Perspective Pose Estimation from Uncertain Omnidirectional Image Data

Christian Gebken Antti Tolvanen Christian Perwass Gerald Sommer

Institut für Informatik, Christian-Albrechts-University of Kiel
Christian-Albrechts-Platz 4, 24118 Kiel, Germany
{chg, ant, chp, gs}@ks.informatik.uni-kiel.de

Abstract

Omnidirectional vision is highly beneficial for robot navigation. We present a novel perspective pose estimation for omnidirectional vision involving a parabolic central catadioptric sensor using small data sets. We incorporate an appropriate and approved stochastic method to deal with uncertainties in the data. Our approach is robust in that it is more accurate than recent methods while using less precise hardware without rigorous calibration.

1. Introduction

Roughly speaking, rigidly moving an object in 3D such that it comes into agreement with 2D-sensory data of a camera, is called 2D-3D pose estimation [3]. It is a well-studied subject in the case of pinhole cameras where sophisticated methods exist, see e.g. [15].

Single viewpoint catadioptric vision sensors combine a conventional camera with one or two mirrors and provide a panoramical view of 360°. Our device is a folded system consisting of two parabolic mirrors and a lens to provide a scaled, approximately orthographic projection from the main mirror. It can equivalently be treated as a single mirror device, see Nayar et al [13].

The most significant advantages of omnidirectional vision are related to navigation. For example, methods of movement estimation from triangulation, topology map and feature flow based methods [1, 4, 7] for localisation give good results on the estimation of movements between frames and the localisation from the visual information. Apart from those methods, 2D-3D pose estimation gives the complete pose information, that is more than a 2D-position in a plane. Since it includes all six possible DOF, it can account for effects like pitch, roll and yaw. Therefore, it represents an appropriate method for navigation, also on uneven surfaces. Furthermore, in the case of omnidirectional pose estimation, the object does not need to be observed within some narrow spatial angle, but may surround the visual sensor itself. This implies a number of advantages. Firstly, an object remains on the image plane under most movements, which is desirable for tasks such as tracking. Secondly, the accuracy of the estimated pose should be superior, as for example in triangulation, which performs best if the used landmarks are seen at right angles. Still, surprisingly few works focus on omnidirectional pose estimation.

It is important to note that the achievable quality from omnidirectional vision sensors ultimately depends on the resolution of the CCD image sensor and on the calibration accuracy; but the major drawback is the small vertical field of view and its non-uniform sampling.

Our objective was to develop accurate pose estimation for omnidirectional vision given few imprecise image points, i.e. 2D-sensory data. By using stochastic estimation, we account for uncertain observations. Besides, one motivation was to take the opportunity to extend approved pinhole methods to the omnidirectional case by exploiting simple existing geometrical relations for parabolic mirrors. In addition, the mathematical framework was already given by an ideally matching algebra.

In the next section, we discuss the pose estimation in some detail. In section 3 we present experimental results. Finally, we give conclusions in section 4.

2. Omnidirectional 2D-3D Pose Estimation

In general, perspective 2D-3D pose estimation consists of determining the orientation and position of an internally calibrated camera (see e.g.[5]), given a 3D-model of an object in a scene and a set of 2D-correspondence points from an image of that scene. The model serves as a reference to an external (world) coordinate system. If we determine the model's position and orientation with respect to the camera coordinate system, we are able to infer the pose of the camera, given by a rigid body motion (RBM). Specifically, we estimate the RBM, such that the model points come to lie on the projection rays of the corresponding image points.

0-7695-2521-0/06/$20.00 (c) 2006 IEEE

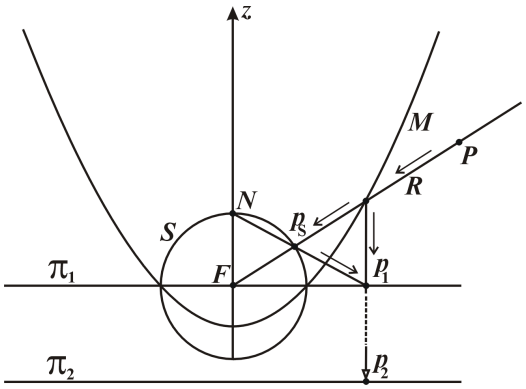

Figure 1. Mapping (cross-section) of a point P: the image planes π_1 and π_2 are identical.

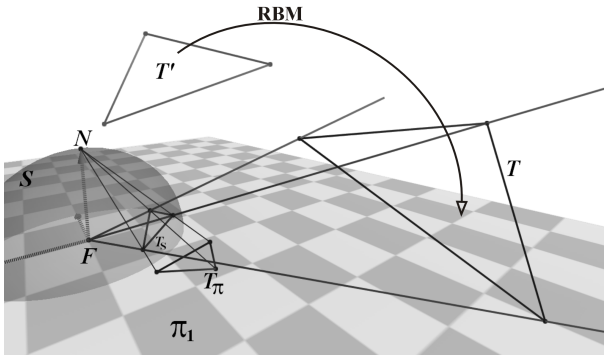

Figure 2. Pose estimation: fitting three model points to the respective projection rays in 3D.

Our method consists of three steps: from those pixels corresponding to model points, we compute the projection rays. As a second step, we use a conventional pose estimation algorithm yielding an initial RBM, which is then improved using a stochastic estimation method. We explain these steps in the following.

2.1. Omnidirectional Projection Rays

The omnidirectional camera setup we consider consists of a camera focussed at infinity, which looks at a parabolic mirror centered on its optical axis. This setup is shown in figure 1. A light ray emitted from point P that would pass through the focal point F of the parabolic mirror M, is reflected parallel to the central axis of the parabolic mirror, to give point p_2. Since all such reflected rays are parallel, a camera placed beneath the mirror focussed at infinity will generate a sharp image on plane π_2. Here, we use the simplification that a projection to sphere S with a subsequent stereographic projection to π_1 produces an identical image on π_1. Accordingly, point P maps to p_s and further to p_1, see figure 1. For details refer to [8, 16].

Given image points, we can now apply these steps in reverse order to obtain the corresponding projection rays. Consider for instance figure 2: three image points form the triangle T_π, which is stereographically back-projected to the triangle T_S on S. In conjunction with F, we obtain the projection rays. The correct RBM then moves the model triangle T', so that it comes to lie on the projection rays.

2.2. Initial Pose Estimation

We compute initial RBM estimates using a 3-point (3P-) technique, which merely considers three model points and associated projection rays. Note that there exist a vast number of 3P-algorithms; in [9] Haralick et al provide a thorough overview of six methods. An example setup is depicted in figure 2, where the model T' has been fitted to the projection rays, yielding T. The desired RBM between T' and T is unique and can be computed easily. The well-known ambiguity in the 3P-fitting itself can usually be lifted, if we consider at least one additional correspondence pair for validation. Fischler et al, for example, used a similar strategy in [6]. Besides, we do multiple 3P-runs on different subsets and build the mean of the respective RBMs. This then serves as a starting value for the stochastic estimation described in the next section.

2.3. Stochastic Estimation Method

In the final pose estimation step we improve on the initial RBM estimate by taking into account the uncertainties of the image points. This can be done using the Gauss-Helmert method from the area of least squares adjustment, which is founded on the homonymic linear model, introduced by Helmert [10] in 1872. As a by-product, one obtains a covariance matrix for the estimated parameters, cf. [12]. Details on this parameter estimation method can be found in [14].

The geometric constraint equation used is derived from the Geometric Algebra of conformal space $\mathbb{G}_{4,1}$ [2, 11]. A similar methodology was chosen by Rosenhahn et al [15]. The products used in the following are the *geometric product*, which is the main algebra product, and the *outer product*, which is in no way related to the outer product of matrices. The geometric product is denoted by juxtaposition and the outer product by $\wedge$.

A similar pose estimation could also be done solely in Euclidian 3D-space, but we obtain certain advantages when working in $\mathbb{G}_{4,1}$: points, spheres, lines and RBMs are basic elements of $\mathbb{G}_{4,1}$ and have thus a natural representation in terms of (sparse) vectors of $\mathbb{R}^{2^5}$. Moreover, incidence relations, as needed to decide whether a point lies on a projec-

tion ray, can be expressed by means of a nullspace regarding a bilinear algebra product.

In order to obtain the point p_s from the image point p_1, we need to perform the stereographic back-projection onto the sphere S. With respect to figure 1, this can also be modeled by performing an inversion in the sphere centered on N with radius $\sqrt{2}\,r$, where r denotes the radius of S. In $\mathbb{G}_{4,1}$ this can be written as $p_s = S_I\, p_1\, S_I$, where S_I denotes the inversion sphere and the product between the entities is the geometric product. The projection ray R can now be constructed as $R = F \wedge p_s \wedge \mathbf{e}_\infty$, where $\mathbf{e}_\infty$ is a particular basis vector of $\mathbb{G}_{4,1}$. Since point P lies on R, we have $R \wedge P = 0 \in \mathbb{G}_{4,1}$. A model point P' is transformed by an RBM represented by V, say, via the operation $V\, P'\, \tilde{V}$, where the *reverse* $\tilde{V}$ is similar to conjugation in quaternions. Therefore, if we have estimated the correct RBM V, a model point P' with corresponding image point p_1 has to satisfy $F \wedge (S_I\, p_1\, S_I) \wedge \mathbf{e}_\infty \wedge (V\, P'\, \tilde{V}) = 0$.

Each algebra product is a bilinear function and can be formulated equivalently using a certain tensor. By contracting the constituent tensors the constraint equation can be written in terms of eight particular components $(\mathsf{v}^1, \ldots, \mathsf{v}^8)$ of V in the following way.

$$\sum_{i=1}^{8}\sum_{j=1}^{8} \mathsf{v}^i\, \mathsf{v}^j\, Q^k{}_{ij} = 0, \quad k \in \{1, \ldots, 5\}. \qquad (1)$$

There exists one tensor $Q^k{}_{ij}$ for each pair of corresponding image and model points. Algebraically, the constraint $R \wedge P$ may only be nonzero in five of its $2^5 = 32$ components, which explains k's range. Note that the eight components of V are an overparameterization of the six DOF of an RBM, such that we need to include the RBM-constraint $V\,\tilde{V} = 1$ in the minimization process, which also turns out to be a bilinear function of the components of V. Such additional constraints can be readily included in the Gauss-Helmert method. The pixel uncertainties are incorporated in terms of 2×2 covariance matrices. Note that due to the linear nature of the underlying stochastic model, the constraints, which are quadratic in the components of V, have to be linearized. Hence, the estimation becomes an iterative process, and we start from the initial 3P-pose, which needs not be particularly accurate for converging stably.

3. Experimental results

Two sets of real world experiments were performed using an imaging system consisting of a `Sony DXC-151AP` camera and `Remote Reality Netvision 360 catadioptrics` mirror device. As intrinsic calibration parameters we used the $40\,$mm mirror radius given by the manufacturer and $18.05\,$mm focal length for the main mirror given by calibration in the experimental setup. The projection was assumed to be ideal orthographic. Images were acquired in 768×576 resolution where the actual size of the omnidirectional image is $570 \times 540\,$pixels. The markers' pixel coordinates were extracted manually, where we assumed a standard deviation of one pixel. The respective 2×2 covariance matrices, as required by the stochastic estimation, were set accordingly.

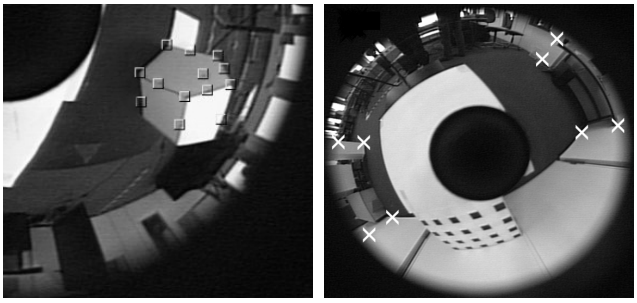

Figure 3. Examples from the experiments: the model house (left) and the room navigation (right) experiment. The respective model points are graphically highlighted.

In the first set of experiments we rotated a model house with $35.1\,$cm (A) and $52.4\,$cm (B) orthogonal distances from the optical axis. The house dimensions in cm are $21 \times 15 \times 21$, see left in figure 3. Note that only a quarter of the image is shown. In order to simplify the acquisition of ground truths the rotation plane was perpendicular to the optical axis. The house was rotated in $10°$ steps from $60°$ to $0°$. We measured the respective errors relative to the $60°$-rotation. The results are listed in table 1. Note that the house appears flat in the $0°$-image, i.e. the usable 3D-model points are nearly coplanar affecting the estimation result. The mean error in the rotation was $1.65°$ and the mean error in the planar distance was $0.43\,$cm. The estimated height of the sensor w.r.t. the house was $27.5 \pm 0.4\,$cm, which is within the measurement error of ground truth $27.55 \pm 0.4\,$cm.

Rotation	Abs. error angle		Abs. error distance	
[°]	A, [°]	B, [°]	A, [cm]	B, [cm]
50 to 60	2.8	0.4	0.65	0.10
40 to 60	2.6	0.2	0.40	0.30
30 to 60	1.3	0.5	0.54	0.33
20 to 60	0.3	0.8	0.70	0.04
10 to 60	0.7	6.4	0.90	1.12
0 to 60	3.0	6.3	3.90	1.20

Table 1. Results of the house pose estimation. Distance A$=35.1\,$cm, B$=52.4\,$cm.

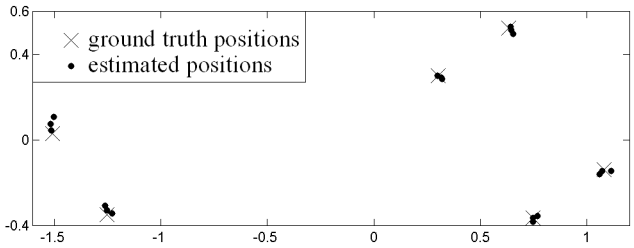

Figure 4. Pose estimation inside the object (room navigation). Units in meters.

In the second set of experiments the sensor was moved to six positions inside a $5.3 \times 2.2\,\mathrm{m}^2$ room. The room model, as indicated by white crosses in figure 3, was defined by four vertical pairs of markers. In each of the six positions the sensor was rotated by $0°$, $30°$ and $70°$. The results are illustrated in figure 4. The error regarding the planar distances to the ground truth positions was $2.45 \pm 1.74\,\mathrm{cm}$. The estimated height of the sensor had an error of $3.22 \pm 1.0\,\mathrm{cm}$. These are comparable results to those given by Aliaga [1]; he obtained an average planar error of $2.8\,\mathrm{cm}$ within a room being very similar in size. Nevertheless, Aliaga made a thorough calibration and used a high-precision 3-CCD chip camera, the resolution of which was 1360×1024. Cauchois et el [4] obtained as good results as we taking the room size into account, which was noticeably smaller. Further sensible comparisons are hard due to the limited number of usable publications.

In addition to the large movements, we studied the accuracy in estimating smaller displacements and rotations in the center of the room separately. The sensor was translated on a line in 1, 5 and $10\,\mathrm{cm}$ steps giving a mean error of $0.5\,\mathrm{cm}$. The error did not depend on the length of the translation. The camera was rotated in $5°$ steps from a starting position to $90°$ giving a mean error of $0.4°$.

4. Conclusions

The aim of this work was to realize 2D-3D pose estimation for omnidirectional vision. We combined geometry and stochastics such that we obtain precise estimates from few image points. Our experiments substantiated that the achieved accuracy in 2D-navigation is somewhat better than those of comparable approaches [1, 4]; in contrast to those we used an of-the-shelf catadioptric device and a low-resolution camera without rigorous calibration. Due to the promising results of our house experiments, it is likely that central catadioptric sensors may be used instead of pinhole cameras in many pose estimation applications. In future research, a robot will be equipped with a state-of-the-art omnidirectional vision sensor for autonomous navigation experiments.

References

[1] D. G. Aliaga. Accurate catadioptric calibration for real-time pose estimation of room-size environments. In *International Conference on Computer Vision (ICCV)*, pages 127–134, 2001.

[2] P. Angles. Construction de revêtements du groupe conforme d'un espace vectoriel muni d'une ≪métrique≫ de type (p, q). *Ann. Inst. Henri Poincaré*, 33(1):33–51, 1980.

[3] A. Ansar and K. Daniilidis. Linear pose estimation from points or lines. In *7th European Conference on Computer Vision (ECCV), Copenhagen, Denmark*, pages 282–296, 2002.

[4] C. Cauchois, E. Brassart, L. Delahoche, and C. Drocourt. Spatial localization method with omnidirectional vision. In *11th IEEE International Conference on Advanced Robotics (ICAR), Coimbra, Portugal*, pages 287–292, 2003.

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

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

[7] J. Gaspar and J. Santos-Victor. Vision-based navigation and enviromental representations with an omni-directional camera. *IEEE Transactions on Robotics and Automation*, 16(6):890–898, 2000.

[8] C. Geyer and K. Daniilidis. Catadioptric projective geometry. *International Journal of Computer Vision*, 45(3):223–243, 2001.

[9] R. M. Haralick, C.-N. Lee, K. Ottenberg, and M. Noelle. Review and analysis of solutions of the three point perspective pose estimation problem. *International Journal on Computer Vision*, 13(3):331–356, 1994.

[10] F. Helmert. *Die Ausgleichsrechnung nach der Methode der kleinsten Quadrate*. Teubner, Leipzig, 1872.

[11] D. Hestenes and G. Sobczyk. *Clifford Algebra to Geometric Calculus: A Unified Language for Mathematics and Physics*. Reidel, Dordrecht, 1984.

[12] K.-R. Koch. *Parameter Estimation and Hypothesis Testing in Linear Models*. Springer, 1997.

[13] S. K. Nayar and V. Peri. Folded catadioptric cameras. In *Conference on Computer Vision and Pattern Recognition (CVPR), Ft. Collins, CO, USA*, pages 2217–, 1999.

[14] C. Perwass, C. Gebken, and G. Sommer. Estimation of geometric entities and operators from uncertain data. In *27. Symposium für Mustererkennung, DAGM 2005, Wien, 29.8.-2.9.005*, number 3663 in LNCS. Springer-Verlag, Berlin, Heidelberg, 2005.

[15] B. Rosenhahn and G. Sommer. Pose estimation in conformal geometric algebra, part I: The stratification of mathematical spaces, part II: Real-time pose estimation using extended feature concepts. *Journal of Mathematical Imaging and Vision*, 22:27–70, 2005.

[16] A. Tolvanen, C. Perwass, and G. Sommer. Projective model for central catadioptric cameras using clifford algebra. In *27. Symposium für Mustererkennung, DAGM 2005, Wien, 29.8.-2.9.005*, volume 3663 of *LNCS*, pages 192–199. Springer-Verlag, Berlin, Heidelberg, 2005.