Photogrammetrie • Fernerkundung • Geoinformation 1/2007, S. 25–33, 6 figs.

A Probabilistic Notion of Camera Geometry: Calibrated vs. Uncalibrated

JUSTIN DOMKE & YIANNIS ALOIMONOS, College Park, Maryland, USA

Keywords: Computer vision, photogrammetry, Correspondence, matching, fundamental matrix, essential matrix, egomotion, multiple view geometry, probabilistic, stochastic

Zusammenfassung: *Berechnung der Fundamental- und der essentiellen Matrix mittels Korrespondenzen zwischen Wahrscheinlichkeitsverteilung statt Punktkorrespondenzen*. Traditionell werden zur Bestimmung der Fundamental- bzw. essentiellen Matrix Punktkorrespondenzen in Bildpaaren gefunden und auf deren Basis die Kamerageometrie bestimmt. Die Schätzung der Geometrie ist gut verstanden, versagt in der Praxis jedoch häufig.

Diese Arbeit verfolgt daher eine andere Strategie. Zu Beginn wird die Wahrscheinlichkeitsverteilung von Punktkorrespondenzen geschätzt, aus der anschließend die Kamerageometrie bestimmt wird. Dadurch ist der Schritt der Korrespondenzfindung wesentlich vereinfacht, was allerdings zu Lasten des Schätzprozess der Kamerageometrie geht. Ein auf dieser Basis entwickelter Algorithmus bestätigt jedoch dieses Vorgehen in umfangreichen Untersuchungen.

Abstract: We suggest altering the fundamental strategy in Fundamental or Essential Matrix estimation. The traditional approach first estimates correspondences, and then estimates the camera geometry on the basis of those correspondences. Though the second half of this approach is very well developed, such algorithms often fail in practice at the correspondence step.

Here, we suggest altering the strategy. First, estimate probability distributions of correspondence, and then estimate camera geometry directly from these distributions. This strategy has the effect of making the correspondence step far easier, and the camera geometry step somewhat harder. The success of our approach hinges on if this trade-off is wise. We will present an algorithm based on this strategy. Fairly extensive experiments suggest that this trade-off might be profitable.

1 Introduction

The problem of estimating camera geometry from images lies at the heart of both Photogrammetry and Computer Vision. In our view, the enduring difficulty of creating fully automatic methods for this problem is due to the necessity to integrate image processing with multiple view geometry. One is given images as input, but geometry is based on the language of points, lines, etc. Bridging this gap – using image processing techniques to create objects useful to multiple view geometry – remains difficult. In both the Photogrammetric and Computer Vision literature, the object at interface between image processing and geometry is generally correspondences, or matched points. This is natural in Photogrammetry, because correspondences are readily established by hand. However, algorithmically estimating correspondences directly from images remains a stubbornly difficult problem.

One may think of most of the previous work on Essential or Fundamental matrix estimation as falling into one of two categories. First, there is a rather mature literature on Multiple View Geometry. This is well summarized in HARTLEY & ZISSERMAN's recent book (HARTLEY & ZISSERMAN 2004), emphasizing the uncalibrated techniques leading to Fundamental Matrix estimation. Specifically, there are techniques for estimating the Fundamental Matrix from the

1432-8364/07/2007/0025 $ 2.25
© 2007 DGPF / E. Schweizerbart'sche Verlagsbuchhandlung, D-70176 Stuttgart

minimum of seven correspondences (Bartoli & Sturm 2004). In the calibrated case, the Essential Matrix can be efficiently estimated from five correspondences (Nistér 2004). Given perfect matches, it is fair to say that the problem is nearly solved.

The second category of work concerns the estimation of the correspondences themselves. Here commonly a feature detector (e. g. the Harris corner detector, Harris & Stephens 1988) is first used to try to find points whose correspondence is most easily established. Next, matching techniques are used to find probable matches between the feature points in both images (e. g. normalized cross correlation, or SIFT features, (Lowe 2004). These are active research areas, and progress continues up to the present.

Nevertheless, no fully satisfactory algorithm exists. Current algorithms often suffer from problems such as change in scale or surface orientation (Schmid et al. 2000). Furthermore, there are many situations in which it is essentially *impossible* to estimate correspondences without using a higher-level understanding of the scene. These include repeated structures in the image, the aperture effect, lack of texture, etc. When humans estimate correspondences, they use this high-level information. Nevertheless, it is unavailable to algorithms.

Research in multiple view geometry, of course, has considered the difficulties in the underlying algorithms for correspondence estimation. As such, robust techniques such as RANSAC (Fischler & Bolles 1981) are traditionally used to estimate a camera geometry from a set of correspondences known to include many incorrect matches. These techniques are fairly successful, but because even 'inlying' correct matches include noise there is a difficulty in discriminating between inlying matches with noise, and outlying, 'wrong', matches. When simultaneously adjusting the camera geometry, and 3-D points in a final optimization, bundle adjustment methods frequently use more sophisticated noise models which smoothly account for error due to both noise, and 'outlying' matches (Triggs et al. 1999).

In this paper, we suggest that it is worth stepping back and reconsidering if correspondences are the correct structure to use at the interface between image processing and multiple view geometry. Point correspondences are natural in Photogrammetry because they are easily estimated by humans. Nevertheless they are very difficult to estimate algorithmically. Here, we suggest instead using *correspondence probability distribution*s. We can see immediately that this makes the image processing side of the problem much easier. If repetitive structure or the aperture effect presents itself, it is simply incorporated into the probability distribution. We will present a simple, contrast invariant, technique for estimating these correspondences from the phase of tuned Gabor filters.

The more difficult side of this strategy concerns multiple view geometry. One must estimate the camera geometry from only distributions of correspondence. As we will see, one can quite easily define a *probability* for any given camera motion, from only these distributions of correspondence. We then present a heuristic non-linear optimization scheme to find the most probable geometry. In practice, this space has a similar structure to the least-squares epipolar error space (Oliensis 2005), in that it contains relatively few local minima.

1.1 *Previous Work*

Other work has asked similar questions. First, there are techniques which generate from images feature points, and local image profiles, without estimating an explicit correspondences (Makadia et al. 2005). These techniques then find motions which are compatible with these features, in the sense that each feature tends to have a compatible feature along the epipolar line in the second image.

Other work has created weaker notions of correspondences, such as the normal flow. If a point is along a textureless edge in one image, local measurements can only constrain it to lie along the same edge in the second image. This constraint is essen-

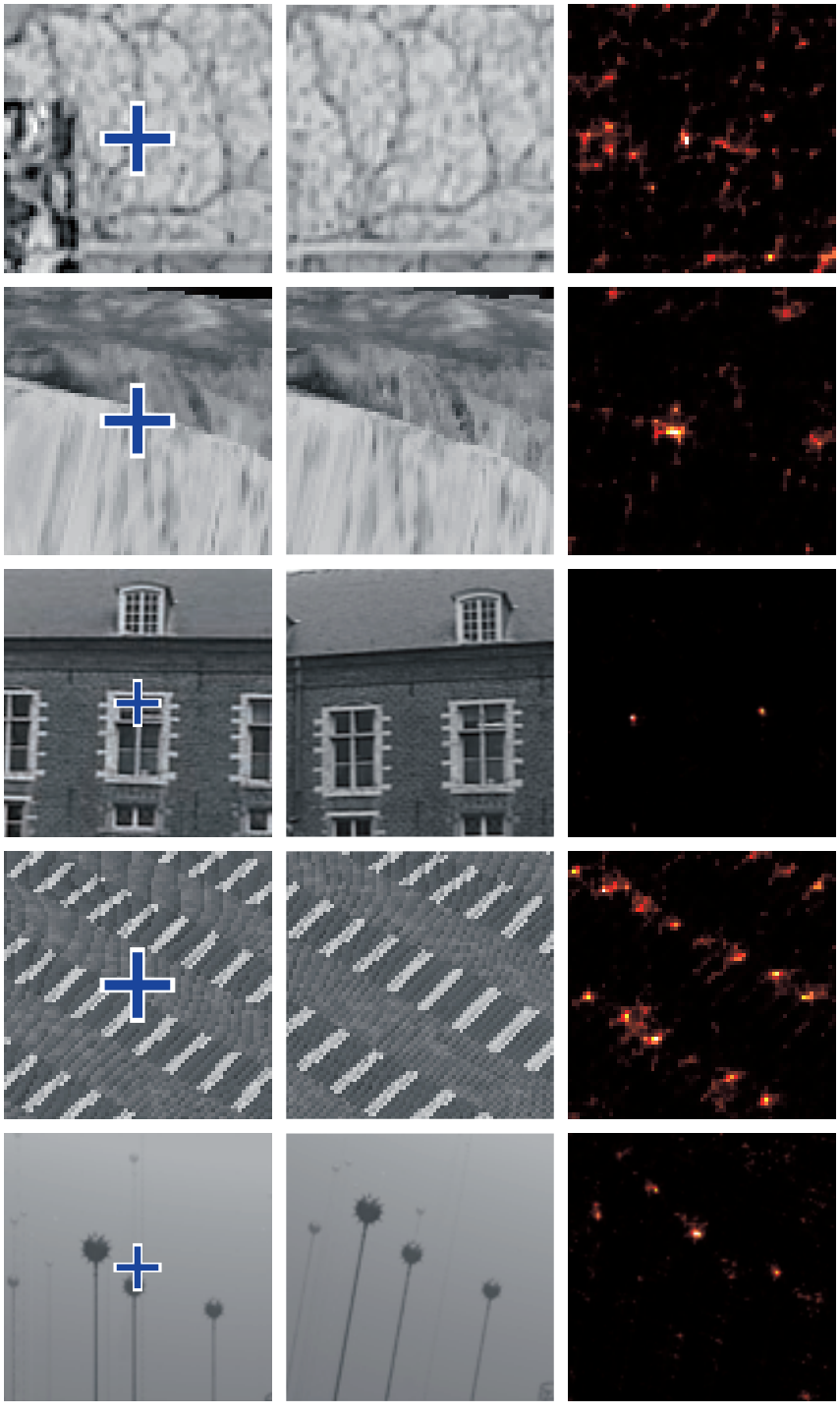


Fig. 1: Correspondence Probability Distributions. Left: First image, with point in consideration marked. Center: Second image: Right: Probability distribution over the points in the second image, with probability encoded as color.

tially the normal flow, and algorithms exist to estimate 3-D motion directly from it (Brodsky et al. 2000). Though these techniques will not suffer from the aperture effect, they cannot cope with situations such as repeated structures in the images. It is also important to notice that the normal flow will give up information unnecessarily at points which do not happen to suffer from the aperture effect.

2 Correspondence Probability Distributions

Given a point s in the first image, we would like the probability that this corresponds most closely to each pixel $\hat{q}$ in the second image. It is important to note that there is no obvious way to use traditional matching techniques here. Whereas traditional techniques try to find the most probably point corresponding to s, we require the relative probabilities of *all* points.

Our approach is based on the phase of tuned Gabor filters. Let $\phi_{l,\gamma}(s)$ denote the phase of the filter with scale l and orientation γ at point s. Now, given a single filter, (l, γ), we take the probability that s corresponds to a given point $\hat{q}$ to be proportional to

$$\exp\left(-[\phi_{l,\gamma}(s) - \phi_{l,\gamma}(\hat{q})]_\pi^2\right) + 1 \qquad (1)$$

The notation $[\phi]_\pi$ here indicates taking the principal angle of ϕ, in the range from $-\pi$ to π. This is necessary to deal with phase wrapping. Combining the probability distributions given by all filters then yields the probability that s corresponds to $\hat{q}$, which we denote by $\rho_s(\hat{q})$.

$$\rho_s(\hat{q}) \propto \prod_{l,\omega}\left(\exp\left(-[\phi_{l,\gamma}(s) - \phi_{l,\gamma}(\hat{q})]_\pi^2\right) + 1\right) \qquad (2)$$

Note, here that $\hat{q}$ corresponds to a particular *pixel* in the second image. Since we are computing probabilities over a discrete grid, we approximate the probability that s corresponds to an arbitrary point, having non-integer coordinates, though the use of a Gaussian function.

$$\rho_s(q) \propto \alpha + \max_{\hat{q}} \rho_s(\hat{q}) \exp(-|q - \hat{q}|^2) \qquad (3)$$

Here, α represents the probability that the information given by the Gabor filters is misleading. This would be the case, for example, were the point s to become occluded in the second image. Notice that adding the constant of α is equivalent to combining the distribution with the 'flat' distribution in which all points q are equally likely. In all experiments described in this paper, we have used $\alpha = 1$.

Correspondence distributions for several images are shown in Fig. 2.

3 Essential and Fundamental Matrix Estimation

Given the correspondence distributions, we will define natural distributions over the space of the Fundamental and Essential Matrices. Because the space of these matrices are of high dimension (7 and 5 respectively), it is impractical to attempt to calculate a full distribution, by sampling. It is possible that future work will directly use these distributions. Nevertheless, we use a simple heuristic optimization to maximize the probability in the Essential or Fundamental Matrix space. This makes it possible to examine the behavior of these distributions more easily.

3.1 *Fundamental and Essential Matrix Probability*

Given the correspondence distribution for a single point s, $\rho_s(\cdot)$, we define a distribution over the space of fundamental matrices.

$$\rho(F) \propto \max_{q:q^T F s = 0} \rho_s(q) \qquad (4)$$

Thus, the probability of a given Fundamental Matrix F is proportional to *the maximum probability correspondence compatible with the epipolar constraint*. Now, to use all correspondence distributions, simply take the product of the distributions given by each point s.

$$\rho(F) \propto \prod_s \max_{q:q^T F s = 0} \rho_s(q) \qquad (5)$$

Substituting our expression for $\rho_s(q)$ from Equation (3), we obtain

$$\rho(F) \propto \prod_s \left[\alpha + \max_{q:q^T F s=0} \max_{\hat{q}} \rho_s(\hat{q}) \exp(-|q-\hat{q}|^2)\right] \tag{6}$$

Rearranging terms, this is

$$\rho(F) \propto \prod_s \left[\alpha + \max_{\hat{q}} \rho_s(\hat{q}) \max_{q:q^T F s=0} \exp(-|q-\hat{q}|^2)\right] \tag{7}$$

Notice here, that we do not need to explicitly find the point q. Only required is $\max_{q:q^T Fs}|q-\hat{q}|$. Notice that this is exactly the minimum distance of the point $\hat{q}$ from the line Fs. Therefore, we can write the probability of F in its final form.

$$\rho(F) \propto \prod_s \left[\max_{\hat{q}} \rho_s(\hat{q}) \exp(-(\hat{q}^T l_{(F,s)})^2) + \alpha\right] \tag{8}$$

Here, $l_{(F,s)}$ is the line Fs normalized such that $r^T l_{(F,s)}$ gives the minimum distance between r and the line Fs on the plane $z = 1$. If F_i is the ith row of F, then

$$l_{(F,s)} = \frac{Fs}{\sqrt{(F_1 s)^2 + (F_2 s)^2}} \tag{9}$$

When searching for the most probable F, a parameterization of the fundamental matrices is required. We found it convenient to use three parameters f, p_x, and p_y representing the focal length, and x and y coordinates of the principal point. Next, keeping the magnitude of the translation vector t fixed to one, we took two parameters to parameterize its axis and angle. Finally, we used 3 parameters to represent the rotation vector ω. This corresponds to a rotation of an angle $|\omega|$ about the axis $\omega/|\omega|$.

$$K = \begin{bmatrix} f & 0 & p_x \\ 0 & f & p_y \\ 0 & 0 & 1 \end{bmatrix} \tag{10}$$

$$E = [t] \times R(\omega) \tag{11}$$

$$F = K^{-T} E K^{-1} \tag{12}$$

Notice there are a total of 8 free parameters, despite the fact that the Fundamental Matrix has only 7 degrees of freedom. Though this presents no problem to the estimation of F, it does mean that an ambiguity is present in the underlying parameters.

To extend this to the calibrated case, we take K to be known. Thus, there are now 5 free parameters: 2 for the translation t, and 3 for the rotation ω. It would be trivial to extend this to the case that only certain calibration parameters were known, or to include a constant for camera skew.

3.2 *Optimization*

To explore the behavior of the probability distributions over the Fundamental and Essential Matrices, we will use a heuristic optimization to try to find arg $\max_F \rho(F)$ and arg $\max_E \rho(E)$, respectively. The optimization proceeds as follows: First, select N random points in the Fundamental or Essential matrix space. Evaluate $\rho(E)$ or $\rho(F)$ at each of these points. Next, take the M highest scoring points, and run a nonlinear optimization, initialized to each of these points. We have used both Simplex and Newton's type optimizations, with little change in performance. The final, highest scoring point is taken as the max.

For the calibrated case, we have found that using $N = 2500$ and $M = 25$ was sufficient to obtain a value very near the global maximum in almost all cases. As in the case for the standard least-squares error surface (Oliensis 2005, Tian et al. 1996), there are generally several, but only several local minima. Usually, a significant number of the nonlinear searches lead to the same (global) point.

In the uncalibrated case, we used $N = M = 100$. (Thus searches are taken from 100 random points.) We found that it was neces-sary to increase M to 100 to obtain reasonable certainty of obtaining the global maximum. At the same time, we found that increasing N did not improve results, and may even be counterproductive. Still, the

space of $\rho(F)$ appears to have more local minima, and even this increased method does not always appear to achieve the global maximum.

4 Experiments

To analyze the performance of the framework, we prepared three different 3-D Models with the POV-Ray software. Each model was chosen for its difficulty, including repetitive structure, lack of texture, or little image motion. The use of synthetic models makes the exact motion and calibration parameters available. For each model, we generated two different image sequences, one with a forward motion, and one with a motion parallel to the image plane.

For each image pair, 10,000 correspondence probability distributions were created. Next, the calibrated and uncalibrated algorithm were both run across a range of input sizes. For each input size, 100 random subsets of the correspondences were generated, and the algorithm was run on each input.

In the calibrated case, the measurement of error is simple. Let the true translation vector be t_0, normalized so that $|t_0| = 1$. Let the vector parameterizing the true rotation matrix be ω_0. The error metrics we use are simply the Euclidean distance between the estimated and true motion vectors, $|t - t_0|$, and $|\omega - \omega_0|$ respectively. For each input size, means are taken over the errors for all resulting motion estimates.

In the uncalibrated case, we must measure the error of a given fundamental matrix F. Commonly used metrics such as the Frobenius norm are difficult to interpret, and allow no comparison to the calibrated case. Instead, we use the known ground truth calibration matrix K to obtain E (Hartley & Zisserman 2004).

$$E = K^T F K \tag{13}$$

Next, Singular Value Decomposition is used to decompose E into the translation and rotational components, $E = [t] \times R(\omega)$. From this, it is simple to recover the underlying motion parameters, t and ω. The error is then measured in the same way as the calibrated case.

Results for the 'Cloud', 'Abyss', and 'Biscuit' models are shown in Figs. 2, 3, and 4 respectively. Several observations are clear from the data. First, motion estimation is always more accurate when the epipole is in the middle of the image than when it is parallel to it. Surprisingly, perhaps, neither the calibrated nor uncalibrated approach clearly outperforms the other. The performance of the uncalibrated approach relative to the calibrated approach is better when the epipole is further from the image.

Two frames from the 'Castle' sequence, along with the epipolar lines are shown in Fig. 5. Two frames from the popular 'Oxford Corridor' sequence are shown in Fig. 6. In both cases, approximately 2000 correspondence distributions were used. Though no ground truth calibration or motion is available, the reader can observe the close correspondence among epipolar lines.

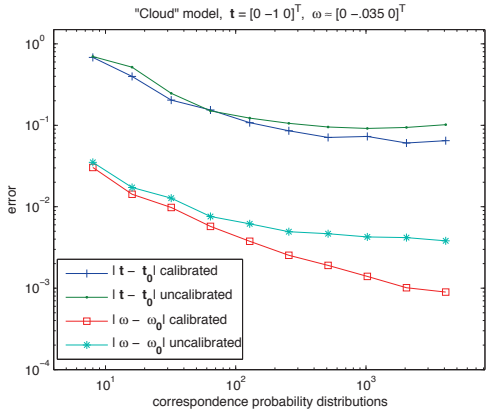
The running time of the algorithm is dominated by the time to generate correspondence distributions. In practice, the motion estimation step runs on the order of a minute on a modern laptop.

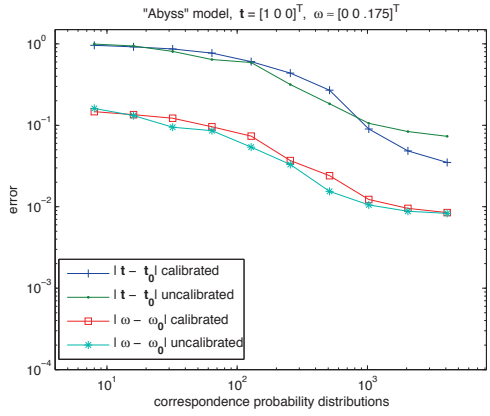
5 Conclusions and Future Work

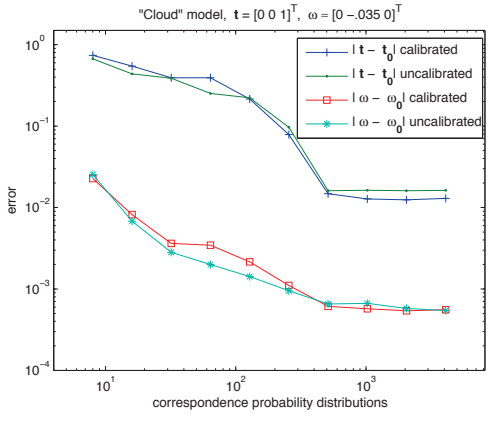
With real cameras, neither the fully calibrated, nor fully uncalibrated approach is fully realistic. In practice, one has some idea of the calibration parameters, even if only from knowledge of typical cameras. At the same time, even when a camera is calibrated, the true calibration is not found *exactly*. It would be quite natural to extend this paper's work to create a unifying approach between the two cases.

Write the prior distribution over the focal lengths by $\rho(f)$. Similarly, we can write the prior distributions of the principal point by $\rho(p_x, p_y)$. Now, we can make the Bayesian nature of this approach more explicit by writing Eqn. 14 as

$$\rho(E|f, p_x, p_y) \propto \max_{q: q^T K^{-T} E K^{-1} s = 0} \rho_s(q) \tag{14}$$







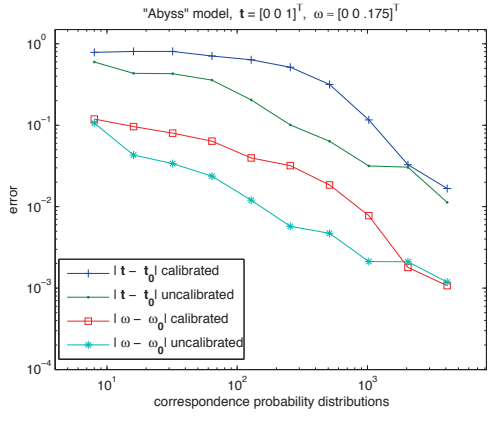
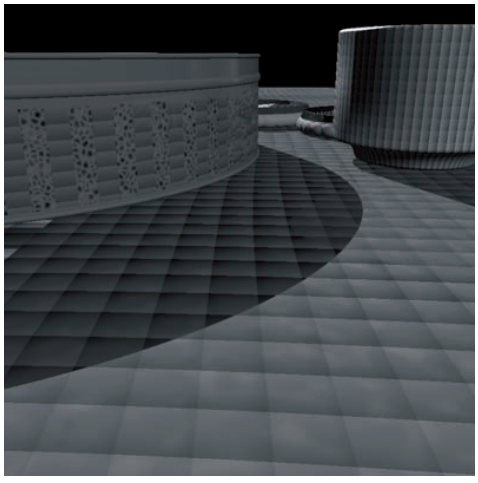
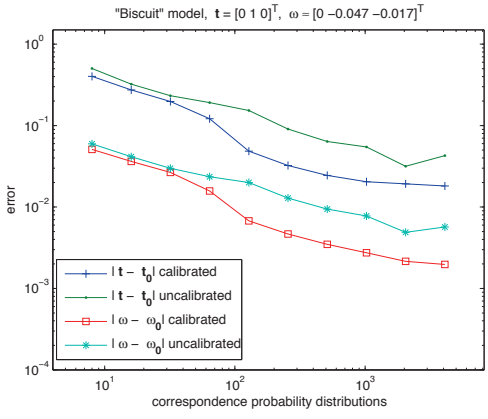


Fig. 2: 'Cloud' model, and mean errors for the two different motions.

Fig. 3: 'Abyss' model, and mean errors for two different motions.





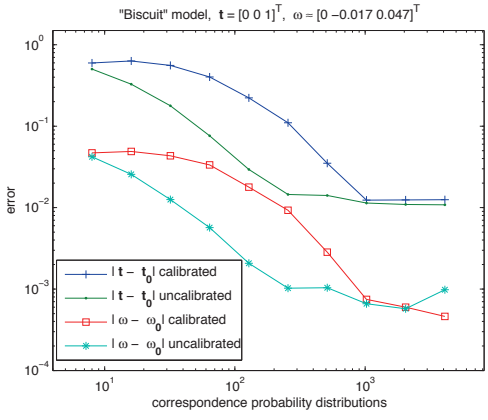


Fig. 4: 'Biscuit' model, and mean errors for two different motions.

Fig. 5: Two frames from the 'Castle' sequence, with epipolar lines overlaid.

Now, in the optimization step, instead of seeking

$$\arg\max_{F} \rho(F), \tag{15}$$

the optimization would be over

$$\arg\max_{E,f,p_x,p_y} \rho(E|f,p_x,p_y)\,\rho(f)\,\rho(p_x,p_y) \tag{16}$$

In this way, in one step, the most likely calibration parameters would be found as well as the most likely motion. This could be particularly useful in the common case that the camera calibration is approximately known, but the focal length changes, perhaps due to change of focus.

Acknowledgements

The authors would like to thank Gilles Tran for providing the 3-D models, The Oxford Visual Geometry Group for provid-

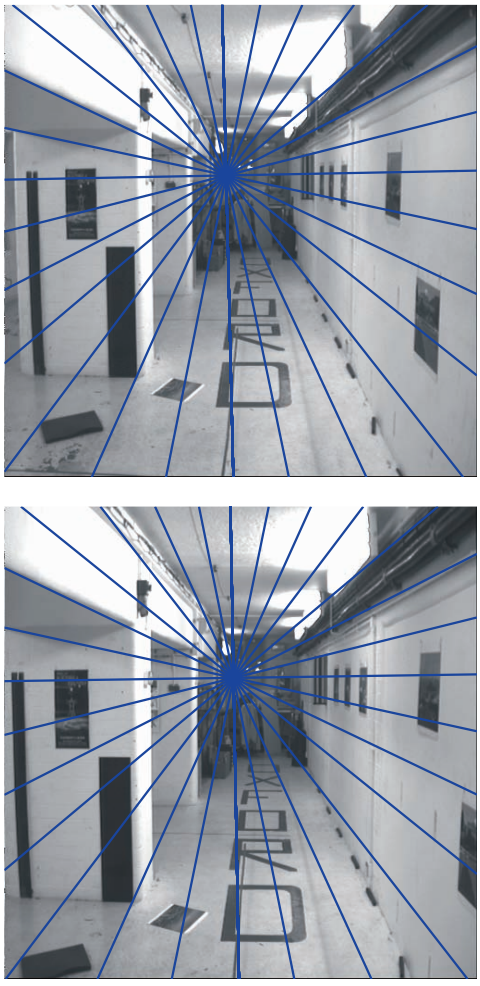


Fig. 6: Two frames from the 'Oxford Corridor' sequence, with epipolar lines overlaid.

ing the 'Oxford Corridor' sequence, and Marc Pollefeys for providing the 'Castle' sequence.

References

Bartoli, A. & Sturm, P., 2004: Non-linear estimation of the fundamental matrix with minimal parameters. – IEEE Transactions on Pattern Analysis and Machine Intelligence **26**(4): 426–432.

Brodsky, T., Fermuller, C. & Aloimonos, Y., 2000: Structure from motion: Beyond the epipolar constraint. – International Journal of Computer Vision **37**(3): 231–258.

Fischler, M.A. & Bolles, R.C., 1981: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. – Commun. ACM **24**(6): 381–395.

Harris, C.G. & Stephens, M., 1988: A combined corner and edge de-tector. – AVC **88:** 147–151.

Hartley, R.I. & Zisserman, A., 2004: Multiple View Geometry in Computer Vision. – 2nd ed., Cambridge University Press, ISBN: 052154 0518.

Lowe, D.G., 2004: Distinctive image features from scale-invariant keypoints. – Int. J. Comput. Vision **60**(2): 91–110.

Makadia, A., Geyer, C. & Daniilidis, K., 2005: Radon-based structure from motion without correspondences. – CVPR.

Nistér, D., 2004: An efficient solution to the five-point relative pose problem. – IEEE Trans. Pattern Anal. Mach. Intell. **26**(6): 756–777.

Oliensis, J., 2005. The least-squares error for structure from infinitesimal motion. – Int. J. Comput. Vision **61**(3): 259–299.

Pollefeys, M., Gool, L.V., Vergauwen, M., Verbiest, F., Cornelis, K., Tops, J. & Koch, R., 2004: Visual modeling with a hand-held camera. – Int. J. Comput. Vision **59**(3): 207–232.

Schmid, C., Mohr, R. & Bauckhage, C., 2000: Evaluation of interest point detectors. – International Journal of Computer Vision **37**(2): 151–172.

Tian, T., Tomasi, C. & Heeger, D., 1996: Comparison of approaches to egomotion computation.

Triggs, B., McLauchlan, P.F., Hartley, R.I. & Fitzgibbon, A.W., 1999: Bundle adjustment – a modern synthesis. – Workshop on Vision Algorithms, pp. 298–372.

Address of the authors:

M.Sc. Justin Domke
e-mail: domke@cs.umd.edu

Prof. M.Sc. Ph.D. Yiannis Aloimonos
e-mail: yiannis@cfar.umd.edu

Computational Vision Laboratory
Center for Automation Research
University of Maryland
College Park, MD, 20740, USA
http://www.cs.umd.edu/domke/
http://www.cfar.umd.edu/yiannis/

Manuskript eingereicht: November 2006
Angenommen: November 2006