# A MULTI-OBJECTIVE PARAMETER ESTIMATOR FOR IMAGE MOSAICING

*Tony Scoleri, Wojciech Chojnacki, Michael J. Brooks*

School of Computer Science, Adelaide University, Adelaide, SA 5005
Cooperative Research Centre for Sensor Signal and Image Processing, Mawson Lakes, SA 5095
{tscoleri,wojtek,mjb}@cs.adelaide.edu.au

## ABSTRACT

We consider the image mosaicing task of estimating the homography that defines the transfer function between a pair of overlapping images. A novel homography estimator is proposed that is multi-objective in character in its dealing with a system of underlying equations. Experiments carried out with both synthetic and real images reveal the estimator to be commensurate in accuracy with the Gold Standard maximum likelihood estimator, but with the advantage of being considerably faster.

## 1. INTRODUCTION

A mosaic is constructed by "stitching" together several overlapping images. Such composite images find application in the generation of panoramic views, virtual reality tours (in particular when mosaics cover a 360° view) and super-resolution images. If two overlapping images of a *planar* scene are taken from different camera positions and/or orientations or when a single camera undergoes *pure rotation*, the images are linked via a planar projective transformation, or homography. Accurate computation of homographies has been the subject of much attention [1–6]. This paper proposes a new homography estimator that is multi-objective in character in that it deals with a system of underlying equations. Experimental results indicate that the method estimates homographies with accuracy commensurate with the Gold Standard maximum likelihood method, but in considerably less time.

## 2. HOMOGRAPHY

A homography is a projective mapping between two image planes and, when homogeneous coordinates are used, is described by a $3 \times 3$ matrix. Suppose that left and right image points $\boldsymbol{m} = [u, v, 1]^T$ and $\boldsymbol{m}' = [u', v', 1]^T$ resulting from a projection of a single point in the scene are related by a homography $\boldsymbol{H}^T = [\boldsymbol{h}_1, \boldsymbol{h}_2, \boldsymbol{h}_3]$. Then

$$\begin{aligned} v'\boldsymbol{h}_3^T\boldsymbol{m} - \boldsymbol{h}_2^T\boldsymbol{m} &= 0, \\ \boldsymbol{h}_1^T\boldsymbol{m} - u'\boldsymbol{h}_3^T\boldsymbol{m} &= 0, \\ u'\boldsymbol{h}_2^T\boldsymbol{m} - v'\boldsymbol{h}_1^T\boldsymbol{m} &= 0. \end{aligned} \tag{1}$$

These equations are linearly dependent and only a combination of any two of them is linearly independent. So one has a choice of using either all three equations or any of three pairs of equations to derive a homography estimate of $\boldsymbol{H}$.

Define a single item of data by concatenating the coordinates of a pair of matching points into a vector $\boldsymbol{x} = [u, v, u', v']^T$. Let $\boldsymbol{\theta} = \mathrm{vec}(\boldsymbol{H}^T)$, where vec denotes *vectorisation* [7]. Starting with all three equations, we let

$$\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\theta}) = [f_1(\boldsymbol{x}, \boldsymbol{\theta}), f_2(\boldsymbol{x}, \boldsymbol{\theta}), f_3(\boldsymbol{x}, \boldsymbol{\theta})]^T,$$

where $f_1, f_2$ and $f_3$ are the corresponding expressions on the left-hand side of (1). System (1) can then be succinctly written as $\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\theta}) = \boldsymbol{0}$. Creating a "pure" measurement matrix $\boldsymbol{U}(\boldsymbol{x}) = [\boldsymbol{u}_1(\boldsymbol{x}), \boldsymbol{u}_2(\boldsymbol{x}), \boldsymbol{u}_3(\boldsymbol{x})]$, where

$$\begin{aligned} \boldsymbol{u}_1(\boldsymbol{x}) &= [0, 0, 0, -u, -v, -1, uv', vv', v']^T, \\ \boldsymbol{u}_2(\boldsymbol{x}) &= [u, v, 1, 0, 0, 0, -uu', -vu', -u']^T, \\ \boldsymbol{u}_3(\boldsymbol{x}) &= [-uv', -vv', -v', uu', vu', u', 0, 0, 0]^T, \end{aligned}$$

we have $\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\theta}) = \boldsymbol{U}(\boldsymbol{x})^T\boldsymbol{\theta}$. If only two equations of (1) are employed, $\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\theta})$ is modified to comprise only two components accordingly.

## 3. FUNDAMENTAL NUMERICAL SCHEME

Given a data set $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$, the simplest way to find $\boldsymbol{\theta}$ is to solve the system $\boldsymbol{f}(\boldsymbol{x}_1, \boldsymbol{\theta}) = \cdots = \boldsymbol{f}(\boldsymbol{x}_n, \boldsymbol{\theta}) = \boldsymbol{0}$. The parameter $\boldsymbol{\theta}$ is sought only up to a non-zero scalar factor, so the problem has 8 degrees of freedom. When $n > 8$ and noise is present in the data points, the system has no non-zero solution. However, an *approximated maximum likelihood* estimate of $\boldsymbol{\theta}$, $\widehat{\boldsymbol{\theta}}_{\mathrm{AML}}$, can always be evolved by minimising the *multi-objective* cost function

$$J_{\mathrm{AML}}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n, \boldsymbol{\theta}) = \sum_{i=1}^{n} \boldsymbol{f}(\boldsymbol{x}_i, \boldsymbol{\theta})^T \boldsymbol{C}(\boldsymbol{x}_i, \boldsymbol{\theta})^{-1} \boldsymbol{f}(\boldsymbol{x}_i, \boldsymbol{\theta})$$

where $\boldsymbol{C}(\boldsymbol{x}_i, \boldsymbol{\theta}) = \partial_{\boldsymbol{x}} \boldsymbol{f}(\boldsymbol{x}_i, \boldsymbol{\theta}) \boldsymbol{\Lambda}_{\boldsymbol{x}_i} \partial_{\boldsymbol{x}} \boldsymbol{f}(\boldsymbol{x}_i, \boldsymbol{\theta})^T$ and $\boldsymbol{\Lambda}_{\boldsymbol{x}_i}$ is a symmetric covariance matrix associated with the image measurement $\boldsymbol{x}_i$. The gradient of $J_{\mathrm{AML}}$ with respect to $\boldsymbol{\theta}$, $\nabla_{\boldsymbol{\theta}} J_{\mathrm{AML}}$, vanishes at the minimiser $\widehat{\boldsymbol{\theta}}_{\mathrm{AML}}$. Direct

computation shows that $\nabla_{\boldsymbol{\theta}} J_{\mathrm{AML}} = 2\boldsymbol{X}_{\boldsymbol{\theta}}\boldsymbol{\theta}$, where

$$\begin{aligned}
\boldsymbol{X}_{\boldsymbol{\theta}} &= \sum_{i=1}^{n} \boldsymbol{U}_i \boldsymbol{\Sigma}_i^{-1} \boldsymbol{U}_i^T - (\boldsymbol{\eta}_i^T \otimes \boldsymbol{I}_l)\boldsymbol{B}_i(\boldsymbol{\eta}_i \otimes \boldsymbol{I}_l), \\
\boldsymbol{U}_i &= \boldsymbol{U}(\boldsymbol{x}_i) = [\boldsymbol{u}_1(\boldsymbol{x}_i), \ldots, \boldsymbol{u}_m(\boldsymbol{x}_i)], \quad m \in \{2,3\}, \\
\boldsymbol{B}_i &= \partial_{\boldsymbol{x}_i} \mathrm{vec}(\boldsymbol{U}_i) \boldsymbol{\Lambda}_{\boldsymbol{x}_i} (\partial_{\boldsymbol{x}_i} \mathrm{vec}(\boldsymbol{U}_i))^T, \\
\boldsymbol{\Sigma}_i &= (\boldsymbol{I}_m \otimes \boldsymbol{\theta}^T)\boldsymbol{B}_i(\boldsymbol{I}_m \otimes \boldsymbol{\theta}), \\
\boldsymbol{\eta}_i &= \boldsymbol{\Sigma}_i^{-1} \boldsymbol{U}_i^T \boldsymbol{\theta}
\end{aligned}$$

with $\boldsymbol{I}_l$ the $l \times l$ identity matrix and $\otimes$ *Kronecker product* [7]. Finding $\widehat{\boldsymbol{\theta}}_{\mathrm{AML}}$ reduces then to solving the *variational equation* $\boldsymbol{X}_{\boldsymbol{\theta}}\boldsymbol{\theta} = \boldsymbol{0}$. One algorithm for numerically solving this equation is the *fundamental numerical scheme* (FNS) presented in Figure 1. This scheme was originally proposed in a version adequate for optimisation of a single-objective cost function [8,9]. Importantly, in the case of three principal equations, the computation of $\boldsymbol{X}_{\boldsymbol{\theta}}$ uses the *rank-constrained generalised inverse of rank* 2 of $\boldsymbol{\Sigma}_i$, $(\boldsymbol{\Sigma}_i)_2^-$, [10] instead of $\boldsymbol{\Sigma}_i^{-1}$ to eliminate potential instabilities that might occur for small noise. The scheme is seeded with the *normalised algebraic least-squares* (NALS) *estimate*, $\widehat{\boldsymbol{\theta}}_{\mathrm{NALS}}$. This estimate results from operating the *algebraic least-squares* (ALS) method on Hartley-normalised data [11]. ALS is a simple non-iterative technique employing singular value decomposition.

1. Set $\boldsymbol{\theta}$ to $\widehat{\boldsymbol{\theta}}_{\mathrm{NALS}}$.
2. Repeat:
   (a) Compute the matrix $\boldsymbol{X}_{\boldsymbol{\theta}}$;
   (b) Compute a normalised eigenvector of $\boldsymbol{X}_{\boldsymbol{\theta}}$ corresponding to the eigenvalue closest to zero (in absolute value);
   (c) Take the computed eigenvector for an update of $\boldsymbol{\theta}$;

   until convergence.

**Fig. 1**. Fundamental numerical scheme.

In the way it uses the rank-constrained generalised inverse of $\boldsymbol{\Sigma}_i$, FNS based on three principal equations resembles Kanatani's method of *renormalisation* [4]. The latter method is more complicated than FNS and computes merely an approximate solution to the variational equation.

## 4. EXPERIMENTAL EVALUATION

In this section, we present the results of comparative tests carried out to evaluate the performance of FNS. Three algorithms were used to compute homographies from both synthetic and real image data. The covariances of data were assumed to be the default $4 \times 4$ identity matrix corresponding to isotropic homogeneous noise in image point measurement. The basic estimation methods considered were:

- ALS = Algebraic least-squares,
- FNS = Fundamental numerical scheme,
- GS = Gold Standard.

GS is an advanced method [5] for minimising the *maximum likelihood* cost function $J_{\mathrm{ML}}$ given in (2) below. GS and FNS are both seeded with the NALS estimate of $\boldsymbol{H}$. Execution of FNS was terminated every time the $J_{\mathrm{AML}}$-values of current and updated iterates were sufficiently close.

### 4.1. Synthetic Image Tests

Repeated experiments were performed in order to collect results of statistical significance. In each test 60 coplanar scene points were generated as in Figure 2 and then projected by two cameras onto two respective images of $500 \times 500$ pixels to provide "true" matches. Homogeneous Gaussian noise with standard deviation $\sigma = 1$ pixel was added to each image point and the contaminated pairs were used as input to the three algorithms. For each of 200 experiments, one ML estimate was calculated against four ALS and AML estimates depending on whether all three equations of (1) or a combination of two have been used. The tables below show averages over the 200 random trials.

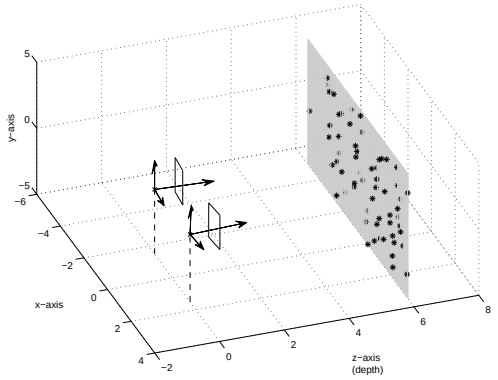


**Fig. 2**. A random planar scene photographed by a pair of cameras with non-parallel optical axes.

Table 1 gives results for a commonly used error measure, the *symmetric transfer error*, given by

$$\sum_i d(\boldsymbol{x}_i, \boldsymbol{H}^{-1}\bar{\boldsymbol{x}}_i')^2 + d(\boldsymbol{x}_i', \boldsymbol{H}\bar{\boldsymbol{x}}_i)^2,$$

where $\bar{\boldsymbol{x}}_i$ and $\bar{\boldsymbol{x}}_i'$ denote true matches and $d(\boldsymbol{x}, \boldsymbol{y})$ is the Euclidean distance between points $\boldsymbol{x}$ and $\boldsymbol{y}$ expressed in inhomogeneous coordinates. The order of the error is comparable to the level of noise introduced in the data points, which matches expectations.

Table 2 considers the $J_{\mathrm{AML}}$ cost function and shows that the $\widehat{\boldsymbol{\theta}}_{\mathrm{AML}}$ estimates achieve very similar cost values.

Perhaps the most critical test comes from using the maximum likelihood function, $J_{\mathrm{ML}}$, defined by

$$\sum_i d(\boldsymbol{x}_i, \hat{\boldsymbol{x}}_i)^2 + d(\boldsymbol{x}_i', \hat{\boldsymbol{x}}_i')^2, \tag{2}$$

| Methods | Homography Equations | | | |
|---|---|---|---|---|
| | $1-2-3$ | $1-2$ | $1-3$ | $2-3$ |
| GS | 2.51 | – | – | – |
| ALS | 2.69 | 2.53 | 2.72 | 2.96 |
| FNS | 2.51 | 2.51 | 2.51 | 2.51 |

**Table 1**. Symmetric transfer error.

| Methods | Homography Equations | | | |
|---|---|---|---|---|
| | $1-2-3$ | $1-2$ | $1-3$ | $2-3$ |
| GS | 111.23 | – | – | – |
| ALS | 121.91 | 112.13 | 123.16 | 136.97 |
| FNS | 111.23 | 111.19 | 111.23 | 110.93 |

**Table 2**. $J_{\mathrm{AML}}$ cost values.

where $\hat{\boldsymbol{x}}_i$ and $\hat{\boldsymbol{x}}_i'$ are corrected data points such that $\hat{\boldsymbol{x}}_i' = \widehat{\boldsymbol{H}}\hat{\boldsymbol{x}}_i$ for all $i$. Inspecting Table 3, we see that FNS estimates produce very competitive cost values in comparison to the GS estimate.

| Methods | Homography Equations | | | |
|---|---|---|---|---|
| | $1-2-3$ | $1-2$ | $1-3$ | $2-3$ |
| GS | 111.20 | – | – | – |
| ALS | 133.17 | 113.51 | 135.57 | 167.01 |
| FNS | 111.21 | 111.20 | 111.28 | 111.24 |

**Table 3**. $J_{\mathrm{ML}}$ cost values.

Finally, a timing test is presented in Table 4. Unsurprisingly, GS turns out to be by far the slowest of the methods. While it may be speeded up via the incorporation of sparse-matrix techniques, it is intrinsically slow given the high-dimensionality of its search strategy.

| Methods | Homography Equations | | | |
|---|---|---|---|---|
| | $1-2-3$ | $1-2$ | $1-3$ | $2-3$ |
| GS | 19.04 | – | – | – |
| ALS | 0.10 | 0.10 | 0.10 | 0.09 |
| FNS | 4.59 | 2.54 | 2.53 | 2.55 |

**Table 4**. Execution time in seconds.

### 4.2. Real Image Tests

Two sequences of images were acquired by rotating a camera about its centre. The images are registered using planar homographies from FNS and composed into single panoramic mosaics, shown in Figures 3 and 4.

#### *4.2.1. Convention Centre Sequence*

Tables 5 and 6 give test averages for the two homography estimations. FNS and GS give the best results and are essentially inseparable. FNS is much faster than GS despite the relatively low number of matches.

| Methods | Homography Equations | | | |
|---|---|---|---|---|
| | $1-2-3$ | $1-2$ | $1-3$ | $2-3$ |
| GS | 89.98 | – | – | – |
| ALS | 93.93 | 96.39 | 97.27 | 94.64 |
| FNS | 89.98 | 89.97 | 89.99 | 89.97 |

**Table 5**. $J_{\mathrm{AML}}$ residuals.

| Methods | Homography Equations | | | |
|---|---|---|---|---|
| | $1-2-3$ | $1-2$ | $1-3$ | $2-3$ |
| GS | 15.65 | – | – | – |
| ALS | 0.26 | 0.08 | 0.09 | 0.09 |
| FNS | 5.16 | 2.19 | 2.20 | 2.21 |

**Table 6**. Execution time in seconds.

#### *4.2.2. War Memorial Sequence*

Three homography estimations were needed in this case and average results are presented in Table 7. Again FNS and GS produce close results with FNS being significantly faster.

| Methods | $J_{\mathrm{AML}}$ | Time (sec) | Iterations |
|---|---|---|---|
| ALS | 3339.07 | 0.13 | 1 |
| FNS | 164.08 | 6.30 | 5 |
| GS | 164.07 | 24.36 | 5 |

**Table 7**. $J_{\mathrm{AML}}$ residual, execution time and iterations. Estimation was carried out using all three equations in (1).

## 5. CONCLUSION

A newly-developed parameter estimation method was proposed for problems in which the relationship between parameters and image data is expressed as a system of equations. Its performance was demonstrated on the estimation of homographies used to build panoramic mosaics. In this application, it is possible to base the estimation process on two or three equations. Our tests revealed that selecting the first two equations provides advantages in terms of speed with no effect on the accuracy of the solution. In general, when compared to the much slower NALS-seeded GS, FNS gives almost identical results, both in terms of $J_{\mathrm{AML}}$ residual and GS's MLE cost function residual.

## 6. REFERENCES

[1] M. Irani, P. Anandan, and S. Hsu, "Mosaic based representations of video sequences and their appli-

**Fig. 3**. **Convention centre**. Each image is $640 \times 426$. Consecutive pairs of images had $50$ and $54$ matches respectively.

**Fig. 4**. **War memorial**. Each image is $429 \times 640$. Consecutive pairs of images had $35, 80$ and $67$ matches respectively.


cations," in *Proc. Fifth Int. Conf. Computer Vision*, 1995, pp. 605–611.

[2] H-Y. Shum and R. Szeliski, "Construction and refinement of panoramic mosaics with global and local alignment," in *Proc. Sixth Int. Conf. Computer Vision*, 1998, pp. 953–958.

[3] I. Zoghlami, O. Faugeras, and R. Deriche, "Using geometric corners to build a 2D mosaic from a set of images," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 1997, pp. 420–425.

[4] K. Kanatani and N. Ohta, "Accuracy bounds and optimal computation of homography for image mosaicing applications," in *Proc. Seventh Int. Conf. Computer Vision*, 1999, pp. 73–78.

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

[6] D. Capel and A. Zisserman, "Automated mosaicing with super-resolution zoom," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, June 1998, pp. 885–891.

[7] P. Lancaster and M. Tismenetsky, *The Theory of Matrices with Applications*, Academic Press, San Diego, 2nd edition, 1985.

[8] 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.

[9] W. Chojnacki, M. J. Brooks, A. van den Hengel, and D. Gawley, "FNS, CFNS and HEIV: A unifying approach," *J. Math. Imaging and Vision*, vol. 23, no. 2, pp. 175–183, 2005.

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

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