GLOBAL UNCERTAINTY IN EPIPOLAR GEOMETRY VIA FULLY AND PARTIALLY DATA-DRIVEN SAMPLING

C. Engels, D. Nistér

Center for Visualization and Virtual Environments, Dept. of Computer Science,
University of Kentucky, Lexington, KY 40507, USA
(engels@vis, dnister@cs).uky.edu

KEY WORDS: Relative orientation, Stucture from Motion, Epipolar Geometry

ABSTRACT:

In this paper we explore the relative efficiency of various data-driven sampling techniques for estimating the epipolar geometry and its global uncertainty. We explore standard fully data-driven methods, specifically the five-point, seven-point, and eight-point methods. We also explore what we refer to as partially data-driven methods, where in the sampling we choose some of the parameters deterministically. The goal of these sampling methods is to approximate full search within a computionally feasible time frame. As a compromise between fully representing posterior likelihood over the space of fundamental matrices and producing a single estimate, we represent the uncertainty over the space of translation directions. In contrast to finding a single estimate, representing the posterior likelihood is always a well-posed problem, albeit an often computionally challenging one. Furthermore, this representation yields an estimate of the global uncertainty, which may be used for comparison between differing methods.

1 . INTRODUCTION

Estimation of the relative orientation between two images is an extensively researched subject in computer vision. Many methods have been proposed and the state of the art is now quite elaborate and mature. In our view, the main requirements on an estimation method are that it

- Is accurate (both locally and globally)
- Is robust
- Is computationally efficient
- Can exploit all constraints, exact and approximate
- Gives a truthful uncertainty estimate (local and global)

It is widely accepted that accuracy is best achieved with iterative refinement, called bundle adjustment [24], according to a cost function that is derived from a realistic model of the problem. However, bundle adjustment is dependent on an initial starting point and only achieves what we refer to as local accuracy, which is the ability to precisely pinpoint a local minimum of the cost function. Perhaps even more important and challenging in computer vision is to, insofar as possible, achieve global accuracy, which is the ability to reliably locate the global minimum of the cost function.

Robustness is achieved by using an appropriate data model that includes data distortions and outliers. Computational efficiency is always desirable, although the requirements are more stringent in some applications than others. It is likewise desirable to use all available constraints, such as camera calibration information.

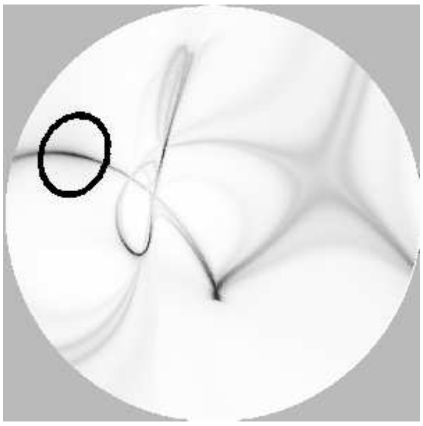

Figure 1: We derive an uncertainty representation for epipolar geometry parameterized by the epipole in the first image. The figure shows an example of the uncertainty representation when the number of point correspondences is too low, leading to intricate patterns of probability mass. The global maximum is circled, but notice the multiple peaks captured by the representation.

Gauging the uncertainty is important, since without a notion of how likely it is that the estimate at hand is in error, it is very hard to take any useful action based upon it. It is best-practice to gauge local uncertainty around an estimate by analyzing the local shape of the cost function around the minimum. However, such an uncertainty measure only makes sense if the global minimum was truly found. Moreover, it assumes that the cost function is unimodal and nicely behaved. This is seldom the case. Due to outliers, noise, the nonlinear nature of the problem, planar scenes and small translation, the cost function may lack a clear global minimum or have several throughs of complicated shape.

Therefore, to assess global uncertainty, an estimation method should ideally provide a representation of the posterior probability distribution over all the regions of parameter space where the probability is significant.

For strong data, producing a single estimate is possible. However, there will always be situations with ambiguous data, in which obtaining a single estimate is essentially an ill-posed problem. On the other hand, provided we have selected an appropriate data model, representing the posterior distribution is always a well-posed problem. Representing the posterior may be computationally difficult, but it is well-posed for any input data.

Our approach draws upon background material in probabilistic Bayesian frameworks and multiple view geometry. Due to space limitations, we by necessity have to assume that the reader has some familiarity with these concepts. The interested reader is referred to [4, 5, 20] for the former and [6, 9, 15] for the latter.

2 . APPROACH

Ideally, we would like to evaluate the likelihood $p(d|w)$ for all possible world states w to derive our representation for the posterior distribution. However, it is impractical to perform full search over a high-dimensional space (in this case five or more dimensions). Such a complete representation would also be unmanageable for a module that needs to use the results for further computation or decision making.

To reach an efficient representation of the likelihood, we will rely on the following observation: If the epipole in the first image is known, the remaining parameters of the fundamental matrix (simply rotation in an uncalibrated setting) are uniquely determined unless all the points from the point correspondences and the epipole lie on a common conic in the second image.

Thus it is natural to represent the likelihood with an explicit representation indexed by the translation direction (epipole in the first image).

The usefulness of treating the translation and rotation differently has been understood by many authors and exploited in different ways, see for example [10, 3, 18, 1]. It is also closely related to the highly popular plane-plus-parallax approach [11, 14, 21, 23, 13], where one relies on the existence of a dominant homography and solves for that in order to guide the search for the translation direction.

3 . DATA DRIVEN SAMPLING

As argued above, we can not search the likelihood over the whole parameter space. Several authors have noted that it can be much more efficient to search the parameter space with data-driven hypothesis generators [2, 25]. We will use hypothesis generation in a similar manner as in RANSAC [7], where minimal samples of correspondences are randomly chosen from the whole set of correspondences. A minimal sample contains the smallest number of data points that will determine the geometric relation up to a finite number of solutions. The samples are made minimal to minimize the risk of including devastating outliers. In this case, a minimal sample contains seven correspondences for the fundamental matrix and five for the essential matrix. We refer to this as fully data-driven sampling, since the correspondences ideally should determine the fundamental matrix. We will also use partially data-driven sampling, where for a given translation direction, we take samples containing the smallest number of correspondences that will determine the remaining parameters of the fundamental matrix up to a finite number of solutions. The samples contain five correspondences to determine the fundamental matrix in the uncalibrated case and three correspondences to determine the essential matrix given translation direction in the calibrated case.

4 . REPRESENTATION

If we can derive an accurate representation of the data likelihood $p(d|w)$ it can be converted into a representation of the posterior by multiplying with the prior. The representation of the posterior can then support any inferences we wish to make based on the data.

We consider the world state w to be represented by the fundamental matrix F and the data d to be represented by all the point correspondences, denoted by X. Bayes' rule then becomes

$$p(F|X) \propto p(X|F)p(F). \tag{1}$$

We store the hypotheses for the fundamental matrix in a two-dimensional array indexed by epipole in the first image. Our goal is to find the best fundamental matrix hypothesis for each cell of the array and the integral likelihood in each cell. Let $\Omega(e)$ denote the set of all fundamental matrices with the epipole e in the first image. The desired output from our approach is

$$F_{opt}(e) = \begin{array}{c} arg\ max \\ F \in \Omega(e) \end{array} p(X|F) \tag{2}$$

and

$$f(e) = \int_{F \in \Omega(e)} p(X|F)dF. \tag{3}$$

for all values of the epipole e. The latter can be computed by a Laplace approximation around the former.

Along the lines of our above motivation, it is assumed that the likelihood $p(X|F)$ has a unique narrow peak in $\Omega(e)$. By assuming that the prior $p(F)$ is smooth in comparison to the extent of the peak, the user of the output can make the approximation

$$p(e|X) \propto \int_{F\in\Omega(e)} p(X|F)p(F)dF \approx p(F_{opt}(e))f(e). \quad (4)$$

In a similar manner, most inferences that one may wish to make based on the data has to do with an integral of some function $g(F)$ times the posterior likelihood. Such integrals

$$\int_e \int_{F\in\Omega(e)} g(F)p(F|X)dFde \quad (5)$$

can be approximated as

$$\frac{\int_e g(F_{opt}(e))p(F_{opt}(e))f(e)de}{\int_e p(F_{opt}(e))f(e)de}. \quad (6)$$

The advantage is that the inferences can be made outside the relative orientation module with any choice of prior $p(F)$ using only $F_{opt}(e)$, $f(e)$ and easy two-dimensional integrals.

If this can be done efficiently and reliably, inferences can be made in an application-dependent manner based on the resulting representation, without major alterations to the core of the computer vision algorithm.

4.1 Prior Likelihood

In the simplest case, the prior likelihood $p(F)$ is set to uniform. In some cases we may have more prior information. For example, if we are calibrating a stereo-head, we typically have approximate knowledge of the location of the epipole and also of the relative rotation. We may also work in the uncalibrated setting, but use the prior to put approximate constraints on the calibration.

4.2 Posterior Likelihood

We use a Sampson approximation (see [9]):

$$s(x, x', F) = \frac{(x'^{\top}Fx)^2}{(Fx)_1^2 + (Fx)_2^2 + (x'^{\top}F)_1^2 + (x'^{\top}F)_2^2} \quad (7)$$

where the homogeneous coordinates for the points are assumed to be normalized such that their last coordinates are one. It approximates the squared sum of magnitudes of the smallest perturbation required to bring the image point correspondence $x \leftrightarrow x'$ into agreement with the epipolar geometry described by the fundamental matrix ($x'^{\top}Fx = 0$). This approximation has been found superior to symmetric epipolar distance and other approximations of similar computational complexity [27].

We model our data likelihood as

$$p(X|F) \propto \left(\prod_{i=1}^{N} \sigma^2(\sigma^2 + s(x_i, x'_i, F))^{-1}\right)^{N^{-k}}, \quad (8)$$

where σ is a scale parameter, which we typically set to one pixel of a CIF image (352×288), N is the number of point correspondences, and $0 \leq k \leq 1$. We determine the value of k experimentally in section 6.4. We have also tried the standard way of assuming that the reprojection errors are conditionally independent given the world configuration ($k = 0$), dogmatically leading to a product of many independent factors, where each factor is related to a single point correspondence. However, we have found that although this produces sensible peak locations of the likelihood, it leads to an unrealistically rapid fall-off around the likelihood peak, resembling a delta-function and not a realistic model of any practical situation.

5 . HYPOTHESIS GENERATORS

The hypothesis generators we use in our experiments are:

- 5-Point (Calibrated)
- 7-Point (Uncalibrated)
- 8-Point (Uncalibrated)
- 3-Point+Epipole (Calibrated)
- 5-Point+Epipole (Uncalibrated)

For fully data-driven sampling in the calibrated case, we use the 5-point method (5pt)[16]. In the uncalibrated case, we use the 7-point (7pt) method and the 8-point (8pt) method [9].

The 3-point+epipole (3pt+e) and 5-point+epipole (5pt+e) methods are partially data-driven generators. The former was presented in [17]. It uses the point constraints and the known epipole to restrict the essential matrix to a 3-dimensional linear space. The calibration constraints are then added, leading to two conics that are intersected, which yields four solutions. This method can be carried out extremely fast in closed form. The latter is related to a classical result, which is that given five point correspondences, the epipoles correspond by a fifth-degree Cremona mapping, also discussed in [26]. This method gives a unique solution. It can for example be implemented by stacking linear constraints from the point correspondences and the known epipole into an 8×9 matrix, subsequently extracting the unique nullvector.

6 . EXPERIMENTS

6.1 Construction of the Likelihood Image

To determine the uncertainty of an estimated epipole, we first computed a quantized posterior likelihood over a hemisphere of epipoles. The sign of the epipole can only be determined using cheirality [9], which we do not enforce. We mapped the hemisphere onto a 300×300 image. In each cell, we computed the optimal fundamental matrix with translation direction in the cell. In the cases of the partially data-driven methods, we deterministically sampled the translation direction over all quantized translations. In the fully data-driven methods, the translation direction was determined by the generated hypothesis. We sampled the entire epipolar space, or about 70000 cells, in multiple sweeps, using random sets of point correspondences for each sample. In the partially data-driven methods, a small perturbation in the translation was added within each cell to more fully represent possible fundamental matrices.

We explored the likelihood images for both synthetic and real data. In the synthetic case, images with known relative orientation were created with a scene volume of random points. The image points were then perturbed with Gaussian noise equivalent to one pixel of a CIF image. Finally, outliers were simulated by uniformly scattering a percentage of the image points in one image. For real data, we tracked Harris corners, using normalized correlation for matching. The camera was calibrated in order to compare calibrated and uncalibrated methods.

6.2 Convergence of the Likelihood

We investigated how quickly each method converges to the likelihood over the entire hemisphere. A straightforward measure of the error in the estimated likelihood is given by

$$error = \int_e (p(e) - \hat{p}(e))de, \tag{9}$$

where p is the true likelihood and $\hat{p}$ is the estimated likelihood. Ideally, a full search over the space of fundamental matrices would be used to create p. Since this is infeasible, we approximated the true likelihood as the maximum found using all five tested methods in an extremely long computation. The final image, shown on the top left of Figure 2, was created with 1000 sweeps, or about 7×10^7 samples per method.

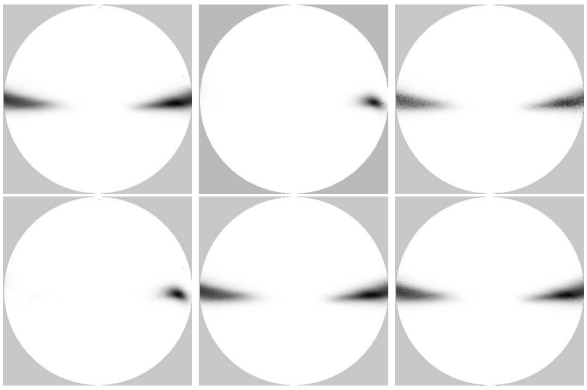

Figure 2: Posterior likelihood images of a scene with sideways translation over 1000 sweeps of the epipolar space. From left to right, top to bottom: true likelihood; 3pt+e method; 5pt+e method; 5pt method; 7pt method; 8pt method.

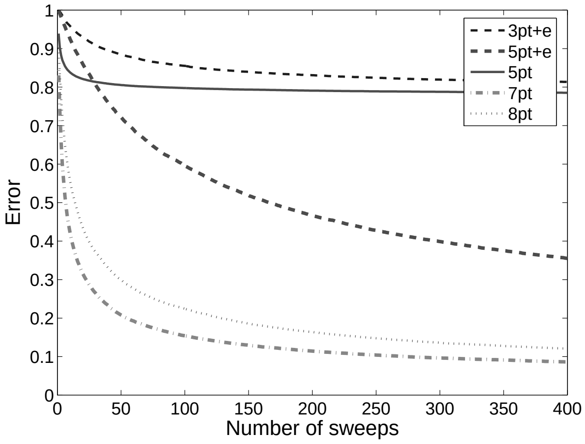

Figure 3: Comparison of convergence rates for the various hypothesis generation methods. Hypothesis generation times are not taken into account.

6.2.1 Comparison of Partially and Fully Data-Driven Methods

We compared the methods by examining the rate of convergence to the likelihood. Since the uncalibrated methods create hypotheses from the space of fundamental matrices, while the calibrated methods generate hypotheses from the more restricted space of essential matrices, the uncalibrated methods uncover a greater probability mass. Because we calibrated the image points, the true solution is an essential matrix, so the mass uncovered by the uncalibrated methods may be overestimated.

We sampled with all methods simultaneously and recorded the errors. Because several methods produce multiple solutions, it was important to ensure that the methods had equivalent numbers of samples. For the 3pt+e and 7pt methods, we disambiguated the solutions by scoring one additional point correspondence and choosing the hypothesis with the highest single point likelihood. For the 5pt method, which may produce up to 10 real solutions representing extra potentially valid solutions such as planar ambiguities, we stored the hypotheses and computed the likelihood of one hypothesis per sampling round.

As seen in Figure 3, the fully data-driven uncalibrated methods explore the greatest probability mass early in the computation, while the 5pt+e method slowly converges to the same value. The calibrated methods converge to a different posterior likelihood, although the fully data-driven method again converges faster than the partially data-driven method.

6.3 Estimation of Confidence Intervals

Once we have the posterior likelihood, we create confidence intervals by finding the global maximum in the posterior likelihood and measuring the fraction of the probability mass that lies within a certain distance of the max-

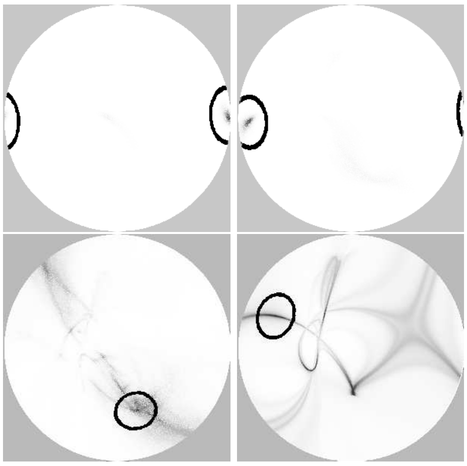

Figure 4: Examples of confidence intervals in an image sequence with a leftward translation. From left to right and top to bottom, the respective probability masses within each circled confidence interval are: 0.865, 0.567, 0.204, 0.065.

imum. That is, we start from a maximal acceptable distance, which then in turn determines the confidence level. Typically, we used a distance of 5 degrees on the sphere. Figure 4 shows examples of confidence intervals in likelihood images. The top two images represent cases with many inlier point correspondences. The bottom left image represents a case with relatively few correspondences and low stability. The bottom right image represents a case that has a critically small number of correspondences. However, these deficiencies are apparent in the representation, due to the small probability mass within the confidence intervals.

6.4 Verification of Confidence Interval

If we construct confidence intervals and collect statistics on the confidence level needed to capture the true epipole, this confidence level should ideally be a uniformily distributed random variable. To explore the sensitivity of our confidence intervals to discrepencies between the assumed data model and the actual data model, we use synthetic data along with our cost function, and measure the deviation from uniform distribution. A synthetic scene with 30% outliers and a known epipole was created.

A 100×100 likelihood image was created using 10 sweeps of the 5pt+e method, and the probability mass required to capture the true epipole was recorded. This was repeated 500 times, and the cumulative distribution function of the mass fractions was plotted. A sublinear cdf indicates overconfidence, while a superlinear cdf indicates underconfidence.

We found the best value for k from Equation (8) to be approximately $1/2$. As seen in Figure 5, this achieves a balance in the confidence estimates, while $k = 1$ leads to underconfidence and $k = 0$ to overconfidence, with a highly peaked likelihood.

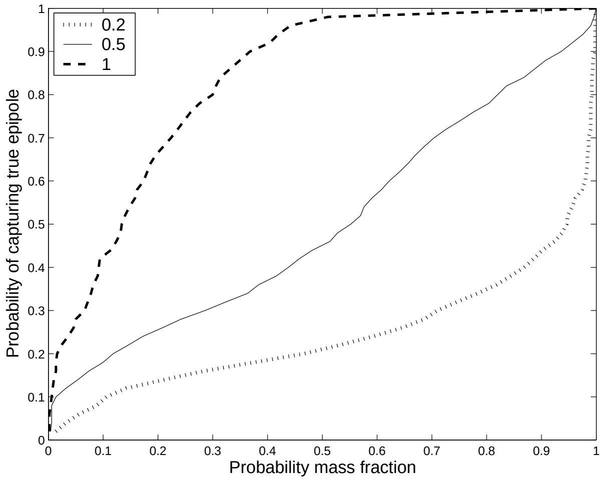

Figure 5: Cumulative distribution functions of confidence levels for varying values of k. Note that $k = 0.5$ most closely matches a uniform random variable.

6.5 Finding Optimal Baseline in an Image Sequence

As a practical test of inference with our uncertainty representation, we aim to find a pair of frames in an image sequence that results in the best possible 3-D reconstruction of a scene. To accomplish this, we search for an optimal baseline between camera positions, such that we have a large translation required for accurate reconstruction while still maintaining a reasonable number of inlier point correspondences. Obtaining a confidence interval between different pairs of images allows us to choose the pair that has the greatest mass fraction in a fixed-size confidence interval, i.e. leads to the greatest confidence in capturing the true epipole to within a fixed angle. In our experiment, we used a video sequence with a camera undergoing sideways translation relative to the scene. We considered all the image pairs that include the first image (frame 0), leaving the second image frame for selection. Figure 6 shows the resulting fractions of the probability mass for each frame. The peak is located at a reasonable baseline spanning four frames. The sharp decline in mass after frame 7 is caused by falling below an acceptable number of inlier point correspondences.

7 . CONCLUSION

We have presented a framework for epipolar geometry estimation that draws upon both multiple view geometry and statistics. The central theme is to derive a representation that faithfully represents the posterior likelihood globally. This is accomplished with a representation parameterized by epipole location in the first image. We have explored

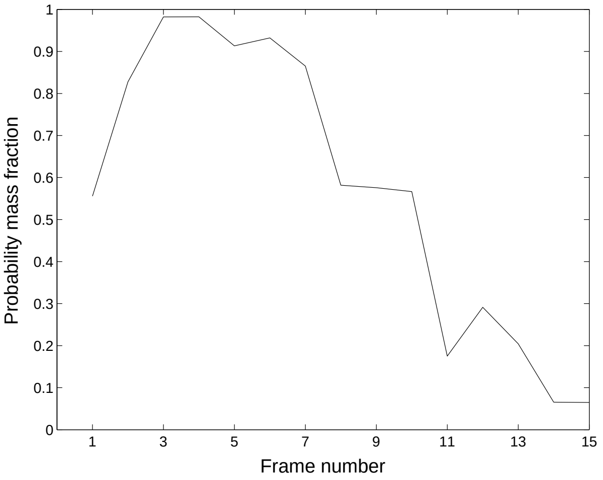

Figure 6: Probability mass lying within confidence interval over a series of video frames.

the efficiency of various fully and partially data-driven hypothesis generators in deriving the representation. We have presented experiments with confidence regions derived from our representation and we have experimentally validated the confidence regions through experiments with synthetic data. This was done by investigating the distribution of the confidence level needed to capture the true epipole in the confidence region, which should ideally be a uniformly distributed random variable. Finally, we have shown on real data how the uncertainty representation helps us accomplish inference tasks that are otherwise difficult, such as selecting which baseline to use when initializing automatic reconstruction from a video-sequence.

References

[1] P. Baker, R. Pless, C. Fermüller and Y. Aloimonos. Eyes from Eyes. *SMILE 00*, Springer-Verlag, p.204-217, 2001.

[2] P. Chang, M. Hebert. Robust tracking and structure from motion through sampling based uncertainty representation. *ICRA*, 2002.

[3] A. Chiuso, R. Brockett, S. Soatto. Optimal Structure from Motion: Local Ambiguities and Global Estimates. *IJCV*, vol. 39 n.3, p.195-228, Sept./Oct. 2000.

[4] J. Clark and A. Yuille. *Data Fusion for Sensory Information Processing Systems*. Kluwer Academic Publishers, ISBN 0-7923-9120-9, 1990.

[5] A. Doucet, N. de Freitas, N. Gordon, eds. Sequential Monte Carlo Methods In Practice. Springer-Verlag New York, 2001.

[6] O. Faugeras. Three-Dimensional Computer Vision. MIT Press, 1993.

[7] M. Fischler and R. Bolles. Random Sample Consensus: a Paradigm for Model Fitting with Application to Image Analysis and Automated Cartography. *Commun. Assoc. Comp. Mach.*, 24:381-395, 1981.

[8] W. Förstner. Uncertainty and Projective Geometry. To appear in: *Handbook of Computational Geometry for Pattern Recognition, Computer Vision, Neurocomputing and Robotics*

[9] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN 0-521-62304-9, 2000.

[10] B.K.P. Horn. Relative Orientation. *IJCV*, vol. 4, pp. 59-78, 1990.

[11] M. Irani, B. Rousso, and S. Peleg. Recovery of Ego-motion Using Region Alignment. *PAMI*, 19(3):268-272, March 1997.

[12] M. Isard and A. Blake. Condensation: conditional density propogation for visual tracking. IJCV, (1), 1998.

[13] R. Kaucic, R. Hartley, N. Dano. Plane-based Projective Reconstruction. *ICCV*, 2001.

[14] R. Kumar, P. Anandan, and K. Hanna. Shape Recovery from Multiple Views: a Parallax Based Approach. *ARPA Image Understanding Workshop*, November 1994.

[15] Y. Ma, S. Soatto, J. Košecká, S. Sastry. An Invitation to 3-D Vision: From Images to Geometric Models. Springer-Verlag New York, 2004.

[16] D. Nistér. An Efficient Solution to the Five-Point Relative Pose Problem. *PAMI*, 26(6):756-770, June 2004.

[17] D. Nistér and F. Schaffalitzky, What do Four Points in Two Calibrated Images Tell Us About the Epipoles?, *ECCV*, Springer Lecture Notes on Computer Science 3022:41-57, 2004.

[18] J. Oliensis. The Least-Squares Error for Structure from Infinitesimal Motion. *ECCV*, May 2004.

[19] G. Qian, R. Chellappa. Structure from Motion Using Sequential Monte Carlo Methods. *ICCV*, 2001.

[20] J. Ó Ruanaidh, W. Fitzgerald. Numerical Bayesian Methods Applied to Signal Processing. Springer-Verlag New York, 1996.

[21] R. Szeliski and P.H.S. Torr. Geometrically Constrained Structure from Motion : Points on Planes. *SMILE98*, pages 171-186, 1998.

[22] P.H.S. Torr; C. Davidson. IMPSAC: Synthesis of Importance Sampling and Random Sample Consensus. *IPAMI*, October 2002.

[23] B. Triggs. Plane+Parallax, Tensors and Factorization. *ECCV*, 2000.

[24] B. Triggs, P. McLauchlan, R. Hartley and A. Fitzgibbon. Bundle Adjustment - a Modern Synthesis. *Springer Lecture Notes on Computer Science*, Springer Verlag, 1883:298-375, 2000.

[25] Z. Tu, S. Zhu, H. Shum Image Segmentation by Data Driven Markov Chain Monte Carlo. *ICCV*, 2001.

[26] T. Werner, Constraints on Five Points in Two Images, *CVPR*, Volume 2, pp. 203-208, 2003.

[27] Z. Zhang. Determining the Epipolar Geometry and its Uncertainty: A Review. *IJCV*, March 1998.