Multiple Homography Estimation with Full Consistency Constraints

Wojciech Chojnacki, Zygmunt L. Szpak, Michael J. Brooks, Anton van den Hengel
School of Computer Science
The University of Adelaide
Adelaide, SA 5005, Australia
Email: {wojciech.chojnacki, zygmunt.szpak, michael.brooks, anton.vandenhengel}@adelaide.edu.au

***Abstract*—A novel approach is presented to estimating a set of interdependent homography matrices linked together by latent variables. The approach allows enforcement of all underlying consistency constraints while accounting for the arbitrariness of the scale of each individual matrix. The input data is assumed to be in the form of a set of homography matrices obtained by estimation from image data with the consistency constraints ignored, appended by a set of error covariances associated with these matrix estimates. A cost function is proposed for upgrading, via optimisation, the input data to a set of homography matrices satisfying the constraints. The function is invariant to a change of any of the individual scales of the input matrices. The proposed approach is applied to the particular problem of estimating a set of homography matrices induced by multiple planes in the 3D scene between two views. Experimental results are given which demonstrate the effectiveness of the approach.**

***Keywords*-multiple homographies, consistency constraints, multi-projective parameter estimation, scale invariance, maximum likelihood, covariance.**

I. Introduction

Estimation of a single homography matrix from image measurements is an important step in 3D reconstruction, mosaicing, camera calibration, metric rectification and other tasks [1]. For some applications, like non-rigid motion detection [2], [3], a whole array of homography matrices, all intrinsically interconnected, are required. The matrices have to satisfy consistency constraints representing the rigidness of the motion and the scene. Moreover, the matrices have to be collectively multi-homogeneous—rescaling any individual matrix should not affect the projective information contained in the whole matrix set. A key problem in estimating multiple homography matrices is to enforce the underlying consistency constraints while accounting for the arbitrariness of the individual scales of the matrices.

As a rule, the consistency constraints are available only in implicit form. The conventional approach to cope with such constraints is to evolve a derivative family of explicit constraints. The new constraints are typically more relaxed than the original ones. Adhering to this methodology, Shashua and Avidan [4] have found that homography matrices induced by four or more planes in the 3D scene between two views span a 4-dimensional linear subspace. Chen and Suter [5] have derived a set of strengthened constraints for the case of three or more homographies in two views. Zelnik-Manor and Irani [2] have shown that another rank-four constraint applies to a set of so-called relative homographies generated by two planes between four or more views. These latter authors have also derived constraints for larger sets of homographies and views.

Once isolated, the explicit constraints can be put to use in a procedure whereby first individual homography matrices are estimated from image data, and next these matrices are upgraded to matrices satisfying the constraints. Following this pattern, Shashua and Avidan as well as Zelnik-Manor and Irani used low-rank approximation under the Frobenius norm to enforce the rank-four constraint. Chen and Suter enforced their set of constraints also via low-rank approximation, but then employed the Mahalanobis norm with covariances of input homographies. All these estimation procedures involve input matrices coming with specific scale factors. The underlying error measures are such that a change of scale factors may *a priori* result in a different set of estimates. Furthermore, the output matrices satisfy only the derivative constraints so their perfect consistency is not guaranteed.

This paper presents an alternative approach to estimating interdependent homography matrices, which ensures that *all* implicit constraints are enforced and the final estimates are unaffected by any specific choice of individual scale factors. A cost function is proposed for upgrading, via optimisation, the input set of homography matrices to a set satisfying the constraints. The function is scale change insensitive. To achieve high statistical accuracy, it incorporates the covariances of the input matrices. The utility of the function is demonstrated in a specific application, namely the problem of estimating a set of homography matrices induced by multiple planes in the 3D scene between two views. Experimental results presented in Section VI validate the whole approach.

II. Multi-projective Parameter Estimation

We first formulate the problem of estimating a set of interdependent homographies as a problem in multi-projective parameter estimation [6]. A general multi-projective parameter estimation problem involves a collection $\mathbf{X}_1, \ldots, \mathbf{X}_I$ of $k \times l$ matrices envisaged as data points, and a collection

$\mathbf{\Theta}_1, \ldots, \mathbf{\Theta}_I$ of $k \times l$ matrices treated as parameters. Each $\mathbf{X}_i$ is assumed to be known only up to an individual multiplicative non-zero factor. The $\mathbf{\Theta}_i$'s are subject to constraints and are meant to represent improved versions of the $\mathbf{X}_i$'s. With $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_I)$ denoting the composite datum and $\mathbf{\Theta} = (\mathbf{\Theta}_1, \ldots, \mathbf{\Theta}_I)$ denoting the composite parameter, the problem under consideration is to fit $\mathbf{\Theta}$ to $\mathbf{X}$ so that the constraints on $\mathbf{\Theta}$ are met. Exemplifying this general problem is the following specific problem of interest:

Fit a set of 3×3 matrices, representing planar homographies engendered by various planes in a 3D scene under common projections on two images, to a given set of 3×3 matrices.

To see how the multi-projective framework applies here, suppose that $\mathbf{P}_1 = [\mathbf{I}_3 \mid \mathbf{0}]$ and $\mathbf{P}_2 = [\mathbf{A} \mid -\mathbf{b}]$ are two fixed camera matrices. Suppose, moreover, that a set of I planes in a 3D scene have been selected. Given $i = 1, \ldots, I$, let the i-th plane from the collection be characterised by a length-4 row vector $[\mathbf{v}_i^\mathsf{T}, v_i^0]$. For each $i = 1, \ldots, I$, the i-th plane gives rise to a planar homography $\mathbf{H}_i$ from view $\mathbf{P}_2$ to view $\mathbf{P}_1$ described by the 3×3 matrix

$$\mathbf{H}_i = v_i^0 \mathbf{A} + \mathbf{b}\mathbf{v}_i^\mathsf{T}.$$

Let $\mathbf{H} = (\mathbf{H}_1, \ldots, \mathbf{H}_I)$ be the composite of all the homography matrices in question. With $\mathbf{a} = \mathrm{vec}(\mathbf{A})$, where vec denotes vectorisation, $\boldsymbol{\eta} = [\mathbf{a}^\mathsf{T}, \mathbf{b}^\mathsf{T}, \mathbf{v}_1^\mathsf{T}, v_1^0, \ldots, \mathbf{v}_I^\mathsf{T}, v_I^0]^\mathsf{T}$, and

$$\mathbf{\Pi}(\boldsymbol{\eta}) = (\mathbf{\Pi}_1(\boldsymbol{\eta}), \ldots, \mathbf{\Pi}_I(\boldsymbol{\eta})), \quad \mathbf{\Pi}_i(\boldsymbol{\eta}) = v_i^0 \mathbf{A} + \mathbf{b}\mathbf{v}_i^\mathsf{T}, \quad (1)$$

$\mathbf{H}$ can be represented as

$$\mathbf{H} = \mathbf{\Pi}(\boldsymbol{\eta}). \quad (2)$$

Since $\boldsymbol{\eta}$ has a total of $4I + 12$ entries and since this number is smaller than $9I$ whenever $I \geq 3$, it follows that $\mathbf{H}$ resides in a proper subset of all *a priori* length-I sequences of 3×3 matrices for $I \geq 3$. Thus (2) can be seen a set of intrinsic constraints on $\mathbf{H}$. Suppose that an estimate $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_I)$ of $\mathbf{H}$ has been generated in some way. For example, for each i, $\mathbf{X}_i$ might be an estimate of $\mathbf{H}_i$ individually obtained from image data. The estimation problem at hand is to upgrade $\mathbf{X}$ to $\mathbf{\Theta} = (\mathbf{\Theta}_1, \ldots, \mathbf{\Theta}_I)$ so that $\mathbf{\Theta} = \mathbf{\Pi}(\boldsymbol{\eta})$ holds for some $\boldsymbol{\eta}$ and $\mathbf{\Theta}$ is close to $\mathbf{X}$ in a meaningful sense. The essence here is to find a criterion and effective means for selecting an appropriate $\boldsymbol{\eta}$.

III. AML Cost Function

The general problem of fitting $\mathbf{\Theta}$ to $\mathbf{X}$ with constraints imposed on $\mathbf{\Theta}$ is best considered as an optimisation problem. Since the input matrices are known only up to individual scales, the output matrices should also be determined only to within individual scales. This can be achieved through the use of *multi-homogeneous* cost functions. A cost function J is multi-homogeneous if

$$J(\times \boldsymbol{\lambda} \mathbf{\Theta}) = J(\mathbf{\Theta})$$

for each I-tuple $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_I)$ with non-zero entries, where $\times \boldsymbol{\lambda} \mathbf{\Theta} = (\lambda_1 \mathbf{\Theta}_1, \ldots, \lambda_I \mathbf{\Theta}_I)$. If a cost function is multi-homogeneous, then it is minimised not only at a single $\mathbf{\Theta}$ but also at all composite multiples $\times \boldsymbol{\lambda} \mathbf{\Theta}$.

To describe a multi-homogeneous cost function relevant to our problem, for each $i = 1, \ldots, I$, let

$$\boldsymbol{\theta}_i = \mathrm{vec}(\mathbf{\Theta}_i), \quad \mathbf{x}_i = \mathrm{vec}(\mathbf{X}_i),$$

with each vector having length kl, and let $\mathbf{P}_{\mathbf{x}_i}^\perp$ denote the $kl \times kl$ symmetric projection matrix given by

$$\mathbf{P}_{\mathbf{x}_i}^\perp = \mathbf{I}_{kl} - \|\mathbf{x}_i\|^{-2} \mathbf{x}_i \mathbf{x}_i^\mathsf{T},$$

with $\mathbf{I}_{kl}$ the $kl \times kl$ identity matrix. Referring to the $\mathbf{X}_i$'s via their vectorisations, suppose that associated with each $\mathbf{x}_i$ is a $kl \times kl$ *raw* covariance matrix $\mathbf{\Lambda}_{\mathbf{x}_i}^0$. Any such $\mathbf{\Lambda}_{\mathbf{x}_i}^0$ is meant to carry the bulk of information about the relative importance of the individual entries of $\mathbf{x}_i$ to within a common scale factor. An explicit expression for $\mathbf{\Lambda}_{\mathbf{x}_i}^0$ in one specific case is given in Section VI; see also [7], [8]. Upon upgrading every $\mathbf{\Lambda}_{\mathbf{x}_i}^0$ to a corresponding *corrected* covariance matrix

$$\mathbf{\Lambda}_{\mathbf{x}_i} = \mathbf{P}_{\mathbf{x}_i}^\perp \mathbf{\Lambda}_{\mathbf{x}_i}^0 \mathbf{P}_{\mathbf{x}_i}^\perp,$$

which reflects the fact that the covariance matrix informs about the spread of, specifically, normalised potential versions of $\mathbf{x}_i$, one can define an *approximate maximum likelihood* (AML) cost function by setting

$$J_{\mathrm{AML}}(\mathbf{\Theta}) = \sum_{i=1}^{I} \|\boldsymbol{\theta}_i\|^{-2} \boldsymbol{\theta}_i^\mathsf{T} \mathbf{\Lambda}_{\mathbf{x}_i}^+ \boldsymbol{\theta}_i,$$

where, for a given matrix $\mathbf{A}$, $\mathbf{A}^+$ denotes the Moore–Penrose pseudo-inverse of $\mathbf{A}$. Here, each summand $\|\boldsymbol{\theta}_i\|^{-2} \boldsymbol{\theta}_i^\mathsf{T} \mathbf{\Lambda}_{\mathbf{x}_i}^+ \boldsymbol{\theta}_i$ represents the covariance-weighted, squared sine of the angle between $\boldsymbol{\theta}_i$ and $\mathbf{x}_i$—an entity which is invariant to multiplying $\boldsymbol{\theta}_i$ and $\mathbf{x}_i$ by individual non-zero scalars, and which gives a scale-invariant discrepancy measure between $\boldsymbol{\theta}_i$ and $\mathbf{x}_i$. As it turns out, the expression for $J_{\mathrm{AML}}(\mathbf{\Theta})$ coincides with the squared *Mahalanobis distance* between any aggregate of normalised, arbitrarily signed forms of the $\mathbf{x}_i$'s and any aggregate of similar forms of the $\boldsymbol{\theta}_i$'s. Moreover, this distance is an approximation—to within an additive constant—of a more refined, maximum likelihood-based Mahalanobis distance between image-based data points underpinning the generation of the $\mathbf{x}_i$'s, and respective points on the geometric primitives described by the $\boldsymbol{\theta}_i$'s (cf. [9]–[12]). With the significance of the AML cost function elucidated, when one now takes into consideration the constraints on $\mathbf{\Theta}$, the corresponding constrained minimiser of J_{AML} can be viewed as a statistically well-founded estimate of $\mathbf{\Theta}$.

IV. Cost Function Optimisation

Let J be a cost function for fitting $\boldsymbol{\Theta}$ to $\mathbf{X}$ of the form

$$J(\boldsymbol{\Theta}) = \sum_{i=1}^{I} \|\boldsymbol{\theta}_i\|^{-2} \boldsymbol{\theta}_i^{\mathsf{T}} \mathbf{A}_i \boldsymbol{\theta}_i,$$

where, for each $i = 1, \ldots, I$, $\mathbf{A}_i$ is a $kl \times kl$ non-negative definite matrix. Clearly, the AML cost function conforms to this profile. Suppose that the constraints on $\boldsymbol{\Theta}$ take the form

$$\boldsymbol{\Theta} = \boldsymbol{\Pi}(\boldsymbol{\eta}), \quad \boldsymbol{\Pi}(\boldsymbol{\eta}) = (\boldsymbol{\Pi}_1(\boldsymbol{\eta}), \ldots, \boldsymbol{\Pi}_I(\boldsymbol{\eta})),$$

where $\boldsymbol{\eta}$ is a length-d vector (we have $d = 4I + 12$ in the case of the constraints given in (1)). Upon introducing the function

$$J'(\boldsymbol{\eta}) = J(\boldsymbol{\Pi}(\boldsymbol{\eta})),$$

the constrained optimisation problem in question reduces to that of optimising J', which is an *unconstrained* optimisation problem.

One way of optimising J' is to use the Levenberg-Marquardt (LM) method. The starting point is to re-express J' as

$$J'(\boldsymbol{\eta}) = \sum_{i=1}^{I} \|\mathbf{f}_i'(\boldsymbol{\eta})\|^2,$$

where, for each $i = 1, \ldots, I$,

$$\mathbf{f}_i'(\boldsymbol{\eta}) = \mathbf{f}_i(\boldsymbol{\pi}_i(\boldsymbol{\eta})),$$
$$\mathbf{f}_i(\boldsymbol{\theta}_i) = \|\boldsymbol{\theta}_i\|^{-1} \mathbf{B}_i \boldsymbol{\theta}_i, \quad \boldsymbol{\pi}_i(\boldsymbol{\eta}) = \mathrm{vec}(\boldsymbol{\Pi}_i(\boldsymbol{\eta})),$$

with $\mathbf{B}_i$ a $kl \times kl$ matrix such that $\mathbf{B}_i^{\mathsf{T}} \mathbf{B}_i = \mathbf{A}_i$; in particular, $\mathbf{B}_i$ may be taken equal to the unique non-negative definite square root of $\mathbf{A}_i$. Let $\mathbf{f}'(\boldsymbol{\eta}) = [\mathbf{f}_1'^{\mathsf{T}}(\boldsymbol{\eta}), \ldots, \mathbf{f}_I'^{\mathsf{T}}(\boldsymbol{\eta})]^{\mathsf{T}}$. The LM technique makes use of the $Ikl \times d$ Jacobian matrix $\partial_{\boldsymbol{\eta}} \mathbf{f}'$ represented as $\partial_{\boldsymbol{\eta}} \mathbf{f}' = [\partial_{\boldsymbol{\eta}} \mathbf{f}_1'^{\mathsf{T}} \mid \cdots \mid \partial_{\boldsymbol{\eta}} \mathbf{f}_I'^{\mathsf{T}}]^{\mathsf{T}}$. For each $i = 1, \ldots, I$,

$$\partial_{\boldsymbol{\eta}} \mathbf{f}_i'(\boldsymbol{\eta}) = \partial_{\boldsymbol{\theta}_i} \mathbf{f}_i(\boldsymbol{\pi}_i(\boldsymbol{\eta})) \partial_{\boldsymbol{\eta}} \boldsymbol{\pi}_i(\boldsymbol{\eta})$$

with

$$\partial_{\boldsymbol{\theta}_i} \mathbf{f}_i(\boldsymbol{\theta}_i) = \|\boldsymbol{\theta}_i\|^{-1} \mathbf{B}_i \mathbf{P}_{\boldsymbol{\theta}_i}^{\perp}, \quad \mathbf{P}_{\boldsymbol{\theta}_i}^{\perp} = \mathbf{I}_{kl} - \|\boldsymbol{\theta}_i\|^{-2} \boldsymbol{\theta}_i \boldsymbol{\theta}_i^{\mathsf{T}}.$$

The algorithm iteratively improves on an initial approximation $\boldsymbol{\eta}_0$ to the minimiser of J' by constructing new approximations with the aid of the update rule

$$\boldsymbol{\eta}_{n+1} = \boldsymbol{\eta}_n - [\mathbf{H}(\boldsymbol{\eta}_n) + \lambda_n \mathbf{I}_d)]^{-1} \partial_{\boldsymbol{\eta}} \mathbf{f}'(\boldsymbol{\eta}_n)^{\mathsf{T}} \mathbf{f}'(\boldsymbol{\eta}_n),$$

where $\mathbf{H} = \partial_{\boldsymbol{\eta}} \mathbf{f}'^{\mathsf{T}} \partial_{\boldsymbol{\eta}} \mathbf{f}'$ and λ_n is a non-negative scalar that dynamically changes from step to step [13].

V. Estimating Multiple Homography Matrices

We now specifically consider the LM-based estimation of multiple homography matrices. In this case, for each $i = 1, \ldots, I$,

$$\boldsymbol{\pi}_i(\boldsymbol{\eta}) = \mathrm{vec}(v_i^0 \mathbf{A} + \mathbf{b}\mathbf{v}_i^{\mathsf{T}}) = v_i^0 \mathbf{a} + \mathbf{v}_i \otimes \mathbf{b},$$

where $\otimes$ denotes the Kronecker product. Taking into account that $\mathbf{v}_i \otimes \mathbf{b} = (\mathbf{I}_3 \otimes \mathbf{b})\mathbf{v}_i = (\mathbf{v}_i \otimes \mathbf{I}_3)\mathbf{b}$, one readily verifies that

$$\begin{aligned}
&\partial_{\mathbf{a}} \boldsymbol{\pi}_i = v_i^0 \mathbf{I}_9, && \partial_{\mathbf{b}} \boldsymbol{\pi}_i = \mathbf{v}_i \otimes \mathbf{I}_3, \\
&\partial_{\mathbf{v}_i} \boldsymbol{\pi}_i = \mathbf{I}_3 \otimes \mathbf{b}, && \partial_{\mathbf{v}_j} \boldsymbol{\pi}_j = \mathbf{0} \quad (i \neq j), \\
&\partial_{v_i^0} \boldsymbol{\pi}_i = \mathbf{a}, && \partial_{v_j^0} \boldsymbol{\pi}_j = \mathbf{0} \quad (i \neq j).
\end{aligned}$$

Representing, for each $i = 1, \ldots, I$, $\partial_{\boldsymbol{\eta}} \mathbf{f}_i'$ as

$$\partial_{\boldsymbol{\eta}} \mathbf{f}_i' = [\partial_{\mathbf{a}} \mathbf{f}_i' \mid \partial_{\mathbf{b}} \mathbf{f}_i' \mid \partial_{\mathbf{v}_1} \mathbf{f}_i' \mid \partial_{v_1^0} \mathbf{f}_i' \mid \cdots \mid \partial_{\mathbf{v}_I} \mathbf{f}_i' \mid \partial_{v_I^0} \mathbf{f}_i'],$$

one finds furthermore that

$$\begin{aligned}
\partial_{\mathbf{a}} \mathbf{f}_i' &= v_i^0 \|\boldsymbol{\pi}_i\|^{-1} \mathbf{B}_i \mathbf{P}_{\boldsymbol{\pi}_i}^{\perp}, \\
\partial_{\mathbf{b}} \mathbf{f}_i' &= \|\boldsymbol{\pi}_i\|^{-1} \mathbf{B}_i \mathbf{P}_{\boldsymbol{\pi}_i}^{\perp} (\mathbf{v}_i \otimes \mathbf{I}_3), \\
\partial_{\mathbf{v}_i} \mathbf{f}_i' = \|\boldsymbol{\pi}_i\|^{-1} \mathbf{B}_i \mathbf{P}_{\boldsymbol{\pi}_i}^{\perp} (\mathbf{I}_3 \otimes \mathbf{b}), &\quad \partial_{\mathbf{v}_j} \mathbf{f}_i' = \mathbf{0} \quad (j \neq i), \\
\partial_{v_i^0} \mathbf{f}_i' = \|\boldsymbol{\pi}_i\|^{-1} \mathbf{B}_i \mathbf{P}_{\boldsymbol{\pi}_i}^{\perp} \mathbf{a}, &\quad \partial_{v_j^0} \mathbf{f}_i' = \mathbf{0} \quad (j \neq i).
\end{aligned}$$

With $\partial_{\boldsymbol{\eta}} \mathbf{f}'$ thus determined, all that is now needed is a suitable initialisation for the LM method.

The initialisation scheme adopted here is based on solving the following problem:

Given $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_I)$ satisfying

$$\lambda_i \mathbf{X}_i = \mathbf{H}_i \tag{3}$$

for each $i = 1, \ldots, I$, where λ_i is a non-zero scalar and $\mathbf{H}_i = v_i^0 \mathbf{A} + \mathbf{b}\mathbf{v}_i^{\mathsf{T}}$, solve for $\mathbf{A}$, $\mathbf{b}$, $\mathbf{v}_i$ and v_i^0 in terms of $\mathbf{X}$.

A simple argument shows the solution of this problem is not unique. A seed $\boldsymbol{\eta}_0 = [\mathbf{a}_0^{\mathsf{T}}, \mathbf{b}_0^{\mathsf{T}}, \mathbf{v}_{1,0}^{\mathsf{T}}, v_{1,0}^0, \ldots, \mathbf{v}_{I,0}^{\mathsf{T}}, v_{I,0}^0]^{\mathsf{T}}$ for the LM method is obtained by modifying a specific solution to the above problem. The modification reflects the fact that the original data set $\mathbf{X}$ admits only an approximate representation as in (3). We omit technical details and merely present the steps of the initialisation procedure. These are as follows:

1) For each $i = 1, \ldots, I$, let $v_{i,0}^0 = 1$.
2) Select i_0 arbitrarily from the range between 1 and I.
3) For each $i \neq i_0$, determine two closest eigenvalues $\mu_{ii_0}^{(1)}$ and $\mu_{ii_0}^{(2)}$ of $\mathbf{X}_i^{-1} \mathbf{X}_{i_0}$, and set $\mu_{ii_0} = (\mu_{ii_0}^{(1)} + \mu_{ii_0}^{(2)})/2$.
4) Take for $\mathbf{b}_0$ the left singular vector of the $3 \times 6(I-1)$ matrix obtained by juxtapositioning the matrices $\mu_{ii_0}^{(1)} \mathbf{X}_i - \mathbf{X}_{i_0}$ and $\mu_{ii_0}^{(2)} \mathbf{X}_i - \mathbf{X}_{i_0}$, $i \neq i_{i_0}$, corresponding to the biggest singular value.
5) For each $i \neq i_0$, replace μ_{ii_0} with the real part of μ_{ii_0}. Also replace $\mathbf{b}_0$ with the vector comprising the real parts of the elements of $\mathbf{b}_0$.
6) Let $\mathbf{a}_0 = \mathrm{vec}(\mathbf{X}_{i_0})$ and $\mathbf{v}_{i_0,0} = \mathbf{0}$.
7) For each $i \neq i_{i_0}$, set $\mathbf{v}_{i,0} = \mathbf{v}_{i_0,0} + \|\mathbf{b}_0\|^{-2} (\mu_{ii_0} \mathbf{X}_i - \mathbf{X}_{i_0})^{\mathsf{T}} \mathbf{b}_0$.

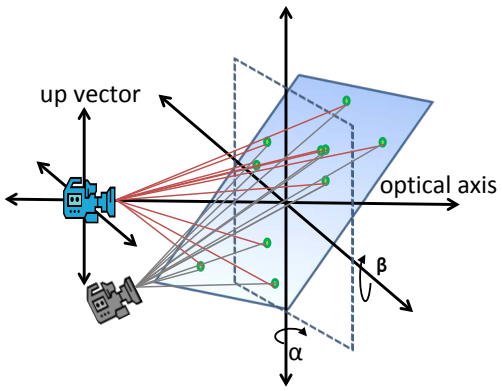


Figure 1. Synthetic data generation procedure.

VI. Experimental Results

The method was tested on both synthetic and real data to determine its performance under varying levels of noise. Repeated experiments were performed in order to collect results of statistical significance.

A. Synthetic Data

The regime adopted was to generate true corresponding points for some stereo configuration and to collect performance statistics over many trials in which random Gaussian noise was added to the image points. Many configurations were investigated and the presented results are typical. Specifically, we conducted experiments by first choosing a realistic geometric configuration for two cameras. Next, we applied a random rotation to a plane that is parallel to the first camera's image plane (see Figure 1). In this manner we generated 4 planes in the 3D scene. Then, 30 points in each plane were randomly selected in the field of view of both cameras, and these were projected onto two 500×500 pixel images to provide "true" image points.

B. Real Data

To investigate the performance of the proposed method on real data, we used the Model House from the Oxford dataset [14]. The dataset consists of 10 different images of a house together with 3D points, projection matrices and 2D point correspondences. We chose the 8th and 9th view of the house and grouped 2D feature points into 4 planar regions in the image: the ground plane as well as the front, side and roof of the house. For each group of 2D points, we determined the corresponding 3D points and fit a plane through them. The 3D points were projected onto this plane, and the planar points were projected onto the image. This ensured that all the feature points in the image belonged to planes in 3D space. Figure 2 depicts the feature points of the Model House used in our experiments.

C. Simulation Procedure

In the case of both synthetic and real data, each image feature point was perturbed by independent homogeneous Gaussian noise at a preset level. For different series of experiments, different noise levels were applied. This resulted in 4 groups of noise-contaminated pairs of corresponding points $\{\mathbf{m}_{n,i}, \mathbf{m}'_{n,i}\}_{n=1}^{N_i}$, $i = 1, \ldots, 4$, corresponding to 4 different planes in the 3D scene.

The estimation methods considered were:

- **DLT** = direct linear transform,
- **FNS** = fundamental numerical scheme,
- **AML** = approximate maximum likelihood,
- **BA** = bundle adjustment.

DLT [1] is a linear method and FNS [9], [15] is an iterative method for estimating a single homography. Both algorithms were run on suitably normalised data. For each $i = 1, \ldots, 4$, two data-dependent normalisation matrices $\mathbf{T}_i$ and $\mathbf{T}'_i$ were applied to $\{\mathbf{m}_{n,i}, \mathbf{m}'_{n,i}\}_{n=1}^{N_i}$ using the rule

$$\tilde{\mathbf{m}}_{n,i} = \mathbf{T}_i \mathbf{m}_{n,i}, \quad \tilde{\mathbf{m}}'_{n,i} = \mathbf{T}'_i \mathbf{m}'_{n,i}$$

to produce individually normalised corresponding points $\{\tilde{\mathbf{m}}_{n,i}, \tilde{\mathbf{m}}'_{n,i}\}_{n=1}^{N_i}$. These normalised groups were used as input to DLT and FNS to produce two sets of 4 homography matrices $(\widehat{\tilde{\mathbf{\Theta}}}_{\mathrm{DLT},i})_{i=1}^4$ and $(\widehat{\tilde{\mathbf{\Theta}}}_{\mathrm{FNS},i})_{i=1}^4$. The final estimates $\widehat{\mathbf{\Theta}}_{\mathrm{DLT}} = (\widehat{\mathbf{\Theta}}_{\mathrm{DLT},i})_{i=1}^4$ and $\widehat{\mathbf{\Theta}}_{\mathrm{FNS}} = (\widehat{\mathbf{\Theta}}_{\mathrm{FNS},i})_{i=1}^4$ were obtained by applying, for each $i = 1, \ldots, 4$, the back-transformation $\mathbf{\Theta}_i \mapsto \mathbf{T}_i'^{-1} \mathbf{\Theta}_i \mathbf{T}_i$ to $\widehat{\tilde{\mathbf{\Theta}}}_{\mathrm{DLT},i}$ and $\widehat{\tilde{\mathbf{\Theta}}}_{\mathrm{FNS},i}$, respectively.

Our proposed AML method used as input the estimates produced by FNS. For reasons of numerical stability, AML required a different data normalisation procedure. Instead of using separate normalisation matrices $\mathbf{T}_i$ and $\mathbf{T}'_i$ for each group of points, we instead combined all the points together to produce two global normalisation matrices $\mathbf{T}$ and $\mathbf{T}'$. These matrices were used to produce globally normalised corresponding points $\{\mathring{\mathbf{m}}_{n,i}, \mathring{\mathbf{m}}'_{n,i}\}_{n=1}^{N_i}$, $i = 1, \ldots, 4$, defined by

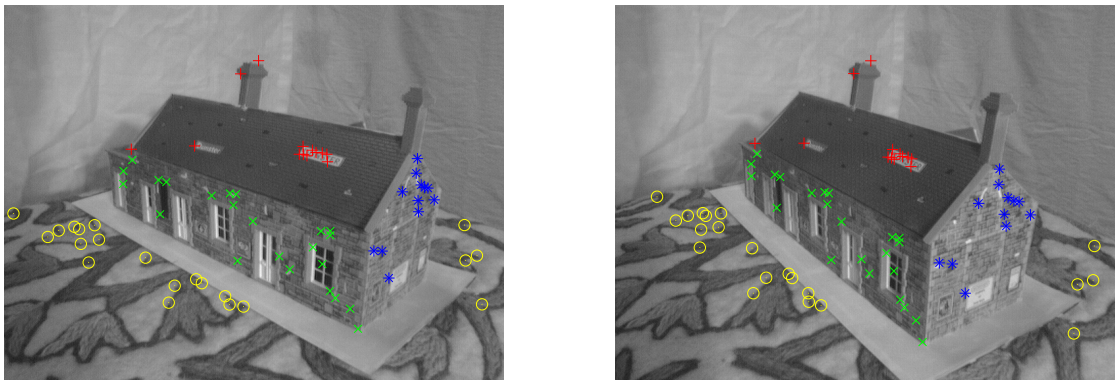
$$\mathring{\mathbf{m}}_{n,i} = \mathbf{T}\mathbf{m}_{n,i}, \quad \mathring{\mathbf{m}}'_{n,i} = \mathbf{T}'\mathbf{m}'_{n,i}.$$

The input of AML took the form of the FNS estimates transferred to the $(\mathring{\mathbf{m}}, \mathring{\mathbf{m}}')$-coordinate system via the rule $\mathbf{X}_i = \mathbf{T}' \widehat{\mathbf{\Theta}}_{\mathrm{FNS},i} \mathbf{T}^{-1}$. The raw covariance matrix for the vectorisation $\mathbf{x}_i$ of $\mathbf{X}_i$, based on $\{\mathring{\mathbf{m}}_{n,i}, \mathring{\mathbf{m}}'_{n,i}\}_{n=1}^{N_i}$, was taken to be

$$\mathbf{\Lambda}^0_{\mathbf{x}_i} = (\mathbf{M}_{\mathbf{x}_i})^+_8,$$

$$\mathbf{M}_{\mathbf{x}_i} = \|\mathbf{x}_i\|^2 \sum_{n=1}^{N_i} \mathbf{U}(\mathring{\mathbf{z}}_{n,i})(\mathbf{\Sigma}(\mathring{\mathbf{z}}_{n,i}, \mathbf{x}_i))^+_2 \mathbf{U}(\mathring{\mathbf{z}}_{n,i})^\mathsf{T}.$$

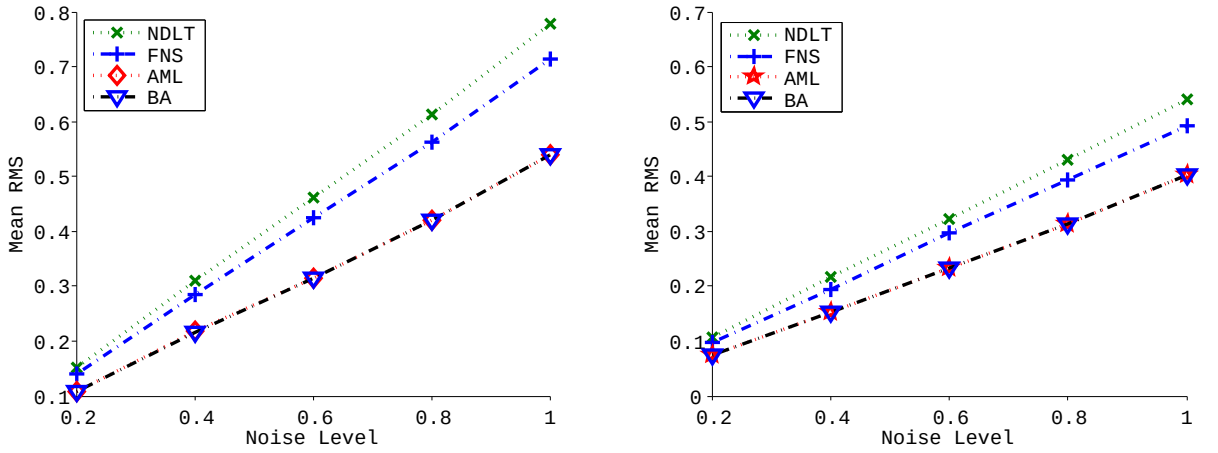
Here $\mathring{\mathbf{z}}_{n,i} = [\mathring{u}_{n,i}, \mathring{v}_{n,i}, \mathring{u}'_{n,i}, \mathring{v}'_{n,i}]^\mathsf{T}$ is the result of combining $\mathring{\mathbf{m}}_{n,i} = [\mathring{u}_{n,i}, \mathring{v}_{n,i}, 1]^\mathsf{T}$ and $\mathring{\mathbf{m}}'_{n,i} = [\mathring{u}'_{n,i}, \mathring{v}'_{n,i}, 1]^\mathsf{T}$ into a



(a) Feature points in view 8.

(b) Feature points in view 9.

Figure 2. Feature points in 4 planar regions of a model house used for real data simulation.



(a) Symmetric transfer error from ground truth on model house data.

(b) Symmetric transfer error from ground truth on synthetic data.

Figure 3. Comparison of methods on real and synthetic data (200 simulations).

single vector; $\mathbf{U}(\mathring{\mathbf{z}}_{n,i})$ and $\mathbf{\Sigma}(\mathring{\mathbf{z}}_{n,i}, \mathbf{x}_i)$ are defined by

$$\begin{aligned}\mathbf{U}(\mathbf{z}) &= -\mathbf{m} \otimes [\mathbf{m}']_\times, \\ \mathbf{B}(\mathbf{z}) &= \partial_{\mathbf{z}}\mathrm{vec}(\mathbf{U}(\mathbf{z}))\mathbf{\Lambda}_{\mathbf{z}}\big[\partial_{\mathbf{z}}\mathrm{vec}(\mathbf{U}(\mathbf{z}))\big]^\mathsf{T}, \\ \mathbf{\Sigma}(\mathbf{z}, \mathbf{x}) &= (\mathbf{I}_3 \otimes \mathbf{x}^\mathsf{T})\mathbf{B}(\mathbf{z})(\mathbf{I}_3 \otimes \mathbf{x})\end{aligned}$$

with $\mathbf{z} = [u, v, u', v']^\mathsf{T}$ derived from $\mathbf{m} = [u, v, 1]^\mathsf{T}$ and $\mathbf{m}' = [u', v', 1]^\mathsf{T}$, and with $\mathbf{x}$ a length-9 vector; given a length-3 vector $\mathbf{a}$, $[\mathbf{a}]_\times$ is the 3×3 antisymmetric matrix such that $\mathbf{a} \times \mathbf{y} = [\mathbf{a}]_\times \mathbf{y}$ for each length-3 vector $\mathbf{y}$; and $\mathbf{A}_r^+$ denotes the *rank-r truncated pseudo-inverse* of the matrix $\mathbf{A}$. The image data covariance matrices $\mathbf{\Lambda}_{\mathring{\mathbf{z}}_{n,i}}$, incorporated in the matrices $\mathbf{\Sigma}(\mathring{\mathbf{z}}_{n,i}, \mathbf{x}_i)$, were all chosen to be in their default form $\mathrm{diag}(1, 1, 1, 1)$, corresponding to isotropic homogeneous noise in image point measurement. The matrices $\mathbf{X}_i$ and covariances $\mathbf{\Lambda}^0_{\mathbf{x}_i}$ were next used as input to the AML method to produce estimates which, upon applying the back-transformation $\mathbf{\Theta} \mapsto \mathbf{T}'^{-1}\mathbf{\Theta}\mathbf{T}$, were taken to be $\widehat{\mathbf{\Theta}}_{\mathrm{AML}} = (\widehat{\mathbf{\Theta}}_{\mathrm{AML},i})_{i=1}^4$.

A BA estimate $\widehat{\mathbf{\Theta}}_{\mathrm{BA}} = (\widehat{\mathbf{\Theta}}_{\mathrm{BA},i})_{i=1}^4$ was generated directly from all the image data $\{\mathring{\mathbf{m}}_{n,i}, \mathring{\mathbf{m}}'_{n,i} \mid n = 1, \ldots, N_i,\ i = 1, \ldots, 4\}$ by minimising the error

$$\sum_{i=1}^{4}\sum_{n=1}^{N_i}\left(d(\mathring{\mathbf{m}}_{n,i}, \mathbf{\Pi}_i(\boldsymbol{\eta})^{-1}\underline{\mathbf{m}}'_{n,i})^2 + d(\mathring{\mathbf{m}}'_{n,i}, \mathbf{\Pi}_i(\boldsymbol{\eta})\underline{\mathbf{m}}_{n,i})^2\right)$$

over all 2D points $\underline{\mathbf{m}}_{n,i}$ and $\underline{\mathbf{m}}'_{n,i}$ and all parameter vectors $\boldsymbol{\eta}$. Here $d(\mathbf{m}, \mathbf{n})$ denotes the Euclidean distance between the points $\mathbf{m}$ and $\mathbf{n}$ dehomogenised at the last, third entry. The initial value of each $\underline{\mathbf{m}}_{n,i}$ and each $\underline{\mathbf{m}}'_{n,i}$ was taken to be $\mathring{\mathbf{m}}_{n,i}$ and $\mathring{\mathbf{m}}'_{n,i}$, respectively, and the initial value of $\boldsymbol{\eta}$ was obtained from the result of AML. Then the $\underline{\mathbf{m}}_{n,i}$'s, $\underline{\mathbf{m}}'_{n,i}$'s, and $\boldsymbol{\eta}$ were recomputed iteratively by an LM scheme adapted to the task of minimising the error given above. With $\widehat{\mathbf{A}}_{\mathrm{BA}}$, $\widehat{\mathbf{b}}_{\mathrm{BA}}$, $\widehat{\mathbf{v}}_{\mathrm{BA},i}$'s, $\widehat{v}^0_{\mathrm{BA},i}$'s derived from the terminal value of $\boldsymbol{\eta}$, the estimates $\widehat{\mathbf{\Theta}}_{\mathrm{BA},i}$ were finally obtained by applying

the back-transformation $\boldsymbol{\Theta} \mapsto \mathbf{T}'^{-1}\boldsymbol{\Theta}\mathbf{T}$ to $\widehat{v}^0_{\mathrm{BA},i}\widehat{\mathbf{A}}_{\mathrm{BA}} + \widehat{\mathbf{b}}_{\mathrm{BA}}\widehat{\mathbf{v}}^{\mathsf{T}}_{\mathrm{BA},i}$.

D. Comparison of Methods

For $\widehat{\boldsymbol{\Theta}} = (\widehat{\boldsymbol{\Theta}}_i)_{i=1}^4$ generated by the DLT, FNS, AML, and BA methods, the common distance used to quantify data-model discrepancies was the mean root-mean-square (RMS) *symmetric transfer error* from noiseless data points $\overline{\mathbf{m}}_{n,i}$ and $\overline{\mathbf{m}}'_{n,i}$

$$\frac{1}{4}\sum_{i=1}^{4}\left(\frac{1}{4N_i}\sum_{n=1}^{N_i}\left(d(\overline{\mathbf{m}}_{n,i}, \widehat{\boldsymbol{\Theta}}_i^{-1}\overline{\mathbf{m}}'_{n,i})^2 + d(\overline{\mathbf{m}}'_{n,i}, \widehat{\boldsymbol{\Theta}}_i\overline{\mathbf{m}}_{n,i})^2\right)\right)^{1/2}.$$

A comparison of the methods on real and synthetic data is shown in Figures 3a and 3b, respectively. The results indicate that the proposed algorithm outperformed DLT and FNS, and achieved a very close approximation to the BA method.

VII. Discussion

The experiments suggest that our method of enforcing consistency on the individually estimated homographies has two clear advantages, namely a significant improvement in the quality of the estimated homographies and the possibility of interpreting the relative positions of the planes generating the homographies. A surprising outcome of the experiments is that the bundle adjustment method resulted in only a marginal improvement compared to our proposed algorithm. In future work we plan to explore further alternative initialisations for bundle adjustment, to determine if AML is indeed as good an approximation of the true homographies as our experiments indicate.

VIII. Conclusion

This paper presents a novel approach to estimating a set of interdependent homography matrices. The consistency across the matrix set is achieved via implicit constraints put on every candidate matrix set. The data is assumed to be in the form of matrices not necessarily satisfying the underlying constraints and appended by error covariances associated with these matrices. A particular approximate maximum likelihood cost function is proposed for upgrading the input matrices to matrices satisfying the constraints. This function is invariant to possible changes of the individual scales of the input matrices. The scale invariance property is an essential element of design, differentiating the introduced function from the cost functions used earlier for multiple homography estimation. The approach is tested on the problem of estimating a set of homography matrices induced by a multiple planes in the 3D scene between two views. The Levenberg-Marquardt algorithm evolved to optimise the proposed cost function produces results demonstrating that the approach is feasible and efficient.

Acknowledgement

This research was partially supported by the Australian Research Council.

References

[1] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge: Cambridge University Press, 2004.

[2] L. Zelnik-Manor and M. Irani, "Multiview constraints on homographies," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 2, pp. 214–223, 2002.

[3] O. Kähler and J. Denzler, "Rigid motion constraints for tracking planar objects," in *Proc. 29th DAGM Symposium*, ser. Lecture Notes in Computer Science, vol. 4713, 2007, pp. 102–111.

[4] A. Shashua and S. Avidan, "The rank 4 constraint in multiple (≥ 3) view geometry," in *Proc. 4th European Conf. Computer Vision*, ser. Lecture Notes in Computer Vision, vol. 1065, 1996, pp. 196–206.

[5] P. Chen and D. Suter, "Rank constraints for homographies over two views: revisiting the rank four constraint," *Int. J. Computer Vision*, vol. 81, no. 2, pp. 205–225, 2009.

[6] W. Chojnacki, R. Hill, A. van den Hengel, and M. J. Brooks, "Multi-projective parameter estimation for sets of homogeneous matrices," in *Proc. Digital Image Computing: Techniques and Applications Conf.*, 2009, pp. 119–124.

[7] K. Kanatani, N. Ohta, and Y. Kanazawa, "Optimal homography computation with a reliability measure," *IEICE Trans. Inf. Syst.*, vol. E83-D, no. 7, pp. 1369–1374, 2000.

[8] P. Chen and D. Suter, "Error analysis in homography estimation by first order approximation tools: a general technique," *J. Math. Imaging Vision*, vol. 33, no. 3, pp. 281–295, 2009.

[9] W. Chojnacki, M. J. Brooks, A. van den Hengel, and D. Gawley, "On the fitting of surfaces to data with covariances," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 11, pp. 1294–1303, 2000.

[10] ——, "FNS, CFNS and HEIV: A unifying approach," *J. Math. Imaging Vision*, vol. 23, no. 2, pp. 175–183, 2005.

[11] K. Kanatani, *Statistical Optimization for Geometric Computation: Theory and Practice*. Amsterdam: Elsevier, 1996.

[12] B. Matei and P. Meer, "A general method for errors-in-variables problems in computer vision," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 2, 2000, pp. 18–25.

[13] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C*. Cambridge: Cambridge University Press, 1995.

[14] [Online]. Available: http://www.robots.ox.ac.uk/~vgg/data.html

[15] T. Scoleri, W. Chojnacki, and M. J. Brooks, "A multi-objective parameter estimator for image mosaicing," in *Proc. IEEE Int. Symposium Signal Processing and its Applications*, vol. 2, 2005, pp. 551–554.