Sampson Distance Based Joint Estimation of Multiple Homographies with Uncalibrated Cameras

Zygmunt L. Szpak*, Wojciech Chojnacki, Anders Eriksson, Anton van den Hengel

School of Computer Science, The University of Adelaide, SA 5005, Australia

Abstract

Two images of a scene consisting of multiple flat surfaces are related by a collection of homography matrices. Practitioners typically estimate these homographies separately thereby violating inherent inter-homography constraints that arise naturally out of the rigid geometry of the scene. We demonstrate that through a suitable choice of parametrisation multiple homographies can be jointly estimated in a manner so as to satisfy all inter-homography constraints. Unlike the cost functions used previously for solving this problem, our cost function does not correspond to fitting one set of homography matrices to another set of homography matrices. Instead, we utilise the Sampson distance for homography matrix estimation and operate directly on image data points. By using the Sampson distance and working directly on data points, we expedite the application of a vast amount of knowledge that already exists for Sampson-distance-based single homography or fundamental matrix estimation. The estimation framework reported in this paper establishes a new baseline for joint multiple homography estimation and at the same time raises intriguing new research questions. The work may be of interest to a broad range of researchers who require the estimation of homography matrices with uncalibrated cameras as part of their solution.

Keywords: multiple homographies, parameter estimation, maximum likelihood, Sampson distance, latent variables

1. Introduction

Whether one is planning to autonomously land a spacecraft on a distant planet, generate three-dimensional maps of the earth, or augment the environment with virtual characters, one will need to understand how different images of the same scene are related. A fundamental result in multiple view geometry states that if a scene consists of a flat surface, then the coordinates of any pair of corresponding points in two images are related by a *homography matrix* [11]. If a scene consists of multiple flat surfaces, such as buildings in an urban environment, then multiple homography matrices (one for each plane) are needed to adequately relate corresponding points in two images (see Figure 1).

When multiple homography matrices are required, practitioners typically estimate them separately [12]. This paper argues that multiple homography matrices should not be estimated separately, but rather should be estimated jointly. Estimating homography matrices separately fails to ensure that the matrices satisfy inter-homography constraints that arise naturally due to the rigidity of the planar scene. Through extensive experiments we show that estimating homography matrices jointly so that they satisfy all inter-homography constraints leads to considerably more accurate results.

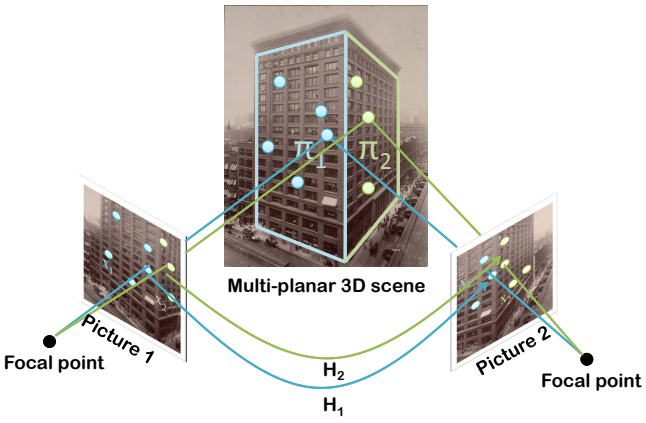


Figure 1: Relationship between homography matrices and planes in three-dimensional space. Homography matrices encapsulate the relationship between the coordinates of images taken from two different views, provided that the scene consists of flat surfaces. For example, the two sides of the building can be represented by two planes, π_1 and π_2. The two planes give rise to two homography matrices, $\mathbf{H}_1$ and $\mathbf{H}_2$. The image coordinates of the side of the building associated with π_i in one view can be mapped with the aid of $\mathbf{H}_i$ to the image coordinates of the same side in the other view. In particular, given the coordinates of a window located on π_i in the first view, one can determine—by exploiting $\mathbf{H}_i$—the coordinates of the same window in the second view.

2. Background

Almost two decades ago Sashua and Avidan [16] reported that a collection of homography matrices induced by multiple planes between two views are inter-related. In particular,

*Corresponding author. Tel.: +61-8-8313-2132; fax: +61-8-8313-4366

Email addresses: zygmunt.szpak@adelaide.edu.au (Zygmunt L. Szpak), wojciech.chojnacki@adelaide.edu.au (Wojciech Chojnacki), anders.eriksson@adelaide.edu.au (Anders Eriksson), anton.vandenhengel@adelaide.edu.au (Anton van den Hengel)

these authors revealed that, for a pair of images, a collection of five or more homography matrices resides in an at most four-dimensional subspace of the parameter space. The corresponding subspace constraint became known as the rank-four constraint. Zelnik-Manor and Irani [17] took Sashua and Avidan's line of inquiry further and showed that homography rank constraints also exist that apply to a whole sequence of images. Both Sashua and Avidan as well as Zelnik-Manor and Irani advocated a straightforward procedure for enforcing rank constraints on a collection of homographies. The essence of their method is to form a composite matrix from vectorised versions of the homography matrices and to enforce the rank constraint using a singular value decomposition. The rank enforcement procedure can be understood as fitting a new set of homography matrices (that reside in a four-dimensional subspace) to an existing set of homography matrices using the Frobenius norm as a measure of distance.

Despite its simplicity the rank enforcement method has not been widely adopted, one reason for this being that it is difficult to detect interest points on more than five planar surfaces in a typical image. Subsequent studies also reported that in many instances the rank constraints were not producing any notable improvements in the accuracy of the underlying homographies. It was believed that the lack of consistent denoising may be due to the fact that singular value decomposition does not take homography covariance information into account.

Chen and Suter [2] extended the work of Sashua and Avidan by proposing an iterative weighted alternating least squares scheme for upgrading separately estimated homography matrices to consistent matrices. By incorporating homography covariance information into the cost function, Chen and Suter intended to surpass the limitations of rank constraint enforcement. Their iterative scheme needed a suitable initialisation and they were able to devise one that required three or more homographies. Regrettably, their cost function was deficient from a theoretical point of view in that it was not scale invariant. Homography matrices are defined only up to an unknown scale factor, yet multiplying the homographies by different scales could affect the solution and convergence of Chen and Suter's iterative scheme. Nevertheless, Chen and Suter's contribution was a valuable one, providing a benchmark against which new methods could be compared.

Another alternative to direct rank-four enforcement was reported by Eriksson and van den Hengel [8]. Like Chen and Suter, these authors also utilised homography covariance information. Unfortunately, they never devised a method to initialise their iterative scheme. Instead, they resorted to random initialisations which occasionally converged to poor solutions.

To overcome the limitations of the aforementioned methods, Chojnacki et al. [7] proposed a homography-covariance-based *scale invariant* cost function. They showed that by a suitable choice of latent variables *all* inter-homography constraints can be satisfied. They used the popular Levenberg–Marquardt scheme to optimise their cost function and also presented an algorithm that can be used to initialise the latent variables when there are two or more homography matrices. This meant that for the first time *dependable* improvements to homography matrix estimates could be observed in the minimal case of two homographies.

3. Contribution

All of the methods that we have reviewed fit one set of homography matrices to another set of homography matrices. In this paper, we propose a departure from what has been attempted before and offer a method for jointly estimating multiple homographies that operates directly on the image points using the Sampson distance. We do not assume that the cameras are calibrated. To enforce all inter-homography constraints, we utilise the same parametrisation as proposed by Chojnacki et al. [7].

Our experiments will demonstrate the utility and superiority of our method, and, in the process, will open up new avenues for compelling future research. Our hope is that practitioners will see the benefit of estimating multiple homographies jointly, and will move away from rather inappropriate separate homography estimation.

4. Estimation problem

We consider the following task:

Problem. Estimate a collection of 3×3 matrices, representing planar homographies engendered by various planes in a 3D scene under common projections on two images, based on a collection of sets of image correspondences, each set being related to a separate homography.

A fundamental aspect of the above problem is that the homography matrices involved are interdependent. To get insight into the nature of the dependencies involved, consider two fixed uncalibrated cameras giving rise to two camera matrices $\mathbf{P}_1 = \mathbf{K}_1\mathbf{R}_1[\mathbf{I}_3, -\mathbf{t}_1]$ and $\mathbf{P}_2 = \mathbf{K}_2\mathbf{R}_2[\mathbf{I}_3, -\mathbf{t}_2]$. Here the length-3 translation vector $\mathbf{t}_k$ and the 3×3 rotation matrix $\mathbf{R}_k$ represent the Euclidean transformation between the k-th ($k = 1, 2$) camera and the world coordinate system, $\mathbf{K}_k$ is a 3×3 upper triangular calibration matrix encoding the internal parameters of the k-th camera, and $\mathbf{I}_3$ denotes the 3×3 identity matrix. 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 have a unit outward normal $\mathbf{n}_i$ and be situated at a distance d_i from the origin of the world coordinate system. Then, for each $i = 1, \ldots, I$, the i-th plane gives rise to a planar homography from view $\mathbf{P}_2$ to $\mathbf{P}_1$ described by the 3×3 matrix

$$\mathbf{H}_i = w_i\mathbf{A} + \mathbf{b}\mathbf{v}_i^\top, \tag{1}$$

where

$$\begin{aligned} \mathbf{A} &= \mathbf{K}_2\mathbf{R}_2\mathbf{R}_1^{-1}\mathbf{K}_1^{-1}, \quad w_i = \mathbf{n}_i^\top\mathbf{t}_1 - d_i, \\ \mathbf{b} &= \mathbf{K}_2\mathbf{R}_2(\mathbf{t}_1 - \mathbf{t}_2), \quad \mathbf{v}_i = \mathbf{K}_1^{-\top}\mathbf{R}_1^{-\top}\mathbf{n}_i. \end{aligned} \tag{2}$$

We note that in the case of calibrated cameras when one may assume that $\mathbf{K}_1 = \mathbf{K}_2 = \mathbf{I}_3$, $\mathbf{t}_1 = \mathbf{0}$, $\mathbf{R}_1 = \mathbf{I}_3$, $\mathbf{R}_2 = \mathbf{R}$, system (2) reduces to

$$\begin{aligned} \mathbf{A} &= \mathbf{R}, \quad w_i = -d_i, \\ \mathbf{b} &= \mathbf{t}, \quad \mathbf{v}_i = \mathbf{n}_i, \end{aligned} \tag{3}$$

with $\mathbf{t} = -\mathbf{R}\mathbf{t}_2$, and equality (1) becomes the familiar *direct nRt* representation

$$\mathbf{H}_i = -d_i\mathbf{R} + \mathbf{t}\mathbf{n}_i^\top$$

(cf. [1, 14]). We stress that all of our subsequent analysis concerns the general uncalibrated case, with $\mathbf{A}$, $\mathbf{b}$, w_i's and $\mathbf{v}_i$'s to be interpreted according to (2) rather than (3).

Let $\mathsf{H} = [\mathbf{H}_1, \ldots, \mathbf{H}_I]$ be the composite of all the homography matrices in question and let $\mathbf{H}$ be the $9 \times I$ matrix given by

$$\mathbf{H} = [\mathbf{h}_1, \ldots \mathbf{h}_I], \quad \mathbf{h}_i = \mathrm{vec}(\mathbf{H}_i),$$

where vec denotes column-wise vectorisation [13]. With $\mathbf{a} = \mathrm{vec}(\mathbf{A})$ and $\otimes$ denoting Kronecker product [13], we have

$$\mathbf{h}_i = w_i\,\mathrm{vec}(\mathbf{A}) + \mathrm{vec}(\mathbf{b}\mathbf{v}_i^\top) = w_i\mathbf{a} + (\mathbf{I}_3 \otimes \mathbf{b})\mathbf{v}_i \tag{4}$$

for each $i = 1, \ldots, I$. This can be concisely written as the matrix equality

$$\mathbf{H} = \mathbf{S}\mathbf{T}, \tag{5}$$

where $\mathbf{S} = [\mathbf{I}_3 \otimes \mathbf{b}, \mathbf{a}]$ is a 9×4 matrix and

$$\mathbf{T} = \begin{bmatrix} \mathbf{v}_1 & \ldots & \mathbf{v}_I \\ w_1 & \ldots & w_I \end{bmatrix}$$

is a $4 \times I$ matrix. An immediate consequence of the representation (5) is that $\mathbf{H}$ has rank at most four. Whenever $I \geq 5$ the requirement that $\mathbf{H}$ should have rank no greater than four places a genuine constraint on $\mathbf{H}$, and hence also on H. This constraint is the *rank-four constraint* of Shashua and Avidan mentioned in the Introduction. The rank-four constraint stipulates that the dimension of the set of all H's is no greater than $4I + 20$ for $I \geq 5$. The ensuing inequality $4I + 20 < 9I$ for $I \geq 5$ makes it clear that H resides in a proper subset of all *a priori* length-I composites of 3×3 matrices for $I \geq 5$.

The dimensionality count for the H's can be refined and the subsequent conclusions sharpened. Letting $\boldsymbol{\eta} = [\mathbf{a}^\top, \mathbf{b}^\top, \mathbf{v}_1^\top, \ldots, \mathbf{v}_I^\top, w_1, \ldots, w_I]^\top$ and

$$\boldsymbol{\Pi}(\boldsymbol{\eta}) = [\boldsymbol{\Pi}_1(\boldsymbol{\eta}), \ldots, \boldsymbol{\Pi}_I(\boldsymbol{\eta})], \quad \boldsymbol{\Pi}_i(\boldsymbol{\eta}) = w_i\mathbf{A} + \mathbf{b}\mathbf{v}_i^\top, \tag{6}$$

H can be represented as

$$\mathsf{H} = \boldsymbol{\Pi}(\boldsymbol{\eta}). \tag{7}$$

In this formulation, $\boldsymbol{\eta}$ appears as the vector of *latent variables* that link all the constituent matrices together and provide a natural parametrisation of the set of all H's. Since $\boldsymbol{\eta}$ has a total of $4I + 12$ entries, the set of all matrices of the form $\boldsymbol{\Pi}(\boldsymbol{\eta})$ has dimension no greater than $4I + 12$. A more refined argument shows that the set of all $\boldsymbol{\Pi}(\boldsymbol{\eta})$'s has in fact dimension equal to $4I + 7$ [5, 6]. Since $4I + 7 < 9I$ whenever $I \geq 2$, it follows that H resides in a proper subset of all length-I composites of 3×3 matrices for $I \geq 2$. This is an improvement over the previous similar statement which was valid only for $I \geq 5$. It is now clear that the requirement that H take the form as per (7) whenever $I \geq 2$ can be seen as an implicit constraint on H, with the consequence that the $\mathbf{H}_i$'s are all interdependent.

In what follows we shall exploit the parametrisation (7) as a handle on the inter-dependencies in the matrices $\mathbf{H}_i$. It is worth stressing that this parametrisation is not minimal in that if $\mathsf{H} = \boldsymbol{\Pi}(\boldsymbol{\eta})$ holds for a particular vector $\boldsymbol{\eta}$, then we also have $\mathsf{H} = \boldsymbol{\Pi}(\boldsymbol{\eta}')$ for any vector $\boldsymbol{\eta}'$ of the form

$$\boldsymbol{\eta}' = [\mathrm{vec}(\beta\mathbf{A} + \mathbf{b}\mathbf{c}^\top), \alpha\mathbf{b}, \alpha^{-1}\mathbf{v}_1 - \alpha^{-1}\beta^{-1}\mathbf{c}, \ldots, \alpha^{-1}\mathbf{v}_I - \alpha^{-1}\beta^{-1}\mathbf{c}, \beta^{-1}w_1, \ldots, \beta^{-1}w_I]^\top,$$

where α and β are non-zero numbers, and $\mathbf{c}$ is a length-3 vector. The arbitrariness in the choice of α, β, and $\mathbf{c}$ reflects five degrees of parametrisation gauge freedom. In particular, if $\mathsf{H} = \boldsymbol{\Pi}(\boldsymbol{\eta})$ holds for a particular vector $\boldsymbol{\eta}$, then it is not true that the entries $\boldsymbol{\eta}$ have to be necessarily interpreted as ones satisfying (2) for the underlying values of $\mathbf{K}_k$, $\mathbf{R}_k$, $\mathbf{t}_k$, d_i, and $\mathbf{n}_i$ ($k = 1, 2$, $i = 1, \ldots, I$). It is also worth stressing that the fact that the parametrisation (7) is not minimal is not a limitation—on the contrary, the parametrisation is perfectly suited to our purpose, as it enforces all underlying constraints on the $\mathbf{H}_i$'s.

5. Background on the homography model

If a planar homography is represented by an invertible 3×3 matrix $\mathbf{H}$ and if $\mathbf{m}' = [u', v', 1]^\top$ is the image of $\mathbf{m} = [u, v, 1]^\top$ by that homography, then $\mathbf{m}' \simeq \mathbf{H}\mathbf{m}$, where $\simeq$ denotes equality up to scale. This relation can equivalently be written as

$$[\mathbf{m}']_\times\mathbf{H}\mathbf{m} = \mathbf{0}, \tag{8}$$

where, for a length-3 vector $\mathbf{a} = [a_1, a_2, a_3]^\top$, $[\mathbf{a}]_\times$ denotes the 3×3 anti-symmetric matrix given by

$$[\mathbf{a}]_\times = \begin{bmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{bmatrix}.$$

With $\theta = \mathrm{vec}(\mathbf{H})$, $\mathbf{z} = [u, v, u', v']^\top$, and $\mathbf{U}(\mathbf{z}) = -\mathbf{m} \otimes [\mathbf{m}']_\times$, we have $[\mathbf{m}']_\times\mathbf{H}\mathbf{m} = \mathbf{U}(\mathbf{z})^\top\theta$, and so (8) can be restated as

$$\mathbf{U}(\mathbf{z})^\top\theta = \mathbf{0}. \tag{9}$$

Since the 9×3 matrix $\mathbf{U}(\mathbf{z})$ has rank 2, the three equations in (9) are linearly dependent and can be reduced—by deleting one of them—to a system of two equations. By leaving specifically the first two equations, the resulting system becomes

$$\mathbf{V}(\mathbf{z})^\top\theta = \mathbf{0},$$

where $\mathbf{V}(\mathbf{z}) = \mathbf{U}(\mathbf{z})\mathbf{I}_{3\times 2}$, $\mathbf{I}_{3\times 2} = [\mathbf{e}_1, \mathbf{e}_2]$, $\mathbf{e}_1 = [1, 0, 0]^\top$, and $\mathbf{e}_2 = [0, 1, 0]^\top$. This system of equations forms the foundation of an approximate maximum likelihood cost function which we describe in the next section.

6. The Sampson distance based AML cost function

Let $\{\{\mathbf{m}_{ij}, \mathbf{m}'_{ij}\}_{j=1}^{J_i}\}_{i=1}^{I}$ be a collection of I sets of image correspondences. For each $i = 1, \ldots, I$ and each $j = 1, \ldots, J_i$, write $\mathbf{m}_{ij} = [u_{ij}, v_{ij}, 1]^\top$ and $\mathbf{m}'_{ij} = [u'_{ij}, v'_{ij}, 1]^\top$, and let

$\mathbf{z}_{ij} = [u_{ij}, v_{ij}, u'_{ij}, v'_{ij}]^\top$. Suppose that each pair $\mathbf{m}_{ij}$, $\mathbf{m}'_{ij}$ comes equipped with a pair of 2×2 respective covariance matrices $\mathbf{\Lambda}_{u_{ij},v_{ij}}$, $\mathbf{\Lambda}_{u'_{ij},v'_{ij}}$. Let

$$\mathbf{\Lambda}_{\mathbf{z}_{ij}} = \begin{bmatrix} \mathbf{\Lambda}_{u_{ij},v_{ij}} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}_{u'_{ij},v'_{ij}} \end{bmatrix}.$$

Suppose that, for each $i = 1 \dots I$, a homography estimate $\mathbf{\Theta}_i$ is to be evolved based on $\{\mathbf{m}_{ij}, \mathbf{m}'_{ij}\}_{j=1}^{J_i}$ in such a way that $\mathbf{\Theta} = [\mathbf{\Theta}_1, \dots, \mathbf{\Theta}_I]$ satisfies the implicit constraint $\mathbf{\Theta} = \mathbf{\Pi}(\boldsymbol{\eta})$ for some $\boldsymbol{\eta}$. For each $i = 1, \dots, I$, let $\theta_i = \mathrm{vec}(\mathbf{\Theta}_i)$. As it turns out, the *approximate maximum likelihood* cost function with which to fulfil the estimation task in a statistically meaningful way is given by

$$J_{\mathrm{AML}}(\mathbf{\Theta}) = \sum_{i=1}^{I} \sum_{j=1}^{J_i} \theta_i^T \mathbf{V}_{ij} \mathbf{\Sigma}_{ij}^{-1} \mathbf{V}_{ij}^\top \theta_i,$$

where

$$\begin{aligned} \mathbf{V}_{ij} &= \mathbf{V}(\mathbf{z}_{ij}), \\ \mathbf{\Sigma}_{ij} &= (\mathbf{I}_2 \otimes \theta_i^\top) \mathbf{B}_{ij} (\mathbf{I}_2 \otimes \theta_i), \\ \mathbf{B}_{ij} &= \partial_{\mathbf{z}} \mathrm{vec}(\mathbf{V}(\mathbf{z}))|_{\mathbf{z}=\mathbf{z}_{ij}} \mathbf{\Lambda}_{\mathbf{z}_{ij}} \left[\partial_{\mathbf{z}} \mathrm{vec}(\mathbf{V}(\mathbf{z}))|_{\mathbf{z}=\mathbf{z}_{ij}} \right]^\top, \end{aligned}$$

and, explicitly,

$$\begin{aligned} \partial_{\mathbf{z}} \mathrm{vec}(\mathbf{V}(\mathbf{z})) = \; - [& \mathrm{vec}((\mathbf{e}_1 \otimes [\mathbf{m}']_\times)\mathbf{I}_{3\times 2}), \mathrm{vec}((\mathbf{e}_2 \otimes [\mathbf{m}']_\times)\mathbf{I}_{3\times 2}), \\ & \mathrm{vec}((\mathbf{m} \otimes [\mathbf{e}_1]_\times)\mathbf{I}_{3\times 2}), \mathrm{vec}((\mathbf{m} \otimes [\mathbf{e}_2]_\times)\mathbf{I}_{3\times 2})]. \end{aligned}$$

Here each summand $\theta_i^T \mathbf{V}_{ij} \mathbf{\Sigma}_{ij}^{-1} \mathbf{V}_{ij}^\top \theta_i$ represents the *Sampson distance* between $\mathbf{\Theta}_i$ and $\{\mathbf{m}_{ij}, \mathbf{m}'_{ij}\}$ [11]. Upon introducing the function

$$J'_{\mathrm{AML}}(\boldsymbol{\eta}) = J_{\mathrm{AML}}(\mathbf{\Pi}(\boldsymbol{\eta})),$$

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

7. Cost function optimisation

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

$$J'_{\mathrm{AML}}(\boldsymbol{\eta}) = \sum_{i=1}^{I} \sum_{j=1}^{J_i} r'^2_{ij}(\boldsymbol{\eta}) = \|\mathbf{r}'(\boldsymbol{\eta})\|^2, \tag{10}$$

where, for each $i = 1, \dots, I$ and each $j = 1, \dots, J_i$,

$$\begin{aligned} r'_{ij}(\boldsymbol{\eta}) &= r_{ij}(\boldsymbol{\pi}_i(\boldsymbol{\eta})), \\ r_{ij}(\theta_i) &= \left(\theta_i^T \mathbf{V}_{ij} \mathbf{\Sigma}_{ij}^{-1} \mathbf{V}_{ij}^\top \theta_i \right)^{1/2}, \\ \boldsymbol{\pi}_i(\boldsymbol{\eta}) &= \mathrm{vec}(\mathbf{\Pi}_i(\boldsymbol{\eta})), \\ \mathbf{r}'(\boldsymbol{\eta}) &= [r'_{11}(\boldsymbol{\eta}), \dots, r'_{1J_1}(\boldsymbol{\eta}), \dots, r'_{I1}(\boldsymbol{\eta}), \dots, r'_{IJ_I}(\boldsymbol{\eta})]^\top. \end{aligned}$$

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

$$\boldsymbol{\eta}_{n+1} = \boldsymbol{\eta}_n - [\mathbf{H}(\boldsymbol{\eta}_n) + \lambda_n \mathbf{I}_{4I+12})]^{-1} [\partial_{\boldsymbol{\eta}} \mathbf{r}'(\boldsymbol{\eta}_n)]^\top \mathbf{r}'(\boldsymbol{\eta}_n),$$

where $\mathbf{H} = (\partial_{\boldsymbol{\eta}} \mathbf{r}')^\top \partial_{\boldsymbol{\eta}} \mathbf{r}'$ and λ_n is a non-negative scalar that dynamically changes from step to step [15]. Here $\partial_{\boldsymbol{\eta}} \mathbf{r}'$ is the Jacobian matrix

$$\partial_{\boldsymbol{\eta}} \mathbf{r}' = [(\partial_{\boldsymbol{\eta}} r'_{11})^\top, \dots, (\partial_{\boldsymbol{\eta}} r'_{1J_1})^\top, \dots, (\partial_{\boldsymbol{\eta}} r'_{I1})^\top, \dots, (\partial_{\boldsymbol{\eta}} r'_{IJ_I})^\top]^\top.$$

To determine $\partial_{\boldsymbol{\eta}} \mathbf{r}'$ explicitly, note that, for each $i = 1, \dots, I$ and each $j = 1, \dots, J_i$,

$$\partial_{\boldsymbol{\eta}} r'_{ij}(\boldsymbol{\eta}) = \partial_{\theta_i} r_{ij}(\boldsymbol{\pi}_i(\boldsymbol{\eta})) \partial_{\boldsymbol{\eta}} \boldsymbol{\pi}_i(\boldsymbol{\eta})$$

with

$$\begin{aligned} \partial_{\theta_i} r_{ij}(\theta_i) &= (r_{ij}(\theta_i))^{-1} \theta_i^\top \mathbf{X}^{(ij)}_{\theta_i}, \\ \mathbf{X}^{(ij)}_{\theta_i} &= \mathbf{M}^{(ij)}_{\theta_i} - \mathbf{N}^{(ij)}_{\theta_i}, \\ \mathbf{M}^{(ij)}_{\theta_i} &= \mathbf{V}_{ij} \mathbf{\Sigma}_{ij}^{-1} \mathbf{V}_{ij}^\top, \\ \mathbf{N}^{(ij)}_{\theta_i} &= (\boldsymbol{\xi}_{ij}^\top \otimes \mathbf{I}_9) \mathbf{B}_{ij} (\boldsymbol{\xi}_{ij} \otimes \mathbf{I}_9), \\ \boldsymbol{\xi}_{ij} &= \mathbf{\Sigma}_{ij}^{-1} \mathbf{V}_{ij}^\top \theta_i. \end{aligned}$$

In the setup described by (6) and (7), we have, in accordance with (4),

$$\boldsymbol{\pi}_i(\boldsymbol{\eta}) = \mathrm{vec}(w_i \mathbf{A} + \mathbf{b}\mathbf{v}_i^\top) = w_i \mathbf{a} + \mathbf{v}_i \otimes \mathbf{b}$$

for each $i = 1, \dots, I$. 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

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

Representing, for each $i = 1, \dots, I$, $\partial_{\boldsymbol{\eta}} r'_{ij}$ as

$$\partial_{\boldsymbol{\eta}} r'_{ij} = [\partial_{\mathbf{a}} r'_{ij}, \partial_{\mathbf{b}} r'_{ij}, \partial_{\mathbf{v}_1} r'_{ij}, \dots \partial_{\mathbf{v}_I} r'_{ij}, \partial_{w_1} r'_{ij}, \dots, \partial_{w_I} r'_{ij}],$$

one finds furthermore that

$$\begin{aligned} \partial_{\mathbf{a}} r'_{ij} &= w_i (r_{ij}(\boldsymbol{\pi}_i))^{-1} \boldsymbol{\pi}_i^\top \mathbf{X}^{(ij)}_{\boldsymbol{\pi}_i}, \\ \partial_{\mathbf{b}} r'_{ij} &= (r_{ij}(\boldsymbol{\pi}_i))^{-1} \boldsymbol{\pi}_i^\top \mathbf{X}^{(ij)}_{\boldsymbol{\pi}_i} (\mathbf{v}_i \otimes \mathbf{I}_3), \\ \partial_{\mathbf{v}_i} r'_{ij} &= (r_{ij}(\boldsymbol{\pi}_i))^{-1} \boldsymbol{\pi}_i^\top \mathbf{X}^{(ij)}_{\boldsymbol{\pi}_i} (\mathbf{I}_3 \otimes \mathbf{b}), \\ \partial_{\mathbf{v}_k} r'_{ij} &= \mathbf{0} \quad (k \neq i), \\ \partial_{w_i} r'_{ij} &= (r_{ij}(\boldsymbol{\pi}_i))^{-1} \boldsymbol{\pi}_i^\top \mathbf{X}^{(ij)}_{\boldsymbol{\pi}_i} \mathbf{a}, \\ \partial_{w_k} r'_{ij} &= \mathbf{0} \quad (k \neq i). \end{aligned}$$

With $\partial_{\boldsymbol{\eta}} \mathbf{r}'$ thus determined, all that is now needed is a means for determining a suitable initial value of $\boldsymbol{\eta}$.

One possible procedure for determining suitable initial values of $\boldsymbol{\eta}$ is presented in Algorithm 1. In essence, the procedure first establishes relative scale factors between a collection of homography matrices using the fact that any homography of the

Algorithm 1: InitialiseLatentVariables

/* Find initial values for $\mathbf{A}$, $\mathbf{b}$, $\mathbf{v}_i$ and w_i in equation (6) */
Input: $\{\mathbf{H}_1, \mathbf{H}_2, \ldots, \mathbf{H}_I\}$
Output: $\boldsymbol{\eta} = [\mathrm{vec}(\mathbf{A})^\top, \mathbf{b}^\top, \mathbf{v}_1^\top, \ldots, \mathbf{v}_I^\top, w_1, \ldots, w_I]^\top$

Set $\mathbf{A} = \mathbf{H}_1$ /* definition of $\mathbf{A}$ */
foreach $i \in \{2, \ldots, I\}$ **do**
    Compute eigenvalues: $\{\lambda_1, \lambda_2, \lambda_3\} = \mathrm{eig}(\mathbf{H}_i^{-1}\mathbf{H}_1)$
    Find closest pair: $\{k', l'\} = \arg\min_{\{k,l\}\ k\neq l}(\lambda_k - \lambda_l)^2$
    Compute average: $\mu_{i1} = (\lambda_{k'} + \lambda_{l'})/2$
end foreach
Set $\mathbf{M} = [\mu_{21}\mathbf{H}_2 - \mathbf{H}_1, \mu_{31}\mathbf{H}_3 - \mathbf{H}_1, \ldots, \mu_{I1}\mathbf{H}_I - \mathbf{H}_1]$
Take for $\mathbf{b}$ the left singular vector of $\mathbf{M}$ corresponding to the largest singular value /* definition of $\mathbf{b}$ */
Set $\mathbf{v}_i = \begin{cases} [0,0,0]^\top & \text{for } i = 1, \\ \|\mathbf{b}\|^{-2}(\mu_{i1}\mathbf{H}_i - \mathbf{H}_1)^\top\mathbf{b} & \text{for } i \in \{2, \ldots, I\} \end{cases}$ /* definition of $\mathbf{v}_i$ */
Set $w_i = 1$ for $i \in \{1, \ldots, I\}$ /* definition of w_i */
return $\boldsymbol{\eta} = [\mathrm{vec}(\mathbf{A})^\top, \mathbf{b}^\top, \mathbf{v}_1^\top, \ldots, \mathbf{v}_I^\top, w_1, \ldots, w_I]^\top$

form $\mathbf{H}_{i_1}^{-1}\mathbf{H}_{i_2}$ has a double eigenvalue. Once the relative scales of the individual homography matrices are determined, singular value decomposition is next used to determine the latent variable $\mathbf{b}$. Finally, straightforward algebraic manipulations are applied to recover all remaining latent variables.

8. A robust AML cost function

Employing the Huber loss function

$$C(r) = \begin{cases} r^2 & \text{if } |r| < b, \\ 2b|r| - b^2 & \text{otherwise,} \end{cases}$$

where b is a tunable positive constant, in place of the quadratic loss function r^2, one can define a robust version of J'_{AML}, denoted $J'_{\mathrm{R-AML}}$, as

$$J'_{\mathrm{R-AML}}(\boldsymbol{\eta}) = \sum_{i=1}^{I}\sum_{j=1}^{J_i} C(r'_{ij}(\boldsymbol{\eta})).$$

This function can be optimised with a suitable modification of the LM scheme presented in the foregoing section. To describe necessary changes, we first note that $C(r)$ can be represented as $C(r) = D^2(r)$, where

$$D(r) = \begin{cases} r & \text{if } |r| \leq b, \\ (2br - b^2)^{1/2} & \text{if } r > b, \\ -(-2br - b^2)^{1/2} & \text{if } r < -b. \end{cases}$$

The function D is continuously differentiable, with the derivative given by

$$\frac{\mathrm{d}D}{\mathrm{d}r}(r) = \begin{cases} 1 & \text{if } |r| \leq b, \\ b(2br - b^2)^{-1/2} & \text{if } r > b, \\ b(-2br - b^2)^{-1/2} & \text{if } r < -b. \end{cases}$$

Introducing the vector

$$\mathbf{d}(\mathbf{r}'(\boldsymbol{\eta})) = [D(r'_{11}(\boldsymbol{\eta})), \ldots, D(r'_{1J_1}(\boldsymbol{\eta})), \ldots, D(r'_{I1}(\boldsymbol{\eta})), \ldots, D(r'_{IJ_I}(\boldsymbol{\eta}))]^\top,$$

we can rewrite $J'_{\mathrm{R-AML}}$ in a least-square form as

$$J'_{\mathrm{R-AML}}(\boldsymbol{\eta}) = \|\mathbf{d}(\mathbf{r}'(\boldsymbol{\eta}))\|^2.$$

With $\mathbf{d}(\mathbf{r}'(\boldsymbol{\eta}))$ taking the role of $\mathbf{r}'(\boldsymbol{\eta})$ in (10), all what is needed to set up an LM scheme for optimising $J'_{\mathrm{R-AML}}$ is an explicit expression for the Jacobian of $\mathbf{d}(\mathbf{r}'(\boldsymbol{\eta}))$. But this is given by

$$\partial_{\boldsymbol{\eta}}\mathbf{d}(\mathbf{r}') = [(\partial_{\boldsymbol{\eta}} D(r'_{11}))^\top, \ldots, (\partial_{\boldsymbol{\eta}} D(r'_{1J_1}))^\top, \ldots, (\partial_{\boldsymbol{\eta}} D(r'_{I1}))^\top, \ldots, (\partial_{\boldsymbol{\eta}} D(r'_{IJ_I}))^\top]^\top,$$
$$\partial_{\boldsymbol{\eta}} D(r'_{ij}) = \frac{\mathrm{d}D}{\mathrm{d}r}(r'_{ij})\,\partial_{\boldsymbol{\eta}} r'_{ij}(\boldsymbol{\eta}),$$

9. Experimental design

We compared our new homography estimation algorithm with established methods on synthetic and real data. Our investigation focused on the direct linear transform (DLT) [11], rank constraint enforcement (RANK) [16], weighted alternating least squares (WALS) [2], approximate maximum likelihood with homography covariances (AML-COV) [7], approximate maximum likelihood with Sampson distance (AML-SMPS), separate bundle adjustment for multiple homographies (BA-SEP), and joint bundle adjustment for multiple homographies (BA-JOINT). With the exception of DLT, data points for all homographies were jointly normalised before they were passed to the estimation methods [3, 4, 10]. DLT operated on separately normalised data points.

Because DLT is the prevailing method used to separately estimate homographies, we took it as the baseline in our experiments. Usually separate bundle adjustment (BA-SEP) is con-

Table 1: Summary of homography estimation methods and their properties

Method	Required homographies	Initialisation	Cost function	Optimisation
DLT	1	N/A	algebraic distance	singular value decomposition
RANK	5 or more	DLT	Frobenius norm	singular value decomposition
WALS	3 or more	DLT + procedure in [2]	Mahalanobis norm	alternating least squares
AML-COV	2 or more	DLT + Algorithm 1	Mahalanobis norm	Levenberg–Marquardt
AML-SMPS	2 or more	DLT + Algorithm 1	Sampson distance	Levenberg–Marquardt
BA-SEP	2 or more	DLT	reprojection error	Levenberg–Marquardt
BA-JOINT	2 or more	DLT + Algorithm 1	reprojection error	Levenberg–Marquardt

sidered a gold standard benchmark for homography estimation. Separate bundle adjustment minimises the reprojection error

$$\sum_{i=1}^{I}\sum_{j=1}^{J_i}\left(d(\mathbf{m}_{ij},\underline{\mathbf{m}}_{ij})^2 + d(\mathbf{m}'_{ij},\mathbf{H}_i\underline{\mathbf{m}}_{ij})^2\right),$$

where $\mathbf{H}_i$ denotes the i-th homography matrix and $\{\{\underline{\mathbf{m}}_{ij}\}_{j=1}^{J_i}\}_{i=1}^{I}$ represents 2D points in the first image. However, as separate bundle adjustment does not enforce homography constraints, it should not really be granted gold standard status. In line with this, as a gold standard benchmark, we utilised joint bundle adjustment which is bundle adjustment employing the parametrisation given in (6) and (7) so that all inter-homography constraints are satisfied. Joint bundle adjustment minimises the gold standard reprojection error

$$\sum_{i=1}^{I}\sum_{j=1}^{J_i}\left(d(\mathbf{m}_{ij},\underline{\mathbf{m}}_{ij})^2 + d(\mathbf{m}'_{ij},\boldsymbol{\Pi}_i(\boldsymbol{\eta})\underline{\mathbf{m}}_{ij})^2\right)$$

over all choices of parameter vectors $\boldsymbol{\eta}$ and 2D points $\{\{\underline{\mathbf{m}}_{ij}\}_{j=1}^{J_i}\}_{i=1}^{I}$. In all experiments we initialised separate bundle adjustment with DLT, and joint bundle adjustment with the procedure from Algorithm 1. In both cases we also seeded $\{\{\underline{\mathbf{m}}_{ij}\}_{j=1}^{J_i}\}_{i=1}^{I}$ with the data points extracted from the first image. A summary of the attributes of all the estimators that we examined is presented in Table 1.

9.1. Synthetic data

We distinguished between two types of synthetic scenes. In the first type of synthetic scene data points corresponding to different homographies overlapped minimally and formed clusters (see Figure 2a), whilst in the second type of synthetic scene data points corresponding to different homographies overlapped substantially (see Figure 2b). The first type of synthetic scene was more realistic; however, the second type of synthetic scene was also of interest because it resulted in homography estimates that had spherical covariance matrices. By a spherical covariance matrix we mean a covariance matrices all of whose non-zero eigenvalues are almost equal.

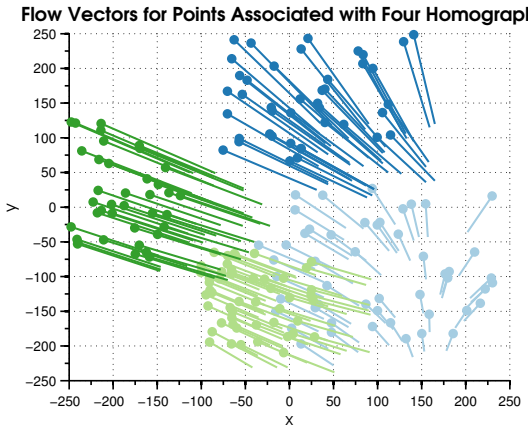
Type 1 synthetic scenes with a specified number of homographies and ground truth correspondences were generated using the following steps:

1. Select a realistic stereo camera configuration and set intrinsic camera parameters thus forming two projection matrices.
2. Generate a plane situated at a random distance in front of the first camera plane.
3. Apply a random rotation and translation to the plane. This rotated and translated plane together with the projection matrices gives rise to a homography.
4. Generate a random sized rectangular region in the first image and sample a random number of points within this region.
5. Transfer the points within the rectangular region in the first image to the second image using the homography from step 3.
6. If insufficient points fall within the boundary of the second image, then discard the homography and repeat the procedure from step 2; otherwise proceed to step 7.
7. Add zero-mean Gaussian noise at a specified standard deviation to the corresponding points.
8. Output the ground truth homography, ground truth corresponding points, and noise perturbed points.

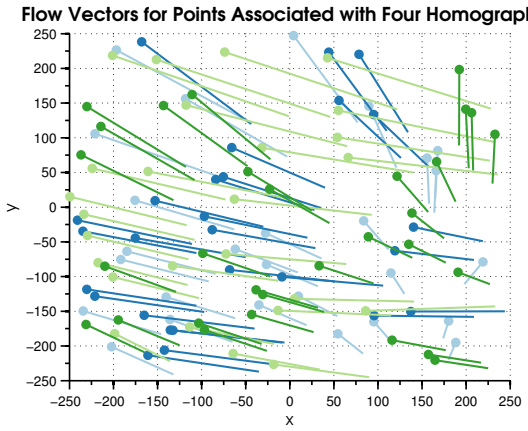
Steps 2 to 8 were repeated for each homography. The fourth step is important because if homographies span varying spatial regions, then they will also be estimated with varying levels of uncertainty. Type 2 synthetic scenes were generated using the same protocol, with the exception that the fourth step was modified so that the rectangular region always spanned the entire image. This modification ensured that homographies were estimated with similar levels of uncertainty.

9.2. Synthetic data with outliers

To analyse the robust variant of our algorithm, we generated numerous type 1 synthetic scenes with two homographies per scene, and ensured that 10% of corresponding points were outliers (false correspondences). Two planar structures were then automatically detected with the use of sequential RANSAC, resulting in two sets of noisy "inliers" per scene. Sequential RANSAC consists of (1) applying standard RANSAC on all data points to find an initial set of inliers, (2) removing these inliers from the dataset, and (3) applying RANSAC again on the remaining data points to find a second set of inliers. Although sequential RANSAC attempts to find inlier sets that are free of any false correspondences, it does not always succeed. The success of RANSAC depends on a suitable choice of threshold parameter to distinguish between inliers and outliers. If the



(a) Points corresponding to four homographies do not overlap, but form clusters.



(b) Points corresponding to four homographies overlap substantially.

Figure 2: Two different paradigms for generating synthetic data. In panel (a) corresponding points belonging to four different homographies do not overlap, but instead form small clusters in the image. Hence, the homography matrix estimates will have varying covariance matrices. In panel (b) corresponding points belonging to four different homographies overlap substantially and are spread out over the entire image. Consequently, the homography matrix estimates will have nearly spherical covariance matrices.

Table 2: Measuring the sphericity of homography covariance matrices associated with the results in Figures 3a and 3b. The sphericity of a homography covariance matrix is determined by computing the ratio of the largest to the smallest (non-zero) eigenvalue of the covariance matrix.

(a) The average sphericity of homography matrices related to the results in Figure 3a.

Generating planes	Noise level (in pixels)				
	$\sigma = 0.2$	$\sigma = 0.4$	$\sigma = 0.6$	$\sigma = 0.8$	$\sigma = 1$
1st plane	809.04	374.62	534.65	421.42	568.84
2nd plane	663.01	412.09	597.27	736.75	250.88
3rd plane	1266.56	833.52	776.65	588.27	697.41
4th plane	445.95	973.53	800.85	1524.96	636.97
5th plane	491.80	526.67	890.49	470.95	484.29
6th plane	550.29	395.86	612.66	698.48	718.71

(b) The average sphericity of homography matrices related to the results in Figure 3b.

Generating planes	Noise level (in pixels)				
	$\sigma = 0.2$	$\sigma = 0.4$	$\sigma = 0.6$	$\sigma = 0.8$	$\sigma = 1$
1st plane	13.16	13.38	12.84	12.79	13.14
2nd plane	13.36	13.19	13.29	13.25	13.34
3rd plane	13.53	13.08	13.53	13.33	12.76
4th plane	13.25	13.12	13.06	12.98	13.13
5th plane	13.10	13.19	13.39	13.01	12.69
6th plane	12.82	12.74	13.14	12.96	12.71

threshold is too strict, very few points will be considered inliers. If the threshold is too liberal, false correspondences may be included. The Huber norm offers a partial remedy to the threshold dilemma. One can use a liberal RANSAC threshold so that many true inliers are correctly identified, followed by a strict Huber norm threshold so that any accidental false correspondences are down-weighted.

9.3. Real data with manually demarcated planar regions

To complement our extensive synthetic experiments with experiments on real data, we collected several images of planar surfaces. Corresponding points between a pair of views were matched by correlation using a Harris corner detector [9]. Planar regions within the images were demarcated using polygons, and data points inside the polygons were grouped together. The number of polygons determined the number of homographies.

9.4. Real data where planar regions are discovered using sequential RANSAC

To further demonstrate the practical utility of our algorithm we utilised two views of Merton College from the Oxford dataset.[1] We used a Harris corner detector to detect 147 interest points and matched corresponding points by correlation. We added a further 132 (47%) false correspondences (pure outliers). Two planar structures were then automatically detected with the aid of sequential RANSAC, resulting in two sets of inliers for two homographies.

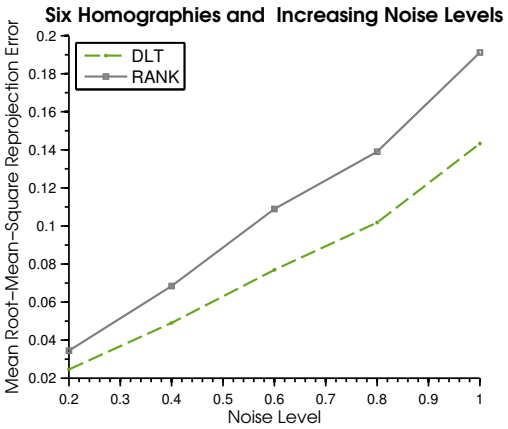
9.5. Quantification

On synthetic data the common distance used to quantify data-model discrepancies was the *mean root-mean-square (RMS) reprojection error from truth*

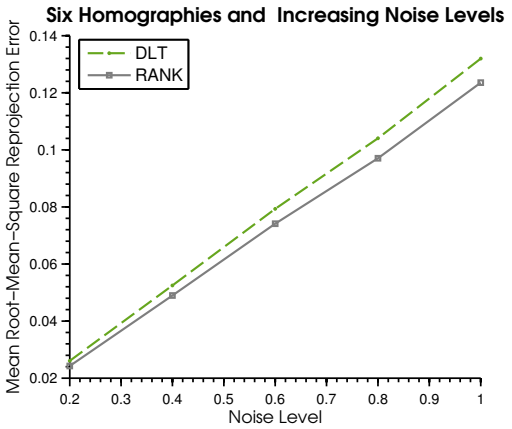
$$\frac{1}{I}\sum_{i=1}^{I}\sqrt{\frac{1}{4J_iK}\sum_{k=1}^{K}\min_{\underline{\mathbf{m}}_{ij}^{(k)}}\sum_{j=1}^{J_i}\left(d(\overline{\mathbf{m}}_{ij}^{(k)},\underline{\mathbf{m}}_{ij}^{(k)})^2+d(\overline{\mathbf{m}}_{ij}^{\prime(k)},\widehat{\Theta}_i\underline{\mathbf{m}}_{ij}^{(k)})^2\right)},$$

where K is the number of experiments, and, for each $k = 1,\ldots,K$, $\{\{\overline{\mathbf{m}}_{ij}^{(k)},\overline{\mathbf{m}}_{ij}^{\prime(k)}\}_{j=1}^{J_i}\}_{i=1}^{I}$ are noiseless data and $\{\{\underline{\mathbf{m}}_{ij}^{(k)}\}_{j=1}^{J_i}\}_{i=1}^{I}$ are arbitrary 2D points over which the minimum is taken in the k-th experiment.

[1] http://www.robots.ox.ac.uk/~vgg/data.html



(a) Points corresponding to six homographies overlapped minimally and formed clusters.



(b) Points corresponding to six homographies overlapped substantially and covered entire image.

Figure 3: Comparison of rank constraint enforcement vs DLT on six homographies with fixed data points and increasing noise level. Results are based on 1500 trials. In experiments reported in panel (a) homography covariance matrices were far from spherical, while in experiments reported in panel (b) the covariance matrices were close to spherical. The sphericity of a homography covariance matrix is determined by computing the ratio of the largest to the smallest (non-zero) eigenvalue of the covariance matrix. The results show that rank constraint enforcement yields improvements only when all homography estimates have spherical covariance matrices, which in turn only happens when the image data points associated with homographies are well spaced out in the image and overlap (see for example Figure 2b).

Table 3: Comparison of estimation methods on data with outliers and multiple structures. Random Sampling and Consensus with DLT (DLT-RANSAC) was used sequentially to determine inliers for two homographies. The homography estimates were further refined using BA-SEP and R-AML-SMPS. Results are based on 1500 trials, with 2 homographies per trial and the noise level of $\sigma = 1$ pixels. Data points associated with homographies overlapped minimally and formed clusters.

Methods	Homography	
	1	2
DLT-RANSAC	0.2512	0.2518
DLT-RANSAC + BA-SEP	0.2395	0.2395
DLT-RANSAC + R-AML-SMPS	0.2043	0.2048

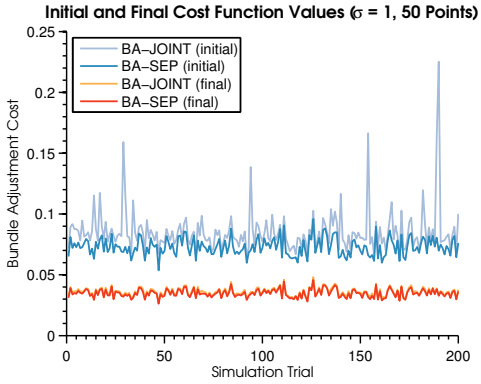
On real data we compared the quality of estimators by using the estimated homographies to transfer textures inside designated polygon regions from the first view to the second view. Imprecisely aligned textures demonstrate unreliable homographies.

10. Results and discussion

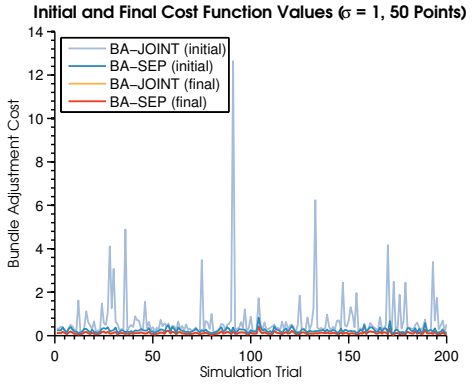
The results of our first experiments demonstrate that SVD-based rank constraint enforcement has limited practical value. Apart from the fact that rank constraints can only be applied when there are five or more underlying planar surfaces (a rare occurrence), the procedure only improves the quality of the homography estimates if the covariance matrices associated with the homographies are spherical. If the homographies are estimated with different degrees of uncertainty, then rank constraint enforcement leads to less accurate homography estimates. This assertion is supported by the results presented in Figure 3. Figure 3a together with Table 2a illustrate that when homography covariance matrices were far from spherical, rank constraint enforcement was consistently worse than DLT. On the other hand, Figure 3b and Table 2b demonstrate that when the homography covariance matrices were all close to spherical, rank constraint enforcement led to improved homography estimates.

An assessment of all our simulations affirms that enforcing homography constraints using our latent variable parametrisation can considerably improve the quality of homography estimates. The results presented in Figure 5 demonstrate close to 10% improvement in the estimation of two homographies for a range of noise levels for both type 1 and type 2 synthetic scenes. However, the experiments also revealed that for the more challenging type 1 synthetic scenes, BA-JOINT, AML-SMPS, and AML-COV occasionally failed to improve upon the BA-SEP estimate because they converged to poor local minima. For example, Table 4 shows that for a noise level of $\sigma = 1$ pixels, BA-JOINT, and AML-SMPS improved the estimate in 97% of trials, whilst AML-COV improved the estimate in 96% of trials. At larger noise levels ($\sigma = 3$ pixels), BA-JOINT, and AML-SMPS improved the estimate in 94% of trials, whilst AML-COV in 80% of trials.

By introducing homography covariance information into their iterative WALS scheme, Chen and Suter hoped to overcome the limitations of SVD-based rank constraint enforcement. They succeeded partially in that WALS with their initialisation can be applied when there are three or more planar surfaces in the scene. However, whilst their scheme sometimes improves upon DLT and BA-SEP, it is prone to converging to poor minima. For example, Figure 6 shows that even though WALS is better than DLT and BA-SEP, it is not as accurate as AML-COV or AML-SMPS. Moreover, the numerical instability of WALS is apparent in Table 5, which shows that for four homographies and the noise level of $\sigma = 5$ pixels, WALS only improved upon BA-SEP in 28% of trials, whereas AML-SMPS



(a) Points corresponding to four homographies overlapped substantially and covered entire image.



(b) Points corresponding to four homographies overlapped minimally and formed clusters.

Figure 4: Comparison of initial and final cost function values for BA-SEP and BA-JOINT. In panel (b) the initial BA-JOINT cost is very high, indicating that whilst Algorithm 1 upgrades a set of separately estimated homographies into a new set of homographies that satisfy all inter-homography constraint, the consistent homographies from the new set may have a high initial reprojection error.

achieved a 82% success rate.

An astute reader may wonder why BA-JOINT occasionally fails to improve upon BA-SEP. The reason for the occasional failure of BA-JOINT has to do with Algorithm 1 which is used to initialise the latent variables in BA-JOINT. For type 2 synthetic scenes Algorithm 1 provides latent variables parametrising fully consistent homographies that also have low reprojection error (see Figure 4a). For type 1 synthetic scenes the algorithm still provides latent variables parametrising fully consistent homographies, but these homographies no longer fit the corresponding points well and have a large reprojection error (see Figure 4b). In the process of minimising a large reprojection error BA-JOINT occasionally converges to a poor local minimum.

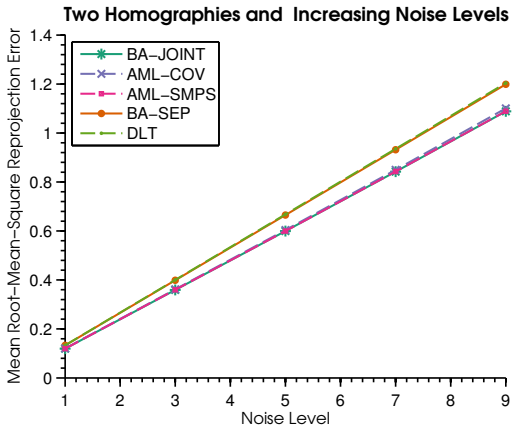
Our examinations have also established that the mean root-mean-square reprojection error decreases as the number of homographies that are jointly estimated increases. This effect is evident in Figure 7 and also in Table 8, where the percentage reduction in reprojection error is quantified. Table 8 shows that the accuracy of AML-SMPS was very close to that of BA-JOINT, ranging from a 23% average error reduction with four homographies, to a 30% average error reduction with eight homographies. We also noticed that as the number of homographies increases the chance of converging to a poor local minimum gradually increases (see Table 6).

Further analysis shows that the percentage reduction in reprojection error that can be achieved by enforcing inter-homography constraints is fairly independent of the number of data points. Figure 8 and Table 9 show that the percentage reduction in reprojection error fluctuated minimally as the number of data points was varied from 10 to 50. However, some of the algorithms are more prone to converging to poor minima when given only few data points. For example, Table 7 shows that with only 10 data points WALS managed to improve upon the BA-SEP solution in only 30.2% of trials, whereas BA-JOINT and AML-SMPS succeed in 98.2% and 78.66% of trials, respectively.

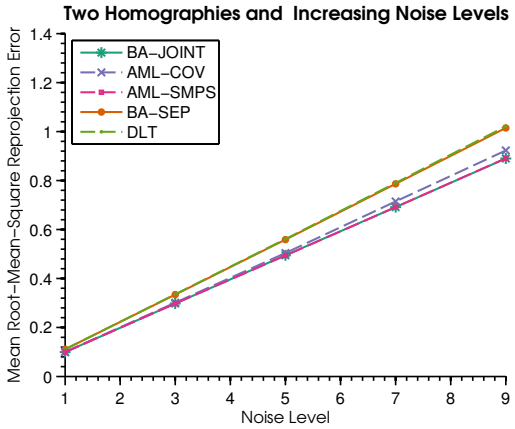
Results on synthetic data with outliers, given in Table 3, demonstrate that the robust variant of the AML cost function leads to superior results when compared to the commonly employed DLT-RANSAC + BA-SEP combination. This means that the theory developed in this paper can be incorporated into various real-world applications which typically rely on the DLT-RANSAC followed by BA-SEP workflow.

Examples of texture transfer on real-world data, depicted in Figures 10, 11, 12, and 13, all demonstrate by the quality of the texture alignments that our method yields more accurate homography estimates. The definite improvement is also evident on the Merton College images in Figure 14.

Currently the main drawback of AML-SMPS is its running time. According to the running times listed in Table 10, AML-COV is by far the fastest, while AML-SMPS is the slowest. In principle, AML-SMPS should be faster than BA-JOINT because it solves a smaller optimisation problem (it does not optimise over corresponding points). However, for the AML-SMPS cost function the Levenberg–Marquardt algorithm takes very many iterations to converge when a current estimate is close to an optimal solution. Figure 9 traces the reduction in cost function value for consecutive iterations of the LM scheme on a sample of 500 simulation trials. The plot shows that typically after approximately 20 iterations the cost function has essentially reached its minimum. For the remaining hundreds of iterations, the cost function plateaus and only decreases marginally between iterations. We believe this uncharacteristically poor performance of the LM algorithm near the minimum may have to do with a combination of factors including (1) the intrinsic rank deficiency of the Jacobian matrix at a local minimum, (2) inherent gauge freedom in the choice of parametrisation, and (3) the Sampson distance itself. We also believe that with an appropriate modification to the LM algorithm the stagnation problem can be overcome, but we leave the pursuit of the best optimisation strategy to future work.



(a) Points corresponding to two homographies overlapped substantially and covered entire image.

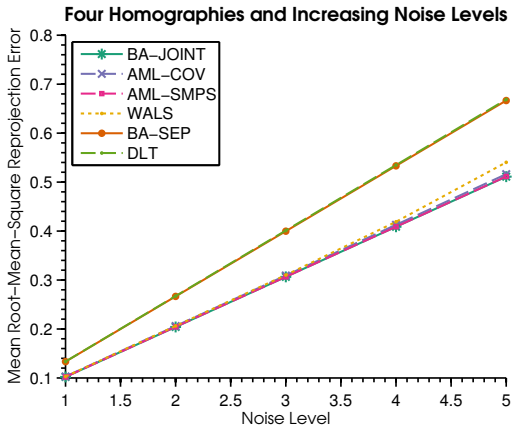


(b) Points corresponding to two homographies overlapped minimally and formed clusters.

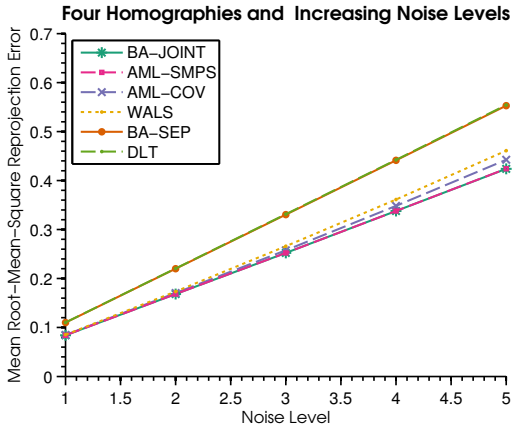
Figure 5: Comparison of estimation methods on two homographies with 50 data points and increasing noise levels. Results are based on 1500 trials. All joint homography estimation methods were seeded with the aid of Algorithm 1. In experiments reported in panel (b) BA-JOINT, AML-SMPS, and AML-COV occasionally converged to poor minima resulting in estimates that were considerably worse than BA-SEP or DLT. These trials were treated as outliers and removed from the plot. A summary of the percentage of trials for which these methods converged to a better solution than BA-SEP is presented in Table 4.

Table 4: Measuring the percentage of trials as per Figure 5b for which various joint homography estimation methods, when seeded with the aid of Algorithm 1, converged to a solution with lower reprojection error than the solution generated by the separate bundle adjustment homography estimation method.

Methods	Noise level (in pixels)				
	$\sigma = 1$	$\sigma = 3$	$\sigma = 5$	$\sigma = 7$	$\sigma = 9$
BA-JOINT	97.93%	94.80%	89.20%	85.00%	80.13%
AML-SMPS	97.80%	94.00%	87.26%	77.93%	66.80%
AML-COV	96.20%	80.80%	60.46%	44.53%	31.73%



(a) Points corresponding to four homographies overlapped substantially and covered entire image.

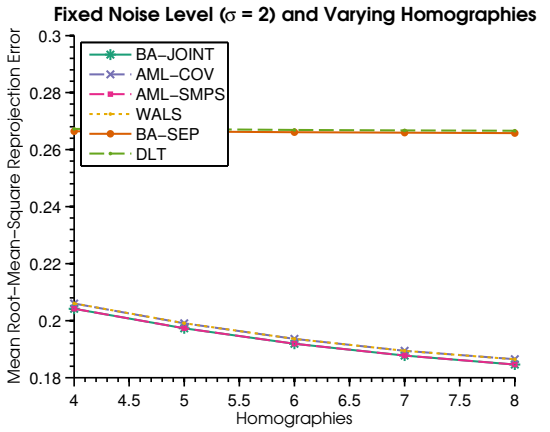


(b) Points corresponding to four homographies overlapped minimally and formed clusters.

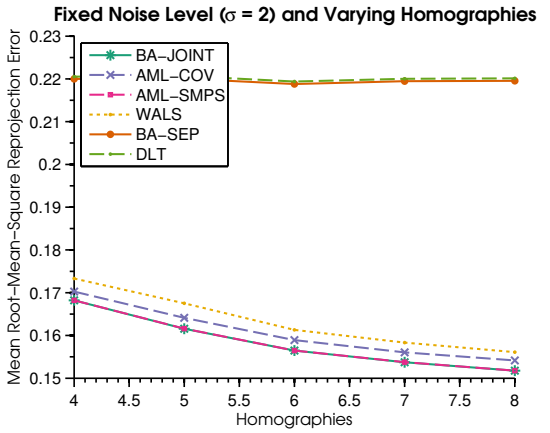
Figure 6: Comparison of estimation methods on four homographies with increasing noise levels. Results are based on 1500 trials. All joint homography estimation methods were seeded with the aid of Algorithm 1. In experiments reported in panel (b) BA-JOINT, AML-SMPS, AML-COV, and WALS occasionally converged to poor minima resulting in estimates that were considerably worse than DLT or BA-SEP. These trials were treated as outliers and removed from the plot. A summary of the percentage of trials for which these methods converged to a better solution than BA-SEP is presented in Table 5.

Table 5: Measuring the percentage of trials as per Figure 6b for which various joint homography estimation methods, when seeded with the aid of Algorithm 1, converged to a solution with lower reprojection error than the solution generated by the separate bundle adjustment homography estimation method.

Methods	Noise level (in pixels)				
	$\sigma = 1$	$\sigma = 2$	$\sigma = 3$	$\sigma = 4$	$\sigma = 5$
BA-JOINT	100.00%	100.00%	99.99%	99.99%	99.99%
AML-SMPS	99.60%	97.60%	94.40%	89.06%	82.86%
AML-COV	99.06%	92.46%	79.20%	64.86%	52.60%
WALS	97.40%	81.26%	60.00%	41.20%	28.13%



(a) Points corresponding to homographies overlapped substantially and covered entire image.

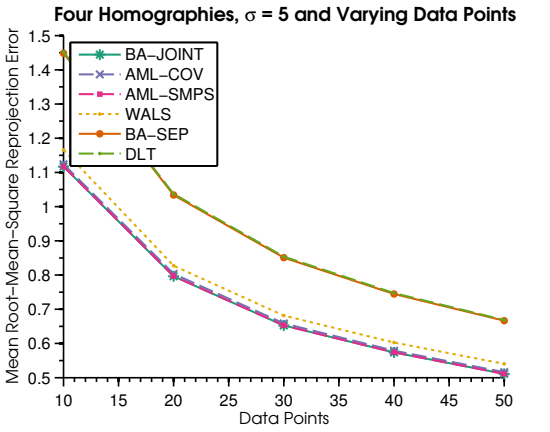


(b) Points corresponding to homographies overlapped minimally and formed clusters.

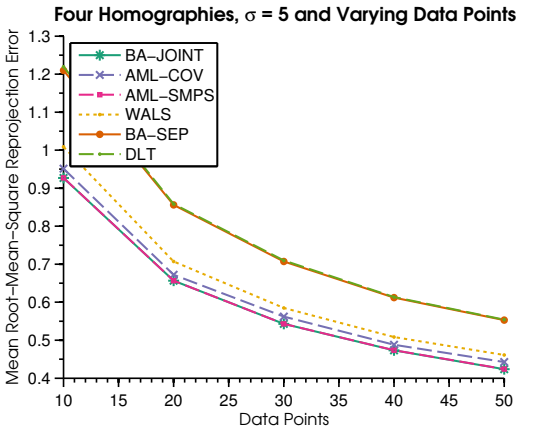
Figure 7: Comparison of estimation methods with variable number of homographies. Results are based on 1500 trials. All joint homography estimation methods were seeded with the aid Algorithm 1. In experiments reported in panel (b) BA-JOINT, AML-SMPS, AML-COV, and WALS occasionally converged to poor minima resulting in estimates that were considerably worse than BA-SEP or DLT. These trials were treated as outliers and removed from the plot. A summary of the percentage of trials for which these methods converged to a better solution than BA-SEP is presented in Table 6.

Table 6: Measuring the percentage of trials as per Figure 7b for which various joint homography estimation methods, when seeded with the aid of Algorithm 1, converged to a solution with lower reprojection error than the solution generated by the separate bundle adjustment homography estimation method. Results are based on 1500 trials with 50 data points and the noise level of $\sigma = 2$ pixels. Data points associated with homographies overlapped minimally and formed clusters.

Methods	Number of homographies				
	4	5	6	7	8
BA-JOINT	100.00%	100.00%	100.00%	100.00%	100.00%
AML-SMPS	97.60%	97.60%	96.46%	96.26%	95.13%
AML-COV	92.46%	88.73%	87.93%	84.40%	83.26%
WALS	81.26%	82.73%	82.60%	79.86%	79.00%



(a) Points corresponding to four homographies overlapped substantially and covered entire image.



(b) Points corresponding to four homographies overlapped minimally and formed clusters.

Figure 8: Comparison of estimation methods on four homographies with fixed noise level and increasing data points. Results are based on 1500 trials. All joint homography estimation methods were seeded with Algorithm 1. In experiments reported in panel (b) BA-JOINT, AML-SMPS, AML-COV, and WALS occasionally converged to poor minima resulting in estimates that were considerably worse than BA-SEP or DLT. These trials were treated as outliers and removed from the plot. A summary of the percentage of trials for which these methods converged to a poor solution is presented in Table 7.

Table 7: Measuring the percentage of trials as per Figure 8b for which various joint homography estimation methods, when seeded with the aid of Algorithm 1, converged to a solution with lower reprojection error than the solution generated by the separate bundle adjustment homography estimation method. Results are based on 1500 trials with four homographies and noise level of $\sigma = 2$ pixels. Data points associated with homographies overlapped minimally and formed clusters.

Methods	Number of data points				
	10	20	30	40	50
BA-JOINT	98.20%	99.33%	99.53%	99.93%	99.86%
AML-SMPS	78.66%	80.40%	82.73%	83.53%	82.86%
AML-COV	46.46%	51.66%	51.20%	51.93%	52.60%
WALS	30.20%	31.86%	26.40%	27.40%	28.13%

Table 8: Improvement over BA-SEP for estimation methods enforcing consistency constraints, expressed in terms of percentage reduction in reprojection error. Results are based on 1500 trials, with 50 data points per trial and the noise level of $\sigma = 2$ pixels.

(a) Data points associated with different homographies overlapped substantially and covered entire image.

Methods	Number of homographies				
	4	5	6	7	8
BA-JOINT	23.360%	25.919%	27.888%	29.403%	30.544%
AML-SMPS	23.355%	25.916%	27.885%	29.400%	30.540%
AML-COV	22.670%	25.264%	27.243%	28.770%	29.858%
WALS	22.658%	25.250%	27.241%	28.768%	29.850%

(b) Data points associated with different homographies overlapped minimally and formed clusters.

Methods	Number of homographies				
	4	5	6	7	8
BA-JOINT	23.541%	26.574%	28.488%	29.948%	30.877%
AML-SMPS	23.534%	26.567%	28.483%	29.942%	30.874%
AML-COV	22.607%	25.415%	27.377%	28.904%	29.784%
WALS	21.221%	23.868%	26.282%	27.856%	28.905%

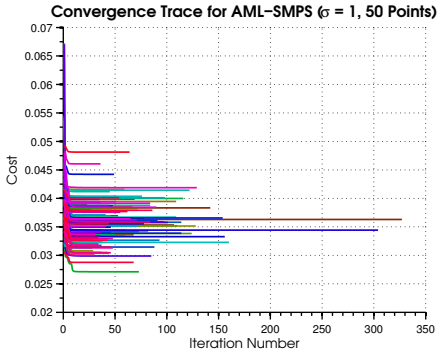
Table 9: Improvement over BA-SEP for estimation methods enforcing consistency constraints, expressed in terms of percentage reduction of reprojection error. Results are based on 1500 trials, with 4 homographies per trial and the noise level of $\sigma = 5$ pixels.

(a) Data points associated with different homographies overlapped substantially and covered entire image.

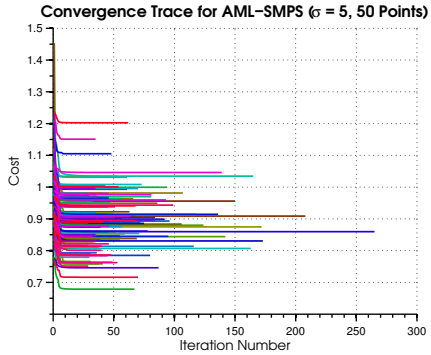
Methods	Number of data points				
	10	20	30	40	50
BA-JOINT	22.865%	22.952%	23.307%	22.935%	23.330%
AML-SMPS	22.865%	22.939%	23.294%	22.946%	23.299%
AML-COV	22.409%	22.287%	22.585%	22.262%	22.542%
WALS	19.528%	19.972%	19.912%	19.063%	18.936%

(b) Data points associated with different homographies overlapped minimally and formed clusters image.

Methods	Number of data points				
	10	20	30	40	50
BA-JOINT	23.414%	23.297%	23.216%	22.634%	23.339%
AML-SMPS	23.390%	23.281%	23.206%	22.611%	23.286%
AML-COV	21.389%	21.468%	20.507%	20.260%	19.977%
WALS	16.643%	17.374%	17.282%	16.929%	16.612%



(a) Data points associated with different homographies overlapped minimally and formed clusters.

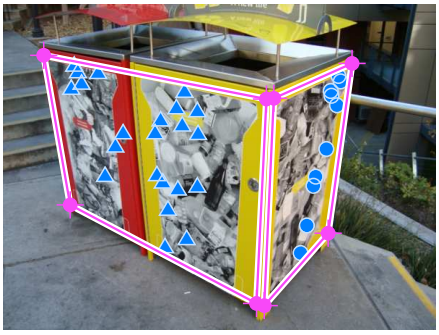


(b) Data points associated with different homographies overlapped substantially and covered entire image.

Figure 9: Panels (a) and (b) contain plots of the AML-SMPS cost function for subsequent iterations of the Levenberg–Marquardt optimisation method for 500 trials. After approximately 20 iterations the cost function has almost reached its minimum, and for the remaining hundreds of iterations the decrease in the cost between iterations is marginal.

Table 10: Median running time of various homography estimation methods for a noise level of $\sigma = 2$ pixels and 50 data points.

Methods	Number of homographies				
	4	5	6	7	8
BA-SEP	3.8209 s (6 iter.)	6.0324 s (6 iter.)	8.2606 s (6 iter.)	11.0717 s (6 iter.)	14.9065 s (6 iter.)
BA-JOINT	4.4657 s (6 iter.)	6.6883 s (6 iter.)	8.9932 s (6 iter.)	11.9537 s (6 iter.)	15.8113 s (6 iter.)
AML-SMPS	25.0437 s (47 iter.)	34.9915 s (51 iter.)	44.1130 s (53 iter.)	52.9921 s (56 iter.)	62.4978 s (58 iter.)
AML-COV	0.0176 s (4 iter.)	0.0277 s (4 iter.)	0.0289 s (4 iter.)	0.0324 s (4 iter.)	0.0358 s (4 iter.)
WALS	2.1205 s (1739 iter.)	3.0876 s (630 iter.)	3.9658 s (562 iter.)	6.0659 s (546 iter.)	13.2215 s (522 iter.)



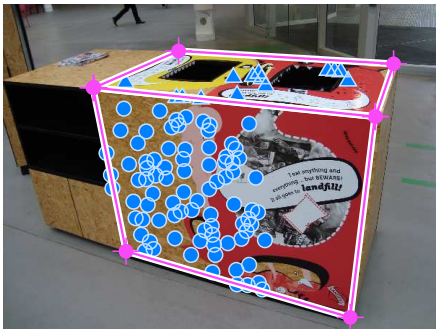
(a) Recycle bin view 1.

(b) Recycle bin view 2.

(c) BA-SEP result.

(d) AML-SMPS result.

Figure 10: Transferring textures inside two designated planar regions on recycle bins from the first viewpoint to the second viewpoint. Panel (c) shows that separate bundle adjustment failed to properly align one of the planes. In contrast, panel (d) shows that jointly estimating both homographies resulted in superior texture alignment.



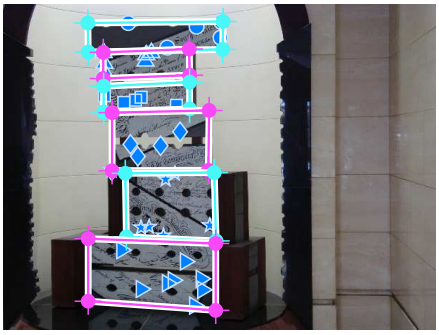
(a) Waste container view 1.

(b) Waste container view 2.

(c) BA-SEP result.

(d) AML-SMPS result.

Figure 11: Transferring textures inside two designated planar regions on waste containers from the first viewpoint to the second viewpoint. In panel (c) the transferred texture on the top plane of the waste container encroaches on the edge of the container, and is not aligned with the red border of the texture on the front-facing plane of the container. In contrast, panel (d) depicts a precise alignment.



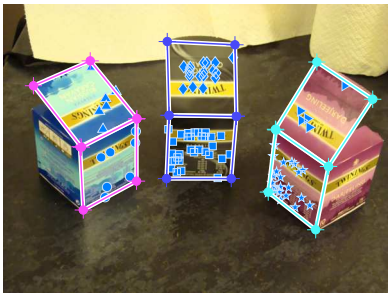
(a) Statue view 1.

(b) Statue view 2.

(c) BA-SEP result.

(d) AML-SMPS result.

Figure 12: Transferring textures inside six designated planar regions on a statue from the first viewpoint to the second viewpoint. Panel (c) shows that the quality of one of the separately estimated homographies was very poor. The results are much better in panel (d) where inter-homography constraints were enforced. Comparing the results in panels (c) and (d) suggest that the poorly estimated homography in panel (c) was incompatible with the remaining five homographies, and that only thanks to enforcing inter-homography constraints that superior results were achieved in panel (d).



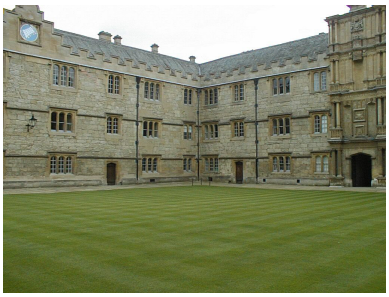
(a) Tea boxes view 1.

(b) Tea boxes view 2.

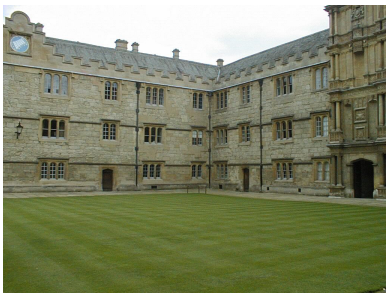
(c) BA-SEP result.

(d) AML-SMPS result.

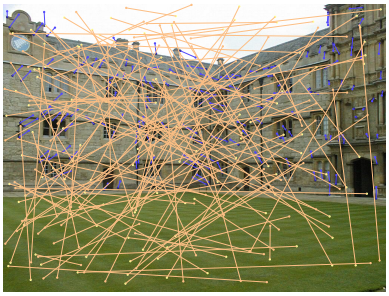
Figure 13: Transferring textures inside six designated planar regions on tea boxes from the first viewpoint to the second viewpoint. In panel (c) the border of the top of the purple tea box does not align properly with the edge of the box. In panel (d) the alignment is substantially better.



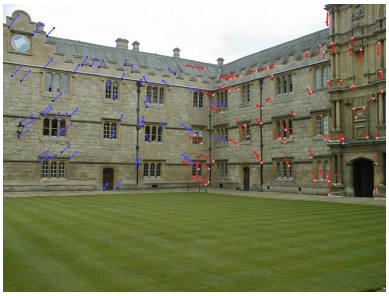
(a) First view of Merton College.



(b) Second view of Merton College.



(c) Flow vectors associated with matched corresponding points (in blue), as well as with outliers (in orange).



(d) Flow vectors associated with two groups of matched inliers resulting from sequential RANSAC.

(e) Result of transferring planar region between (a) and (b) using homography estimated with BA-SEP.

(f) Result of transferring planar region between (a) and (b) using homography estimated with AML-SMPS.

Figure 14: Qualitative results of robust homography estimation with and without consistency constraints. A small planar region containing a window in panel (a) is mapped using an estimated homography to its corresponding region in panel (b). The brightness of the mapped region is increased to make the region more distinguishable. Panel (e) shows that without enforcing consistency constraints the ledge below the window does not align properly. In contrast, panel (f) illustrates that enforcing consistency constraints remedies the misalignment.

11. Conclusion

After giving a comprehensive account of the historical developments in multiple homography estimation, we introduced a new multiple homography estimation method that utilises the dependable Sampson distance and operates directly on corresponding points. In contrast, all prior estimators fit a set of homography matrices to another set of homography matrices. Our experiments attest that jointly estimating multiple homographies yields considerable accuracy gains over separately estimated homographies.

Acknowledgements

This research was partially supported by the Australian Research Council.

References

[1] Baker, S., Datta, A., Kanade, T., 2006. Parameterizing homographies. Tech. Rep. CMU-RI-TR-06-11. Robotics Institute. Carnegie Mellon University, Pittsburgh, PA.

[2] Chen, P., Suter, D., 2009. Rank constraints for homographies over two views: revisiting the rank four constraint. Int. J. Computer Vision 81, 205–225.

[3] Chojnacki, W., Brooks, M.J., 2007. On the consistency of the normalized eight-point algorithm. J. Math. Imaging Vision 28, 19–27.

[4] Chojnacki, W., Brooks, M.J., van den Hengel, A., Gawley, D., 2003. Revisiting Hartley's normalized eight-point algorithm. IEEE Trans. Pattern Anal. Mach. Intell. 25, 1172–1177.

[5] Chojnacki, W., van den Hengel, A., 2011. A dimensionality result for multiple homography matrices, in: Proc. Eighth Int. Conf. Computer Vision, pp. 2104–2109.

[6] Chojnacki, W., van den Hengel, A., 2013. On the dimension of the set of two-view multi-homography matrices. Complex Anal. Oper. Theory 7, 465–484.

[7] Chojnacki, W., Szpak, Z., Brooks, M.J., van den Hengel, A., 2010. Multiple homography estimation with full consistency constraints, in: Proc. Digital Image Computing: Techniques and Applications Conf., pp. 480–485.

[8] Eriksson, A., van den Hengel, A., 2011. Optimization on the manifold of multiple homographies, in: Proc. 12th Int. Conf. Computer Vision Workshops, pp. 242–249.

[9] Harris, C.G., Stephens, M., 1988. A combined corner and edge detector, in: Proc. Fourth Alvey Vision Conference, pp. 147–151.

[10] Hartley, R., 1997. In defense of the eight-point algorithm. IEEE Trans. Pattern Anal. Mach. Intell. 19, 580–593.

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

[12] Kanatani, K., Ohta, N., Kanazawa, Y., 2000. Optimal homography computation with a reliability measure. IEICE Trans. Inf. Syst. E83-D, 1369–1374.

[13] Lütkepol, H., 1996. Handbook of Matrices. John Wiley & Sons, Chichester.

[14] Ma, Y., Soatto, S., Kosecká, J., Sastry, S.S., 2005. An Invitation to 3-D Vision: From Images to Geometric Models. 2nd ed., Springer, New York.

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

[16] Shashua, A., Avidan, S., 1996. The rank 4 constraint in multiple ($\geq$ 3) view geometry, in: Proc. 4th European Conf. Computer Vision, pp. 196–206.

[17] Zelnik-Manor, L., Irani, M., 2002. Multiview constraints on homographies. IEEE Trans. Pattern Anal. Mach. Intell. 24, 214–223.