FBA**

end

A. Synthetic Data

The synthetic experiments were generated from of a series of 11 cameras placed in a circular configuration to ensure correspondences can be determined reliably.

A set of coplanar scene points are generated on the ground-plane of the circle enabling the calculation of the homographies $\mathbf{H}^{\pi}_{i,j}$ and a set of 20 scene points are randomly located within a radius of the circle center and offset from the ground-plane facilitating the calculation of the parallax vector ($\mathbf{d}$).

The scene points were projected into 512×512 images where zero-mean Gaussian noise of varying standard deviations was applied to the ground-plane points (β^2) and the off-plane points (σ^2). A series of 50 trials are completed at each error level using **Direct** and **MC** as linear estimates and **Direct+PBA**, **MC+PBA** and **FBA** as refined estimates.

Figure 1 presents the results of these experiments. It can be seen that all variations of the plane based approach produce stable results in the presence of noise. The Direct reconstruction based methods perform equally well or slightly better than the Matching Constraint methods. The use of both PBA and FBA result in the least sensitivity to error in the ground plane point observations. FBA appears to be able to absorb the effect of error in the ground plane point observations to the extent that it becomes insignificant compared to the effect of error in the off-plane points.

All the correspondences were used for these experiments; however, a further set of experiments were performed (on the same configuration) with an average track length of 3 showing almost identical results. These have been omitted due to space restrictions.

B. Real Images

The first image sequence was acquired from [10] and consists of four 528×384 images with 22 different points manually identified with sub-pixel accuracy in the scene. Figure 2 shows two images of the *PETS* sequence with selected ground plane points labeled in red and off-plane points marked in green. The correspondences related to these points are on average slightly more then three views in length. Ground plane homographies are fitted between all the image pairs with an average fitting error of 0.02 pixels squared.

The second image sequence was acquired of a driveway and consists of four 720×576 images with 20 different points manually identified in the scene. Two images of this sequence are shown in Figure 3, with points marked in a similar fashion to Figure 2. The correspondences related to these points are exactly two views in length. Ground plane homographies are fitted between all the image pairs with an average fitting error of 0.02 pixels squared.

The ground truth error in each case was acquired by calibrating each camera using fiducial markings in the scene, manually identified in the images, then using Tsai's method [1] to determine the camera's internal and external properties.

Table III-B presents the average squared reprojection errors measured in the real images tests. Note that the reduction in error when the Plane base Bundle-Adjustment is applied is much more pronounced than in the synthetic tests. There also exists a more significant difference between the Direct and MC methods; however, the relative performance does not adhere to a simple relationship, but depends on the particular image sequence, as well as the specific schemes employed for reconstruction. The Direct method appears to perform well more consistently than the MC method. In both image sequences, the Direct+PBA+FBA method produced results that were superior to the ground truth produced using Tsai's method.

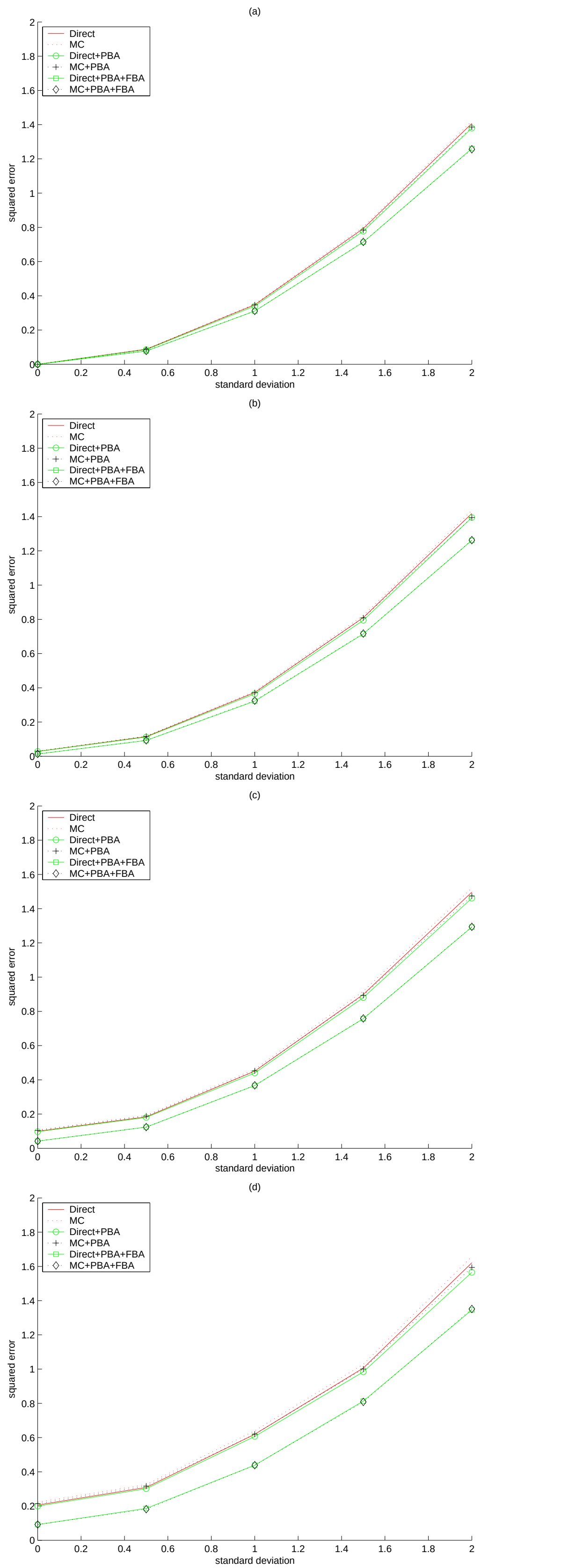

Fig. 1. Average squared reprojection error using a full set of correspondences for $\sigma^2 \in \{0.0, \ldots, 2.0\}$ and (a) $\beta^2 = 0.0$, (b) $\beta^2 = 0.5$, (c) $\beta^2 = 1.0$, (d) $\beta^2 = 1.5$.

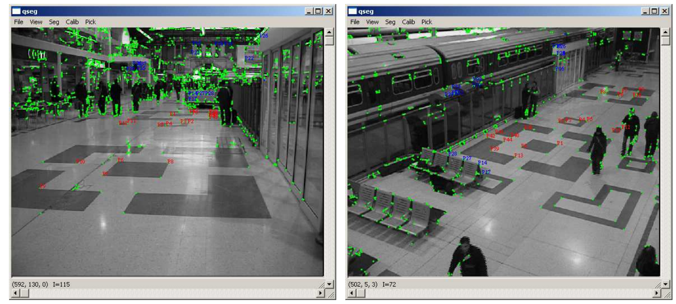

Fig. 2. Images 1 and 4 from the *PETS* data set.

Fig. 3. Images 3 and 4 from the *driveway* data set.

	driveway		*PETS*	
	Direct	**MC**	**Direct**	**MC**
Linear	3.9830	3.6050	5.4604	7.7048
Lin+PBA	2.7466	2.7223	2.3609	2.6521
Lin+PBA+FBA	0.3392	1.0266	0.1180	0.0914
Ground Truth	0.4969		0.1251	

TABLE I
AVERAGE SQUARED REPROJECTION ERROR OF PLANE-BASED RECONSTRUCTION SCHEMES APPLIED TO THE *driveway* AND *PETS* DATASETS.

IV. CONCLUSIONS & FUTURE WORK

In this paper we have implemented a number of algorithms for the task of calibrating a network of security cameras relative to a common ground-plane. The results of the linear schemes applied to real and synthetic data demonstrate the potential to determine a projective reconstruction for a set of cameras utilizing very few correspondences between the images correspondences.

The results from the synthetic experiments indicate that linear plane-based algorithms perform well in the presence of noise, with the error increasing gradually at higher noise levels. The algorithm remains stable with noise added to the ground plane-correspondences in which case the full projective bundle-adjustment seems capable of absorbing most of the error introduced into the plane-based solution.

The experiments performed on real images show that the approach works well for scenes with highly disparate views and in the case of the *driveway* sequence very minimal sequential overlap between the images. The final projective bundle-adjustment in both cases was a necessity in order to absorb some of the error associated with the selection of the ground-plane homographies.

It would be very difficult to reconstruct the *driveway* data set by any other method seeing as sequential structure from motion methods [8] are very sensitive to errors in pair-wise structure initialization where as the plane-based method allows for this initialization to be performed simultaneously across all the cameras (the only alternative to this is sparse factorization [11]). The plane-based method is therefore shown to be a practical approach for this type of scenario.

In future work we plan to utilize the framework for reconstructing networks of cameras relative to the ground-plane to form the basis for a semi-automated system for camera network management. The user of the system would first be required provide the ground-plane correspondences enabling the determination of the plane homographies. This task can be simplified by incorporating the use of wide baseline matching methods [12], [13], [14].

By then incorporating the use of object tracks gathered temporally the estimation of the parallax parameters can be improved incrementally as more track data becomes available. In this respect the resulting external calibration will be most accurate in the regions of the scene where off-plane feature correspondences are available. This in turn conditions the resulting reconstruction to be most accurate for track hand-over purposes.

V. Acknowledgments

This project was supported by the Australian Government Department of the Prime Minister and Cabinet.

References

[1] R. Tsai, “A versatile camera calibration technique for high-accuracy 3-D machine vision metrology using off-the-shelf TV cameras and lenses,” in *Radiometry – (Physics-Based Vision)*, L. Wolff, S. Shafer, and G. Healey, Eds. Jones and Bartlett, 1992.

[2] H. S. Sawhney, “Simplifying motion and structure analysis using planar parallax and image warping,” *In Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp. 709–720, 1994.

[3] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2^{nd} ed. Cambridge University Press, 2003.

[4] R. Kaucic, R. Hartley, and N. Dano, “Plane-based projective reconstruction,” in *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, vol. 1, 2001, pp. 420–427 vol.1.

[5] C. Rother and S. Carlsson, “Linear multi view reconstruction and camera recovery using a reference plane,” *International Journal of Computer Vision*, vol. 49, no. 2, pp. 117–141, 2002.

[6] M. I. A. Lourakis and A. A. Argyros, “Efficient, causal camera tracking in unprepared environments,” *Comput. Vis. Image Underst.*, vol. 99, no. 2, pp. 259–290, 2005.

[7] B. Triggs, “Plane + parallax, tensors, and factorization,” in *The Sixth European Conference on Computer Vision*, 2000, pp. 522–538.

[8] P. A. Beardsley, P. H. S. Torr, and A. Zisserman, “3D model acquisition from extended image sequences,” in *Proceedings of the 4th European Conference on Computer Visionz*, 1996, pp. 683–695.

[9] R. I. Hartley, “In defense of the eight-point algorithm,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 6, pp. 580–593, 1997.

[10] N. I. I. W. on Performance Evaluation of Tracking and Surveillance, “http://www.cvg.rdg.ac.uk/PETS2006/data.html.”

[11] D. Martinec and T. Pajdla, “Structure from many perspective images with occlusions,” in *The seventh European Conference on Computer Vision*, 2002, pp. 355–369.

[12] B. Georgescu and P. Meer, “Point matching under large image deformations and illumination changes,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, pp. 674–688, 2004.

[13] P. Torr and C. Davidson, “Impsac: Synthesis of importance sampling and random sample consensus,” in *The Sixth European Conference on Computer Vision*, 2000, pp. 819–833.

[14] K. Mikolajczyk and C. Schmid, “Scale & affine invariant interest point detectors,” *International Journal of Computer Vision*, vol. 60, no. 1, pp. 63–86, 2004.