

Estimating Global Uncertainty in Epipolar Geometry for Vehicle-Mounted Cameras

David Nistér and Christopher Engels

Center for Visualization and Virtual Environments, University of Kentucky
1 Quality St. Suite 800, Lexington, KY 40507

ABSTRACT

We present a method for estimating the global uncertainty of epipolar geometry with applications to autonomous vehicle navigation. Such uncertainty information is necessary for making informed decisions regarding the confidence of a motion estimate, since we must otherwise accept the estimate without any knowledge of the probability that the estimate is in error. For example, we may wish to fuse visual estimates with information from GPS and inertial sensors, but without uncertainty information, we have no principled way to do so. Ideally, we would perform a full search over the 7-dimensional space of fundamental matrices to yield an estimate and its related uncertainty. However, searching this space is computationally infeasible. As a compromise between fully representing posterior likelihood over this space and producing a single estimate, we represent the uncertainty over the space of translation directions in a calibrated framework. In contrast to finding a single estimate, representing the posterior likelihood is always a well-posed problem, albeit an often computationally challenging one. Given the posterior likelihood, we derive a confidence interval around the motion estimate. We verify the correctness of the confidence interval using synthetic data and show examples of uncertainty estimates using vehicle-mounted camera sequences.

Keywords: Structure from Motion, Relative Orientation, Epipolar Geometry

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,¹ 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

Further author information: D.N.: E-mail: dnister@cs.uky.edu C.D.E.: E-mail: engels@vis.uky.edu

others. It is likewise desirable to use all available constraints, such as camera calibration information.

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.

2. PRELIMINARIES

2.1. The Bayesian Paradigm

According to Bayes' rule, the posterior probability $p(w|d)$ for the world state w given the data d is

$$p(w|d) = \frac{p(d|w)p(w)}{p(d)}. \quad (1)$$

If we accept the Bayesian paradigm and the model embodied in the world parameterization w , prior probability $p(w)$ and data likelihood $p(d|w)$, this governs how the prior probability $p(w)$ should be updated into posterior probability $p(w|d)$ in response to the data. The normalization factor $p(d)$ can be expanded into the integral

$$p(d) = \int p(d|w)p(w)dw. \quad (2)$$

Working numerically, it is straightforward to normalize any distribution in this manner. We will therefore suppress the scale factor and write Bayes' rule as

$$p(w|d) \propto p(d|w)p(w), \quad (3)$$

where \propto denotes equality up to a scale factor independent of w . Our goal is to compute a representation of the posterior probability $p(w|d)$ over all world states w that is as truthful as possible. 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.

2.2. Data Fusion

If there are several data sources with conditionally independent data likelihoods $p(d_1|w), \dots, p(d_n|w)$ Bayes' rule generalizes to

$$p(w|d) \propto p(w) \prod_{i=1}^n p(d_i|w). \quad (4)$$

Each data source contributes a factor $p(d_i|w)$. Computing the factor $p(d_i|w)$ separately for each data source and then computing the posterior through Equation 4 is called weakly coupled data fusion.² One can also leverage the values of one factor to compute the representation of another more efficiently. This can be done in a sequential or iterative fusion and is called strongly coupled data fusion.

2.3. Decision Theory

We choose a function $l(D, w)$ that represents the loss of making the decision D when the true state of the world is w . The conditional risk $r(D|d)$ of making the decision D given the data d is

$$r(D|d) = \int l(D, w)p(w|d)dw. \quad (5)$$

We wish to minimize it in order to make the optimal decision

$$D_{opt}(d) = \underset{D}{arg \ min} \ r(D|d). \quad (6)$$

If we accept the loss function $l(D, w)$, this is the best decision we can make and $r(D_{opt}(d)|d)$ is the expected loss associated with making it.

2.4. Epipolar Geometry

Image points are represented by homogeneous 3-vectors x and x' in the first and second view, respectively. World points are represented by homogeneous 4-vectors X . A perspective view is represented by a 3×4 camera matrix P indicating the image projection $x \sim PX$, where \sim denotes equality up to scale. A view with a finite projection centre can be factored into $P = K[R | t]$, where K is a 3×3 upper triangular calibration matrix holding the intrinsic parameters and R is a rotation matrix. Let the camera matrices for the two views be $K_1[I | 0]$ and $P = K_2[R | t]$. Let $[t]_{\times}$ denote the skew symmetric matrix

$$[t]_{\times} = \begin{bmatrix} 0 & -t_3 & t_2 \\ t_3 & 0 & -t_1 \\ -t_2 & t_1 & 0 \end{bmatrix} \quad (7)$$

so that $[t]_{\times} x = t \times x$ for all x . Then the fundamental matrix is

$$F \equiv K_2^{-\top} [t]_{\times} R K_1^{-1}. \quad (8)$$

The fundamental matrix encodes the well known coplanarity, or epipolar constraint

$$x'^{\top} F x = 0. \quad (9)$$

If K_1 and K_2 are known, the cameras are said to be calibrated. In this case, we can always assume that the image points x and x' have been premultiplied by K_1^{-1} and K_2^{-1} , respectively, so that the epipolar constraint simplifies to

$$x'^{\top} E x = 0, \quad (10)$$

where the matrix $E \equiv [t]_{\times} R$ is called the essential matrix. Any rank-2 matrix is a possible fundamental matrix. An essential matrix has the additional property that the two non-zero singular values are equal. This leads to the following important cubic constraints on the essential matrix, adapted from:³⁻⁶

THEOREM 2.1. *A real non-zero 3×3 matrix E is an essential matrix if and only if it satisfies the equation*

$$E E^{\top} E - \frac{1}{2} \text{trace}(E E^{\top}) E = 0. \quad (11)$$

3. MOTIVATION AND APPROACH

Ideally, we would like to evaluate the data 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 the module that needs to use the results for inferencing. Instead, we need to rely on the fact that if the data is to be useful in decision making, it should concentrate the posterior probability mass to a limited region. We can then exploit this property by finding a way to traverse only the significant mass using importance sampling.⁷⁻⁹ The spread of the posterior may be characterized using its entropy

$$H = - \int p(w|d) \log p(w|d) dw. \quad (12)$$

Clearly, if the posterior was nearly uniform over a high-dimensional space, we would have a highly difficult computational problem without a very useful answer. In this case, the goal is for the representation to capture that the entropy is high, indicating the data may not be useful.

To reach an efficient representation of the likelihood, we will rely on the following observation: If we know the translation direction (epipole in the first image), finding the remaining parameters of the fundamental matrix is typically stable, even for problems that are otherwise degenerate. This typically holds in both calibrated and uncalibrated situations. On the other hand, it is quite common to encounter situations where finding the translation direction is difficult even given the correct rotation. If the translation magnitude is small compared to the distance to the scene seen in the images, any translation direction will seem to fit the data. Another situation where any translation direction is possible is with a planar scene in the uncalibrated setting. The homography induced by the plane can be adjoined with any epipole in the first image, as long as the epipole in the second image is chosen such that the epipoles correspond by the homography. With known calibration, the ambiguity with a planar scene is reduced from two degrees of freedom to a two-fold ambiguity, with only the true motion and a false solution corresponding to reflection of one of the camera centres across the plane. The planar scene is an example of a so-called critical surface. More generally, the critical surfaces are ruled quadrics through the two projection centres. If the scene and camera configuration is critical, there are multiple solutions for the camera motion. If the configuration is close to critical, the solution is unstable. It is also common with bas-relief ambiguity, where it is hard to disambiguate translation from rotation.

The above examples will defeat any solver that insists on selecting a unique solution. Note however that in all of the above examples, if we know the translation direction, the rotation (or in the uncalibrated case the five remaining parameters of the fundamental matrix) is still well determined. This can in fact be put in more rigorous terms:

THEOREM 3.1. *If the epipole in the first image is known, the remaining parameters of the fundamental matrix are uniquely determined unless all the points from the point correspondences lie on a conic in the first image.*

Proof: If one of the point correspondences include the epipole in the first image, the epipole in the second image is immediately determined by that point correspondence, and the epipolar line homography is then also uniquely determined since the points in the second image can not lie on two lines through the epipole (if they did, they would form a degenerate conic). Hence, we can assume that no point in the first image coincides with the epipole. Then, the cross-ratios based at the epipole in the first image, formed by any quadruplet of points from the point correspondences, can be uniquely measured. According to classical theory, these cross-ratios have to be identical in both images.¹⁰ Moreover, it follows from Steiner's theorem¹¹ that the cross-ratios uniquely determine the epipole in the second image unless the epipole and the points are all conconic in the second image. Again, the epipolar line homography is uniquely determined since the points would otherwise form a line pair in the second image. \square

Note that for instability to occur, the points have to be close to a conic in the image. This is much more uncommon than for the points to be close to a critical surface in space, which can occur even with point correspondences that are dense in the images. If a conic can be fit to all the image points, we are in a situation where

the correspondence extraction has almost completely broken down, perhaps due to motion blur, textureless scene structure or other adverse conditions. In such cases, the likelihood will be very flat over all fundamental matrices and hence both hard and meaningless to represent in full. Moreover, the translation direction will also be highly uncertain. Hence, if a representation indexed by translation direction is used, the uncertainty will manifest itself in a flat distribution over translation direction. The weakness of the data will therefore be recognized even with the limited representation.

Thus, in summary, it is natural and powerful 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.¹²⁻¹⁵ It is also closely related to the high popularity of the plane-plus-parallax approach,¹⁶⁻²⁰ where one relies on the existence of a dominant homography and solves for that in order to guide the search for the translation direction.

Quite often, an even stronger assumption than ours has been used in the literature, namely that the parameters of the fundamental matrix apart from the epipole can be uniquely determined without specifying the epipole. Note that this means assuming that the rotation is separable from the translation, which is not always the case in practical situations. For example, with bas-relief ambiguity, it is not possible to determine rotation reliably without assuming that the translation is known. Also, in cases with strong parallax, there is not necessarily a dominant homography, or at the very least the homography may need to be refined in a manner that depends on the translation direction.

4. DATA-DRIVEN SAMPLING

As argued above, we can not search the likelihood over the whole parameter space. One approach to searching the likelihood in a selective manner is through some Markov Chain Monte Carlo Method (MCMC). A Markov chain is then used to stochastically traverse the parameter space. Transitions from the current state are randomly proposed. Typically, the proposal distribution falls off gradually with distance according to some topology around the current state. The chain prefers transitions to higher likelihood states, but does not necessarily get stuck in local maxima, since transitions to lower likelihoods are also often accepted. The transition probabilities are constructed so that at steady state the probability of the state equals the likelihood. Intuitively, the Markov chain is constructed to explore the significant regions of parameter space. However, it accomplishes this by probing the likelihood locally in all the dimensions of the parameter space, which can be prohibitively slow. Moreover, efficiency is dependent on the significant regions being relatively well connected. If there are disconnected modes in the posterior with very low likelihoods in between, the expected time before the chain transitions between modes is very long.

The generality of the approach is part of the problem. Using a raw local proposal density does not fully exploit the structure of the problem at hand. Several authors have noted that it can be much more efficient to search the parameter space with data-driven hypothesis generators.^{21,22} In particular for geometric problems, there is a very strong structure that can be exploited to efficiently construct a sample set that includes samples close to the true solution. The geometric structure is in fact so strong that it is not necessary to depend on a current state. Instead, a global indication of the significant regions of parameter space can potentially be achieved. This can be thought of as using data-driven heuristics to derive a density for importance sampling.⁹

We will use hypothesis generation in a similar manner as in RANSAC,²³ 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.

The partially data-driven sampling is closely related to the work of.²⁴ There, partial samples are taken. Each partial sample leaves a number of degrees of freedom manifested as a set of possible parameter values. Votes are cast for all the possible parameter values in a manner similar to a Hough-transform. What we do here is slightly different. Independent partial samples are used for each translation direction. We argue that this is more powerful. The reason is that if all the samples for the correct translation direction contain outliers, there is some hope that some other sample with a similar translation direction is correct and produces a satisfactory solution. Independent samples for each translation direction minimizes the risk that all relevant samples are contaminated. It therefore achieves a higher redundancy with the same number of likelihood evaluations.

Note that in the above taxonomy of fully and partially data-driven sampling, using a least-squares method over all the correspondences can be thought of as taking a single over-constrained sample.

5. 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

$$p(F|X) \propto p(X|F)p(F). \quad (13)$$

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

$$F_{opt}(e) = \underset{F \in \Omega(e)}{arg \ max} p(X|F) \quad (14)$$

and

$$f(e) = \int_{F \in \Omega(e)} p(X|F)dF. \quad (15)$$

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

$$p(e|X) \propto \int_{F \in \Omega(e)} p(X|F)p(F)dF \approx p(F_{opt}(e))f(e). \quad (16)$$

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

$$\int_e \int_{F \in \Omega(e)} g(F)p(F|X)dFde \quad (17)$$

can be approximated as

$$\frac{\int_e g(F_{opt}(e))p(F_{opt}(e))f(e)de}{\int_e p(F_{opt}(e))f(e)de}. \quad (18)$$

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.

5.1. 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.

5.2. Posterior Likelihood

We use a Sampson approximation (see²⁵):

$$s(x, x', F) = \frac{(x'^T F x)^2}{(F x)_1^2 + (F x)_2^2 + (x'^T F)_1^2 + (x'^T F)_2^2} \quad (19)$$

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 ($x'^T F x = 0$). This approximation has been found superior to symmetric epipolar distance and other approximations of similar computational complexity.²⁶

We model our data likelihood as

$$p(X|F) \propto \left(\prod_{i=1}^N \sigma^2 (\sigma^2 + s(x_i, x'_i, F))^{-1} \right)^{N-k}, \quad (20)$$

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 7.4. 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.

6. 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).²⁷ In the uncalibrated case, we use the 7-point (7pt) method and the 8-point (8pt) method.²⁵

The 3-point+epipole (3pt+e) and 5-point+epipole (5pt+e) methods are partially data-driven generators. The former was presented in.¹⁰ 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.²⁸ 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.

7. EXPERIMENTS

7.1. 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,²⁵ 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 random samples per cell, 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.

7.2. 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

$$error = \int_e (p(e) - \hat{p}(e)) de, \quad (21)$$

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 1, was created with 1000 samples per cell, or about 7×10^7 samples per method.

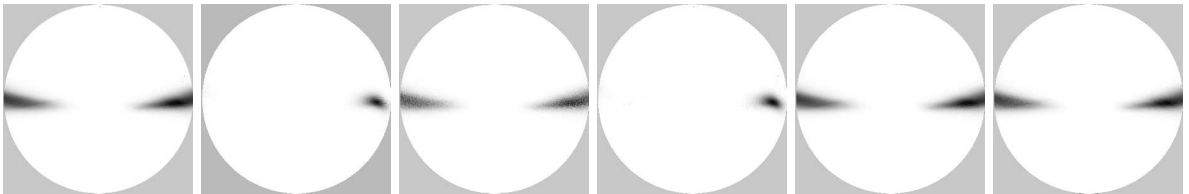


Figure 1. Posterior likelihood images of a scene. From left to right: true likelihood; 3pt+e method; 5pt+e method; 5pt method; 7pt method; 8pt method. Note that using the calibrated methods results in a much more constrained probability mass.

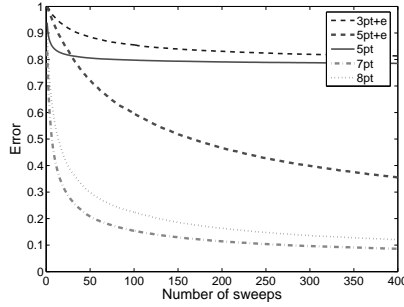


Figure 2. Comparison of convergence rates for the various hypothesis generation methods. Hypothesis generation times are not taken into account.

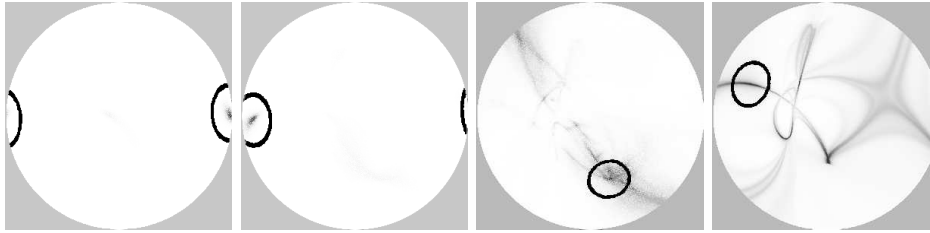


Figure 3. Examples of confidence intervals in an image sequence with a leftward translation. From left to right and top to bottom, the respective probability masses within each circled confidence interval are: 0.865, 0.567, 0.204, 0.065.

7.2.1. 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 2, 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.

7.3. 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 maximum. 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 3 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.

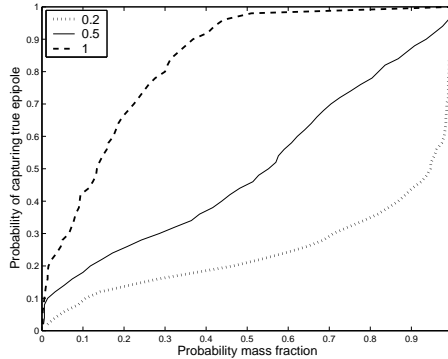


Figure 4. Cumulative distribution functions of confidence levels for varying values of k . Note that $k = 0.5$ most closely matches a uniform random variable.

7.4. 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 uniformly distributed random variable. To explore the sensitivity of our confidence intervals to discrepancies 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 samples per cell for 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 (20) to be approximately $1/2$. As seen in Figure 4, this achieves a balance in the confidence estimates, while $k = 1$ leads to underconfidence and $k = 0$ to overconfidence, with a highly peaked likelihood.

7.5. Application for Vehicle-Mounted Cameras

We now motivate the use of global uncertainty via a structure from motion problem. Specifically, we want to estimate the motion of a vehicle using a rigidly attached camera, which we perform as in.²⁹ In some cases, such as largely planar scenes, this task becomes difficult, as multiple hypotheses may fit the epipolar geometry. If an incorrect hypothesis is chosen, the estimated motion can break and lead to an incorrect state from which the motion estimate is unlikely to recover. The posterior likelihood and associated uncertainty provide a powerful tool for both understanding the behavior of the system and avoiding incorrect states. Figure 5 shows a correct reconstruction of the motion from a sideways-mounted camera. The image on the left shows one of the source images with feature tracks. As expected, most of the probability mass is contained within the confidence interval. Figure 6 shows an example of what can go wrong. The motion should be sideways, as in the previous case, but instead the estimate is of an unstable backwards motion. The likelihood image immediately gives a hint as to what happened: although there are hypotheses in the correct direction, these are widely distributed, and there are larger local peaks in the forward/backward directions. However, the uncertainty reflects the weakness, possibly allowing us to avoid a bad motion either by skipping the frame or relying more heavily on other motion estimators, e.g. from global positioning or inertial data, if available.

8. 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 the efficiency of various fully and partially data-driven hypothesis generators in deriving the

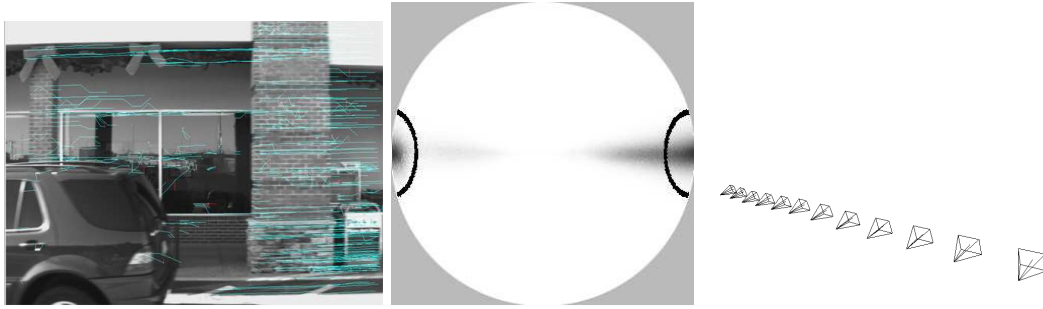


Figure 5. Example of a case with a correct motion estimate. The confidence interval encapsulates 55% of the probability mass.

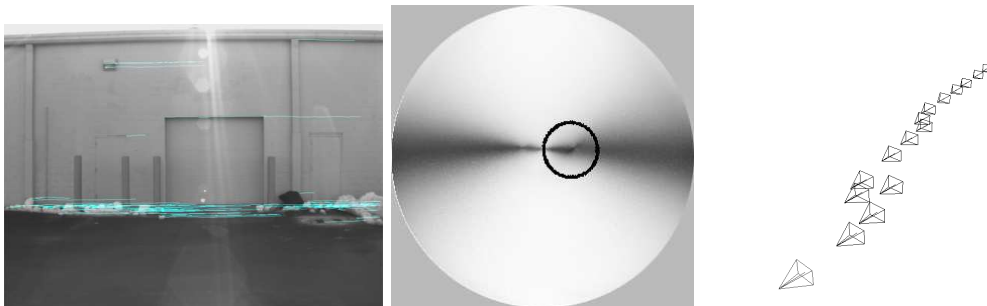


Figure 6. Example of a planar case with an incorrect motion estimate. Although the true motion is approximately the same as in Figure 5, there is a local peak in the probability mass, leading to a roughly backwards motion estimate. However, the weakness of the estimate is correctly reflected by the uncertainty, as the probability mass lying within the confidence interval is only 0.06.

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 tasks that are otherwise difficult, such as determining the validity of a motion estimate.

REFERENCES

1. B. Triggs, P. McLauchlan, R. Hartley, and A. Fitzgibbon, "Bundle adjustment – A modern synthesis," in *Vision Algorithms: Theory and Practice*, W. Triggs, A. Zisserman, and R. Szeliski, eds., LNCS, pp. 298–375, Springer Verlag, 2000.
2. J. J. Clark and A. L. Yuille, *Data Fusion for Sensory Information Processing Systems*, Kluwer Academic Publishers, Norwell, MA, USA, 1990.
3. P. Stefanovic, "Relative orientation - a new approach," *ITC Journal* **3**, pp. 417–448, 1973.
4. O. Faugeras, *Three-dimensional computer vision: a geometric viewpoint*, MIT Press, Cambridge, MA., 1993.
5. S. J. Maybank, "Classification based on the cross ratio," in *Applications of Invariance in Computer Vision*, pp. 453–472, 1993.
6. J. Philip, "A non-iterative algorithm for determining all essential matrices corresponding to five point pairs," *The Photogrammetric Record* **15**(88), pp. 589–599, 1996.
7. A. Doucet, N. de Freitas, and e. N. Gordon, *Sequential Monte Carlo Methods In Practice*, Springer-Verlag, New York, 2001.
8. J. J. K. O Ruanaidh and W. J. Fitzgerald, *Numerical Bayesian methods applied to signal processing*, Statistics and Computing Series, Springer, New York, 1996.

9. P. H. S. Torr and C. Davidson, "Impsac: Synthesis of importance sampling and random sample consensus," *IEEE Trans. Pattern Analysis and Machine Intelligence* **25**, pp. 354–364, Mar. 2003.
10. D. Nister and F. Schaffalitzky, "What do four points in two calibrated images tell us about the epipoles?," in *European Conference on Computer Vision*, pp. Vol II: 41–57, 2004.
11. J. Semple and G. Kneebone, *Algebraic projective geometry*, Oxford University Press, 1952.
12. B. K. P. Horn, "Relative orientation," *International Journal of Computer Vision* **4**, pp. 59–78, Jan. 1990.
13. A. Chiuso, R. Brockett, and S. Soatto, "Optimal structure from motion: Local ambiguities and global estimates," *International Journal of Computer Vision* **39**(3), pp. 195–228, 2000.
14. J. Oliensis, "The least-squares error for structure from infinitesimal motion," *International Journal of Computer Vision* **61**(3), pp. 259–299, 2005.
15. P. Baker, R. Pless, C. Fermuller, and Y. Aloimonos, "Eyes from eyes," in *3D Structure from Multiple Images of Large-Scale Environments*, p. 204 ff., 2000.
16. M. Irani, B. Rousso, and S. Peleg, "Recovery of ego-motion using region alignment," *IEEE Trans. Pattern Anal. Mach. Intell.* **19**(3), pp. 268–272, 1997.
17. R. Kumar, P. Anandan, and K. Hanna, "Shape recovery from multiple views: a parallax based approach," in *ARPA Image Understanding Workshop, Monterey, CA, Nov. 1994*, Morgan Kaufmann Publishers, (2929 Campus Drive, Suite 260, San Mateo, California 94403o), November 1994.
18. R. Szeliski and P. H. S. Torr, "Geometrically constrained structure from motion: Points on planes," in *3D Structure from Multiple Images of Large-Scale Environments*, pp. 171–186, 1998.
19. B. Triggs, "Plane + parallax, tensors, and factorization," in *European Conference on Computer Vision*, pp. I: 522–538, 2000.
20. R. Kaucic, R. Hartley, and N. Dano, "Plane-based projective reconstruction," 2001.
21. P. Chang, "Robust tracking and structure from motion with sampling method," 2002.
22. Z. Tu and S. C. Zhu, "Image segmentation by data-driven markov chain monte carlo," in *International Conference on Computer Vision*, pp. II: 131–138, 2001.
23. M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM* **24**(6), pp. 381–395, 1981.
24. C. F. Olson, "A general method for geometric feature matching and model extraction," *International Journal of Computer Vision* **45**(1), pp. 39–54, 2001.
25. R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, ISBN: 0521540518, second ed., 2004.
26. Z. Zhang, "Determining the epipolar geometry and its uncertainty: A review," Tech. Rep. 2927, Sophia-Antipolis Cedex, France, 1998.
27. D. Nister, "An efficient solution to the five-point relative pose problem," *IEEE Trans. Pattern Analysis and Machine Intelligence* **26**, pp. 756–777, June 2004.
28. T. Werner, "Constraint on five points in two images," in *IEEE Computer Vision and Pattern Recognition or CVPR*, pp. II: 203–208, 2003.
29. D. Nister, O. Naroditsky, and J. Bergen, "Visual odometry," in *IEEE Computer Vision and Pattern Recognition or CVPR*, pp. I: 652–659, 2004.