DEALING WITH DEGENERACY IN ESSENTIAL MATRIX ESTIMATION

Peter Decker and Dietrich Paulus

Active Vision Group
Institute for Computational Visualistics
University of Koblenz-Landau
Universitätsstr. 1, 56070 Koblenz, Germany

Tobias Feldmann

Institute for Algorithms and Cognitive Systems
University of Karlsruhe (TH)
Kaiserstr. 12, 76131 Karlsruhe, Germany

ABSTRACT

Estimation of 3-D egomotion from video input, also known as visual odometry, is an important issue for many applications today. Augmented reality (AR) and robotic systems for example rely heavily on correct pose and motion estimation. In this paper we discuss egomotion estimation from a single camera. We focus on the estimation of the essential matrix and problems which arise from degenerate configurations when using the well known normalized 8-point algorithm. Lately, the BEEM algorithm [1] has been published, which is a combined approach of several RANSAC methods. It tries to guide essential matrix generation away from degenerate configurations. We argue, that there are still cases which are not covered by the BEEM approach and encourage the combination with an improved method for detecting degenerate configurations (IDD).

Index Terms— essential matrix, epipolar geometry, degenerate configuration, BEEM

1. INTRODUCTION

The task of obtaining 3-D egomotion data from a single camera which moves in a rigid environment is known as visual odometry [5]. This information is helpful when trying to solve e. g. the SLAM (simultaneous localization and mapping) set of problems, because it is less prone to some errors which only affect usual odometry (like slipping wheels of a robot).

A common way to retrieve the rotation $\boldsymbol{R}$ and translation $\boldsymbol{t}$ performed by a camera between two consecutive frames is to estimate the essential matrix $\boldsymbol{E} = [\boldsymbol{t}]_\times \boldsymbol{R}$, with $[\boldsymbol{t}]_\times$ denoting the cross product matrix. $\boldsymbol{E}$ describes a relation between all corresponding points $\boldsymbol{p}_i, \boldsymbol{q}_i$ from the two images by the epipolar constraint (1).

$$\begin{aligned} \boldsymbol{q}_i{}^{\mathrm{T}}\, \boldsymbol{E}\, \boldsymbol{p}_i &= 0 \\ \boldsymbol{q}_i{}^{\mathrm{T}}\, [\boldsymbol{t}]_\times\, \boldsymbol{R}\, \boldsymbol{p}_i &= 0 \end{aligned} \tag{1}$$

Both, rotation and translation of the camera between the two frames can be computed from $\boldsymbol{E}$ [3]. The normalized 8-point algorithm [4] is a simple and fast way to obtain a valid $\boldsymbol{E}$-matrix. Matching methods for feature points will usually not return perfect results, i. e. the returned point sets will be noisy or contain outliers. Therefore, RANSAC-like paradigms [2] [9] [10] are commonly used to filter the results. These methods use a randomly determined minimal set of points to compute an essential matrix and measure the quality of the obtainted solution. The quality is defined by the size of the so called support set, that is the set of all corresponding points satisfying equation (1) within a certain threshold.

The structure of the paper is as follows: In section 2 we discuss two different kinds of problems which arise from degenerate data. We motivate a combined approach to solve these problems and present an improved method for detecting degenerate configurations. Results are given in section 3. At last, we give a conclusion in section 4.

2. UNDERSTANDING DIFFERENT DEGENERATE CONFIGURATIONS

A degenerate configuration occurs, when multiple solutions for a problem exist. This may be the result of a data set which is insufficient to determine a unique solution or by wrong assumptions about the number of degrees of freedom (DOF) describing the model, i. e. the motion of the camera. Multiple solutions may be *mathematically correct*, but to obtain the true rotation and translation performed by the camera, one needs to find a solution which additionally *maps reality*.

The 8-point algorithm has some known issues with degenerate point sets when estimating the essential matrix. Wrong solutions will occur when the utilized point correspondences for estimation of the essential matrix lie in a small area of the image only, or arise from features on a single dominant plane [7]. These will not fully describe the geometry of the images.

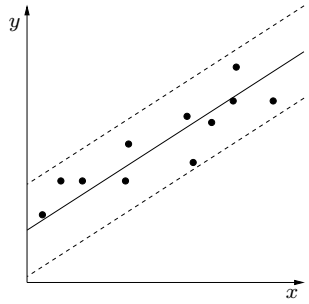
2.1. Avoiding degenerate solutions

Ilan Shimshoni introduced the BEEM alogrithm [1], which tries to avoid degenerate solutions caused by improper point correspondences while estimating the essential matrix. He therefore introduces an additional step of local exploration for his multi state approach. In this step he instantiates a new model for the essential matrix from a union of point correspondences chosen from the support set of the current model and from the rest of the correspondences. By comparing the size of both models' support sets, the one to continue with is chosen. This step enables the algorithm to escape from a degenerate configuration as described above: An essential matrix which is computed from inlier point correspondences in a small region of the image will possibly yield a good support set (thus, beeing classified as mathematically correct), but does not describe the epipolar geometry correctly. If another inlier correspondence from outside this area is taken into account additionally, the essential matrix will hopefully describe the geometry of the *whole* image and have an even larger support set. The same holds for features which arise from one plane in the scene only.

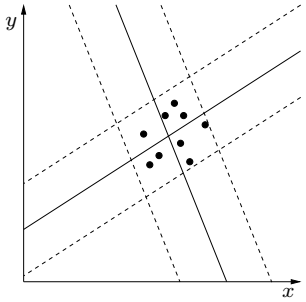
2.2. Detecting degenerate configurations

Besides the already mentioned degenerate solutions caused by improper point correspondences, there exists another form of degeneracy which can ruin epipolar estimation. As Torr, Zisserman and Maybank point out in [8], degeneracy needs to be modeled in order to detect outliers which would be missed otherwise. The scenario is as follows: If the movement of the camera has less degrees of freedom than the essential matrix (e. g. it is a pure translation), a correctly estimated essential matrix may increase the size of its support set by adding outlier *without* failing mathematical correctness. This is due to the fact, that pure translation only has three DOF, while the estimated essential matrix offers five DOF. Accordingly, any outlier could be added to the computation of $\boldsymbol{E}$ and the result would still appear to be correct.

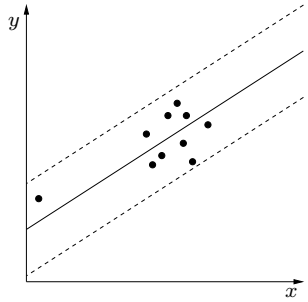
In [8] a concrete example is given, which helps to understand this problem: Assume a RANSAC like method is used to fit a line into a set of 2-D points. Figure 1(a) shows a case, where a unique solution for a line is found. The dashed lines denote the threshold for points to be classified as inlier for this model. In Figure 1(b) a degenerate set of points is shown, where several solutions for a line exist. This is due to the fact, that the 2-D datapoints describe a point rather than a line, which we are looking for. Since a point is a model of a lower dimension, multiple solutions for lines are possible. Figure 1(c) shows a case similar to (b), except for an additional single outlier. When fitting a line into this data set, there is a single solution (the displayed line) which covers all points, including the outlier. But this solution does not describe reality appropriately. A correct description of the data present would be, that it describes a point rather than a line. Using a point as a solution would also identify the single outlier in this case.



(a) Non-degenerate data, unique and correct solution



(b) Degenerate, lower dimension data, multiple solutions



(c) Degenerate data with an outlier, seemingly correct but wrong solution

Fig. 1. Example from [8] illustrating the problem of degeneracy caused by data of a lower dimension than expected.

2.3. Improving the model selection

The above described problem of fitting a line into a 2-D data set is similar to finding an essential matrix, which describes a rotation and translation in 3-D with 5 DOF. If the camera has only undergone a pure translation for example, one might find an essential matrix which is consistent with all correct point correspondences and some outliers, because a pure translation has less DOF than the model for $\boldsymbol{E}$. Since a correct model for the essential matrix in case of pure translation can never beat this solution with respect to the size of the support set, it makes sense to test for lower order models and compare the results. In [8], the lower order model with the largest support set is selected regardless of its DOF, if its support set has a size of at least 95% of the essential matrix' support set.

We argue, that by treating all lower order models equally the same problem can occur again in a lower dimension. Regarding the example given in Figure 1, imagine fitting a plane into 3-D data describing a point with 1 outlier. By selecting the lower order model with the largest support set (probably a line through the point and the outlier in this case) as suggested, one disregards the possibility of another model

with even less DOF being the correct representation for the given data (a point). In case of essential matrix estimation this means, that one might accept a pure translation as a degenerate model with respect to epipolar geometry in case of no motion, by adding an outlier and thus increasing the size of the support set but yielding a wrong solution. We suggest building a dependency graph of possible degenerate models and recursively testing them. A small example is shown in Figure 2: If either 'pure rotation' or 'pure translation' models can achieve a support set of at least 95% of the best epipolar geometry model found, the 'no motion' model is tested against them. To be chosen as the correct model, 'no motion' does not need to outperform 'pure rotation' or 'pure translation', it just needs to achieve at least 95% of their support sets' size, because it is a case of a degenerate model with respect to them.

By using this knowledge of dependencies, we rule out models of higher order than the real data present, which are mathematically consistant with most data and additional outliers, but do not map reality.

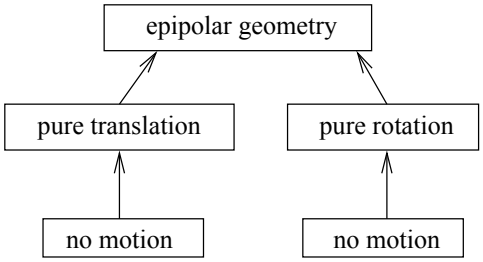


Fig. 2. Dependency graph of possible degenerate models for robot motion. Arrows denote a 'is a possible degenerate model with respect to' relation.

3. EXPERIMENTS AND EVALUATION

To prove our concept, we evaluated trajectory computation from epipolar estimation in case of a single camera moving in a rigid environment. As the point feature detector, SIFT [6] was used. We compared BEEM stand alone, as well as combined with the improved method for detecting degenerate configurations. We tested and evaluated our approach on real sequences for a robotic project. Images are shown in Figure 3. The reconstruced trajectory with the combined approach yield much better results (Figure 5 and 6). On several occasions, lower order models were correctly chosen over the estimated epipolar geometry or other lower order models with more inlier.

Since it is hard to acquire exact ground truth for real camera movement, we also rendered a sequence of images with the Autodesk Maya renderer and evaluated it (Figure 4[1]). We

[1]Maya scene by Christophe Desse and Matthew Thain, `http://www.3drender.com/challenges/index.htm` (visited 08/29/2007).

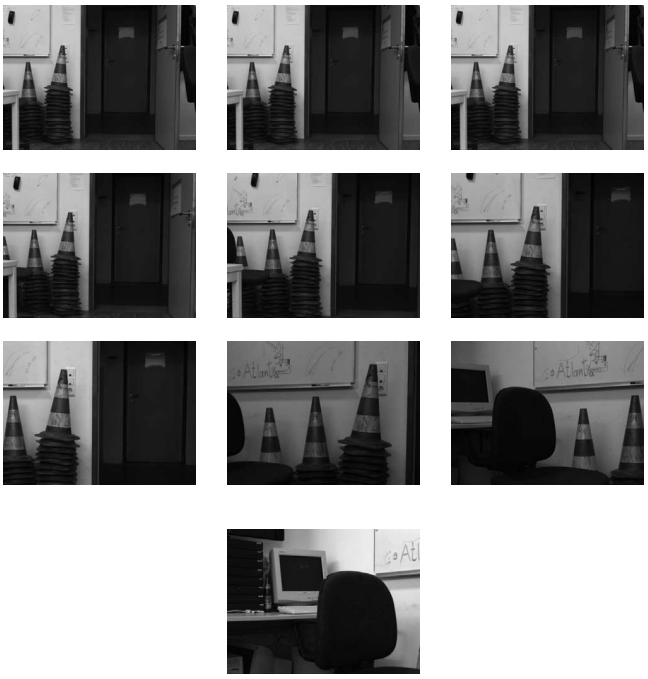


Fig. 3. Image sequence 'lab'. Camera stands still, then moves forward and turns left.

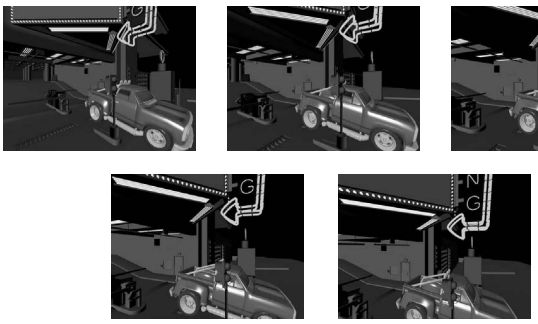


Fig. 4. Image sequence 'neon&chrome'.

measured the error of the estimated translation direction (Δ_t, in degree), rotation axis $\Delta_{\text{R Axis}}$ and rotation angle $\Delta_{\text{R Angle}}$ (axis angle representation, in degree). Because there is still some random element in the BEEM algorithm and our model selection, we did 100 runs and computed the mean μ as well as the root mean square deviation σ of the errors. The results are shown in Table 1. Application of the IDD method leads to significantly less errors.

We are aware of the fact that reconstructing whole trajectories without any kind of bundle adjustment or other techniques to incorporate more than just a pair of frames at a time can not lead to accurate results. But we think, that any improvement at any point before will lead to overall better results.

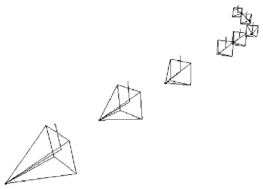


Fig. 5. Reconstructed trajectory of the 'lab' image sequence using our method. Illustrated are the optical center with the field of vision for the estimated camera poses. For the first three frames the pose remains the same, because 'no motion' is detected.

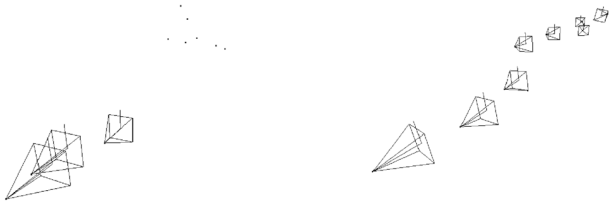


Fig. 6. Two reconstructed trajectories of the 'lab' image sequence without IDD. On the left, the algorithm failes to detect 'no motion' during the first three frames yielding a wrong trajectory of ten camera positions. The trajectory on the right is disturbed because of not considering a pure translation.

4. CONCLUSION

In this paper, we motivated the use of methods to identify two different kinds of problem which occur with degenerate data in epipolar estimation. Two known solutions for these problems were revised and their combination substantiated. Additionally, we improved the method of robust detection of degenerate configurations by enhancing the model selection phase. Experimental data with exact ground truth demonstrated, that the combined approach yields much better results.

Error	With IDD	Without IDD
$\mu(\Delta_t)$	$6,1351$	$19,334$
$\sigma(\Delta_t)$	$3,5150$	$15,378$
$\mu(\Delta_{\text{R Axis}})$	$2,2093$	$12,102$
$\sigma(\Delta_{\text{R Axis}})$	$1,7568$	$11,307$
$\mu(\Delta_{\text{R Angle}})$	$0,68897$	$1,6129$
$\sigma(\Delta_{\text{R Angle}})$	$0,65182$	$1,7704$

Table 1. Difference from ground truth in a rendered scene. The left column shows the results when the improved method for detecting degenerate configurations (IDD) was applied.

5. REFERENCES

[1] Ilan Shimshoni, "Balanced exploration and exploitation model search for efficient epipolar geometry estimations (beem)," in *9th European Conference on Computer Vision (ECCV 2006)*, 5 2006, vol. II, pp. 151–164, BEEM, Code-Demo available at Shimshonis Webpage.

[2] O. Chum, J. Matas, and J. Kittler, "Locally optimized ransac," in *DAGM 2003: Proceedings of the 25th DAGM Symposium*, J. van Leeuwen G. Goos, J. Hartmanis, Ed., Heidelberger Platz 3, 14197, Berlin, Germany, 9 2003, number 2781 in Lecture Notes in Computer Science (LNCS), pp. 236–243, Springer-Verlag.

[3] Berthold K. P. Horn, "Recovering baseline and orientation from essential matrix," 1990.

[4] Richard I. Hartley, "In defense of the eight-point algorithm," *Pattern Analysis and Machine Intelligence*, vol. 19, no. 6, pp. 580–593, 6 1997.

[5] David Nistér, Oleg Naroditsky, and James R. Bergen, "Visual odometry," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2004, pp. 652–659.

[6] David G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[7] O. Chum, Tomá Werner, and J. Matas, "Two-view geometry estimation unaffected by a dominant plane," in *Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR)*, Cordelia Schmid, Stefano Soatto, and Carlo Tomasi, Eds., Los Alamitos, USA, 6 2005, vol. 1, pp. 772–780, IEEE Computer Society.

[8] P. H. S. Torr, A. Zisserman, and S. J. Maybank, "Robust detection of degenerate configurations for the fundamental matrix," *Fifth International Conference on Computer Vision*, pp. 1037–1042, 1995.

[9] Martin A. Fischler and Robert C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[10] O. Chum and J. Matas, "Matching with prosac – progressive sample consensus," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, Cordelia Schmid, Stefano Soatto, and Carlo Tomasi, Eds., Los Alamitos, USA, 6 2005, vol. 1, pp. 220–226, IEEE Computer Society.