

Pose Estimation from Uncertain Omnidirectional Image Data Using Line-Plane Correspondences

Christian Gebken, Antti Tolvanen, and Gerald Sommer

Institut für Informatik, CAU Kiel
Christian-Albrechts-Platz 4, 24118 Kiel, Germany
{chg, ant, 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 line-plane correspondences. We incorporate an appropriate and approved stochastic method to deal with uncertainties in the data.

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 for which sophisticated methods exist, see e.g. [13].

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 [10].

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, 6] for localization give good results on the estimation of movements between frames and the localization 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 degrees of freedom (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. First, an object remains on the image plane under most movements, which is desirable for tasks such as tracking. Second, 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 little research was done on omnidirectional pose estimation.

Our objective was to develop accurate pose estimation for omnidirectional vision given imprecise image features, i.e. 2D-sensory data. The motivation was

to take the opportunity to extend approved pinhole methods to the omnidirectional case by exploiting simple existing geometrical relations for parabolic mirrors. The stochastic is one of the fundamental aspects of this work; to account for invariable uncertainties in observational data we consequently decided on a least squares adjustment parameter estimation. The concept of our approach is a well-tried amalgamation of geometry with stochastic via Geometric Algebra.

One assumption we make is to have 3D-models of the interesting objects we observe in the images. This can be an ordinary object like a table or it is a model describing the environment. Secondly, we assume to know the one-to-one correspondences between the model features and the image features.

Note that with this contribution we extend our previous work by using line models instead of point models. The matching image entities are therefore lines. Recognition and localization is simpler for lines than for points, since those are intrinsically higher-dimensional structures. Localization is more precise for lines, as well. Regarding regular structures, like a skyscraper, it is more efficient to have line models than to store the corners of each single window. We can state that the existence of key points, e.g. corners, mostly inheres with the existence of lines which are then the preferable entities.

In the next section, we discuss the pose estimation and all related topics 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 [5], given a 3D-model of an object in a scene and a set of 2D-correspondence features (points, lines, curves) 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 lines come to lie on the projection planes of the underlying image lines.

We use the Geometric Algebra $\mathbb{G}_{4,1}$ of the conformal embedding of Euclidean 3D-space as introduced in [2, 8]. A similar pose estimation could also be done solely in Euclidian 3D-space, but we obtain certain advantages when working in $\mathbb{G}_{4,1}$: geometric entities as points, spheres, planes or lines and geometric operators as an inversion or an RBM are basic elements of $\mathbb{G}_{4,1}$. They have thus a natural representation in terms of (sparse) vectors of \mathbb{R}^{2^5} . Moreover, incidence relations, as needed to decide whether a line lies on a projection plane, can be evaluated by means of bilinear algebra products. Nevertheless, for understandability to unfamiliar readers, and since it is not the main subject of this work, we make explicit use of Geometric Algebra in just one passage. In practice we employ the framework of Geometric Algebra throughout all steps of

our method. A general introduction to the estimation of geometric entities and operators from uncertain data using Geometric Algebra can be found in [12].

Our method consists of three steps: from those pixels corresponding to model lines, we compute the projection planes with associated uncertainties. In a second step, a simple algorithm is used to do prior rotation estimation being a first and rough guess at the rotational part of the desired RBM. As a result the model will be aligned such that its lines are nearly parallel to the respective projection planes. Next, an iterative method estimates the entire pose now taking also the plane uncertainties into account.

Before we explain those steps we give an overview regarding catadioptric imaging with a parabolic mirror.

2.1 Omnidirectional Imaging

Despite our interest in the mapping of lines to the image we begin with the point case. The omnidirectional camera setup we consider consists of a camera focused 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

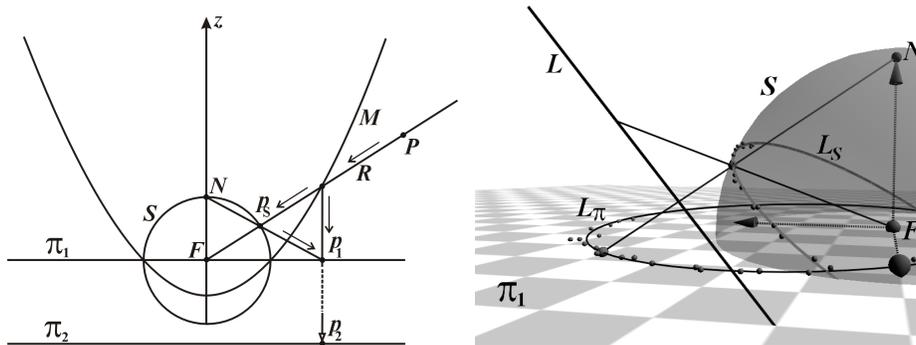


Fig. 1. **Left:** Mapping (cross-section) of a point P : the image planes π_1 and π_2 are identical. **Right:** Mapping of line L to L_π via great circle L_S on S . As an example, scattered image data belonging to L_π is shown.

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 focused 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. Together with the right side of figure 1 it is intuitively clear that infinitely extended lines form great circles on S . Moreover, a subsequent

can thus be restricted to three components representing the normal of the circle plane.

We assume that all image points initially have the same 2D-uncertainty given by a 2×2 identity covariance matrix, i.e. we assume an pixel error of one in row and column. Since the planes have to be estimated from the stereographically back-projected image points, see figure 1, we must move the involved image points to the projection sphere S . This is done by an inversion of the image points in a certain sphere. The points thereby obtain distinct 3D-uncertainties accounting for the imaging geometry. The mapping of a far image point to a point close to the North Pole N of S , for example, is less affected by noise and will thus inhere with a higher confidence, see figure 1. Mathematically, the uncertainties are computed using standard error propagation, where we profit from the inversion being an element of $\mathbb{G}_{4,1}$.

Since our estimation method is capable of providing a covariance matrix regarding the estimated entity, we obtain a 3×3 covariance matrix for each plane. Those are then to be reinput to our pose estimation algorithm.

2.3 Prior Model Alignment

Estimation problems mostly require a linearization of condition or constraint functions and one usually ends up with an iterative method, as we do. This raises the need for a sufficiently good initial estimate regarding the iterations. The prior model alignment provides such a starting point at very low costs. Moreover, it shortens the overall computation time. We like to rotate the model such that the unit direction vectors $\{\hat{r}_{1..N}\}$ of its lines lie on the respective planes. Here, a prerequisite is to have the normal vectors $\{\hat{n}_{1..N}\}$ of all planes belonging to visible model lines. We search for a rotation matrix R such that $(\forall i): \hat{n}_i^T R \hat{r}_i = 0$.

By Rodrigues's formula (1840) we know that the rotation matrix R regarding a rotation of angle θ around unit vector $\hat{a} = (a_1, a_2, a_3)^T$ can be expressed by an exponential map of $A = ((0, a_3, -a_2)^T (-a_3, 0, a_1)^T (a_2, -a_1, 0)^T)$: $R = \exp(\theta A)$ which is $R = I_3 + \sin \theta A + (1 - \cos \theta) A^2$. For small angles we obtain $R = I_3 + \theta A$. With this relation and due to the skew symmetric structure of $A' = \theta A$ it is possible to solve for $\mathbf{a}' = (\theta a_1, \theta a_2, \theta a_3)^T$, where each line-plane pair gives one line $\hat{n}_i^T A' \hat{r}_i = -\hat{n}_i^T \hat{r}_i$ in an overdetermined system of linear equations. Every run of this procedure yields a rotation matrix, the concatenation of which gives the desired rotation matrix R . Once, the rotated lines are close enough to the planes w.r.t. some threshold the procedure can be stopped.

2.4 Stochastic Estimation Method

In this section we concisely introduce our two parameter estimation methods, the common *Gauss-Markov* method and the most generalized case of *least squares adjustment*, the *Gauss-Helmert* method. Both are founded on the respective homonymic linear models, cf. [9]. The word 'adjustment' puts emphasis on the

fact that an estimation has to handle redundancy in observational data appropriately, e.g. to weight unreliable data to a lesser extend. The principle of least squares adjustment, i.e. to minimize the sum of squared weighted errors Δy_i , is often denoted as

$$\sum_i \Delta y_i^\top \Sigma_{y_i}^{-1} \Delta y_i \longrightarrow \min, \quad (1)$$

where Σ_{y_i} is a covariance matrix assessing the confidence of y_i .

Let $\{\mathbf{b}_{1..N}\}$ ³ denote a set of N observations. Each observation \mathbf{b}_i is associated with an appropriate covariance matrix $\Sigma_{\mathbf{b}_i}$ denoting the confidence. An entity, parameterized by a vector \mathbf{p} , is to be fitted to the observational data. Consequently, we define a condition function $\mathbf{g}(\mathbf{b}_i, \mathbf{p})$, which is supposed to be zero if the observations and the entity in demand fit. If we know an already good estimate $\hat{\mathbf{p}}$ we can make a linearization yielding $(\partial_{\mathbf{p}} \mathbf{g})(\mathbf{b}_i, \hat{\mathbf{p}}) \Delta \mathbf{p} + \mathbf{g}(\mathbf{b}_i, \hat{\mathbf{p}}) \approx 0$, hence with $\mathbf{U}_i = (\partial_{\mathbf{p}} \mathbf{g})(\mathbf{b}_i, \hat{\mathbf{p}})$ and $y_i = -\mathbf{g}(\mathbf{b}_i, \hat{\mathbf{p}})$: $\mathbf{U}_i \Delta \mathbf{p} = y_i + \Delta y_i$, which exactly matches the linear *Gauss-Markov* model. The minimization of equation (1) in conjunction with the Gauss-Markov model leads to the *best linear unbiased estimator*⁴. Note that we have to leave the weighting out in equation (1), since our covariance matrices $\Sigma_{\mathbf{b}_i}$ do not match the Σ_{y_i} . Subsequently, we derive a model which includes the weighting.

If we take our observations as estimates, i.e. $\{\hat{\mathbf{b}}_{1..N}\} = \{\mathbf{b}_{1..N}\}$, we can make a complete Taylor series expansion of first order at $(\hat{\mathbf{b}}_i, \hat{\mathbf{p}})$ yielding

$$(\partial_{\mathbf{p}} \mathbf{g})(\hat{\mathbf{b}}_i, \hat{\mathbf{p}}) \Delta \mathbf{p} + (\partial_{\mathbf{b}} \mathbf{g})(\hat{\mathbf{b}}_i, \hat{\mathbf{p}}) \Delta \mathbf{b}_i + \mathbf{g}(\hat{\mathbf{b}}_i, \hat{\mathbf{p}}) \approx 0.$$

Similarly, with $\mathbf{V}_i = (\partial_{\mathbf{b}} \mathbf{g})(\hat{\mathbf{b}}_i, \hat{\mathbf{p}})$ we obtain $\mathbf{U}_i \Delta \mathbf{p} + \mathbf{V}_i \Delta \mathbf{b}_i = y_i$, which exactly matches the linear *Gauss-Helmert* model. Note, that the error term Δy_i has been replaced by the linear combination $\Delta y_i = -\mathbf{V}_i \Delta \mathbf{b}_i$: the Gauss-Helmert differs from the Gauss-Markov model, because the observations have become random parameters and are thus allowed to undergo small changes $\Delta \mathbf{b}_i$ to compensate for errors. But changes have to be kept minimal, as observations represent the best available. This is achieved by replacing equation (1) with

$$\sum_i \Delta \mathbf{b}_i^\top \Sigma_{\mathbf{b}_i}^{-1} \Delta \mathbf{b}_i \longrightarrow \min, \quad (2)$$

where $\Delta \mathbf{b}_i$ is now considered as error vector. The minimization of (2) subject to the Gauss-Helmert model can be done using Lagrange multipliers, cf. [9].

Due to outstanding convergence properties we start iterating with the Gauss-Markov method. At the optimum we start the slower Gauss-Helmert method which ultimately adjusts the estimate according to the given uncertainties $\Sigma_{\mathbf{b}_i}$.

³ We use the abbreviation $\{\mathbf{b}_{1..N}\}$ for a set $\{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_N\}$.

⁴ It has been shown in [9] that different approaches, namely *least squares*, *maximum likelihood* and the linear approach, equally lead to the best linear unbiased estimator.

2.5 Perspective Line-Plane Pose Estimation

Here we derive geometric constraint equations for the stochastic estimation methods presented in the previous section. The respective expressions come from the Geometric Algebra of conformal space $\mathbb{G}_{4,1}$. A similar methodology was chosen by Rosenhahn et al [13]. 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 .

Let P be a projection plane, see section 2.2. For any line L lying on P , we have $P \wedge L = 0 \in \mathbb{G}_{4,1}$. A model line L' is transformed by an RBM represented by V , say, via the operation $V L' \tilde{V}$, where the *reverse* \tilde{V} is similar to conjugation in quaternions. Therefore, if we have estimated the correct RBM V , a model line L' with corresponding projection plane P has to satisfy $P \wedge (V L' \tilde{V}) = 0$.

Due to the numerical representation of $\mathbb{G}_{4,1}$, we can identify our elements P , L' and V with particular vectors $\mathbf{p} \in \mathbb{R}^3$, $\mathbf{l}' \in \mathbb{R}^6$ and $\mathbf{v} \in \mathbb{R}^8$. For example, p simply denotes the normal vector of the plane represented by P . Moreover, each algebra product is a bilinear function and can be formulated equivalently using a certain tensor, cf. [12]. By contracting the constituent tensors the condition function \mathbf{g} of the previous section can be written in the following way

$$\mathbf{g}^t(\mathbf{p}, \mathbf{v}) := \sum_{i,j,k,l} v^i v^j p^k l'^l Q^t_{ijkl} = 0, \quad t \in \{1 \dots 4\}. \quad (3)$$

Algebraically, the constraint $P \wedge L$ may only be nonzero in four of its $2^5 = 32$ components, i.e. we have $t \in \{1 \dots 4\}$. The observations and parameters are \mathbf{p} and \mathbf{v} , respectively. Hence, differentiating would yield the matrices \mathbf{V} and \mathbf{U} required in section 2.4. 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 our parameter estimation methods.

3 Experimental Results

Two real world experiments were performed using an imaging system consisting of a **Kamerawerk Dresden Loglux i5** camera and **Remote Reality Netvision 360** catadioptric sensor with a parabolic mirror. The aim of the experiments was to test object pose estimation and navigation and the robustness of the used methods in these tasks. As intrinsic calibration parameters we used the 40 mm mirror radius and 16.7 mm focal length for the main mirror given by the manufacturer. The projection of the sensor was assumed to be exactly orthographic and the whole mirror was assumed to be visible in the image. Images were acquired in 1280×1024 resolution where the actual size of the omnidirectional image is the area of a circle with 492 pixel radius corresponding to the 40 mm mirror radius. The radius and the center of the image were determined from

the sum of images used in the experiments. No other calibration was done. The image lines were extracted manually with seven points/line.

In the first experiment a model house was moved with a robot arm to 21 different locations. The robot arm gives ground truth of the translations between the different locations with millimeter accuracy. The magnitude of these translations was between 7.7 cm and 62.4 cm and the distance of the model house to the optical center of the catadioptric sensor was between 31.4 cm and 82.8 cm. The house dimensions in cm are $21 \times 15 \times 21$. From the 21 acquired images the RBMs of the model house from the optical center were estimated. These estimates were used to get the relative translation estimates between the different model house positions. The results are given in table 1.

Table 1. The errors of the house pose estimation.

	Abs. error [mm]	Rel. error [%]	Angle error [°]
mean	10.4	3.5	0.9
std	4.8	1.7	0.4
min	0.9	0.4	0.12
max	21.3	11.5	2.4

In the second experiment the sensor was moved to 25 different positions in a hallway. The model was defined by lines clearly visible in most of the images. The other criterion was reasonable measurability needed to create the model. The walls were assumed to be perpendicular to the floor and all corners to be right angled. With these assumptions we get roughly 2 cm accuracy for positions of the model lines. The model consisted of total 51 lines from which on average 20 lines were visible in an image. The maximum orthogonal distance of these lines was 18.1 m, minimum 3.8 m and the sensor movements were made on 8×2 m² area inside the model. The results for the error in the position are given in table 2 for Gauss-Markov (G-M) and Gauss-Helmert (G-H) methods in 2D and 3D. Figure 3 on left represents the results for G-M and G-H methods and the groundtruth (Truth) in 3D. In addition to the pose estimation with the given parameters we studied the robustness of the used methods in respect to the change of the focal length of the mirror (see figure 3 on right). It can be seen that the G-H method is always better than G-M and slightly more robust. 2D estimation works always better as the error source is one dimension smaller and the estimation relies mostly on the vertical world lines whose image remains almost unchanged with the change of focal length. Using 0.1 mm steps for the focal length gives the most accurate results for G-M 3D 5.8 cm with focal length 16.8 mm and for G-H 3D 4.2 cm with focal length 16.9 mm.

Comparisons for 3D results are hard due to the limited number of usable publications. The results in 2D are comparable results to those given by Aliaga

Table 2. The errors of the navigation.

	Mean error [cm]	RMS error [cm]	min [cm]	max [cm]
G-M 3D	7.6	9.4	3.6	32.2
G-M 2D	5.1	7.7	0.4	32.0
G-H 3D	6.4	6.5	2.7	8.3
G-H 2D	3.5	3.9	0.5	5.7

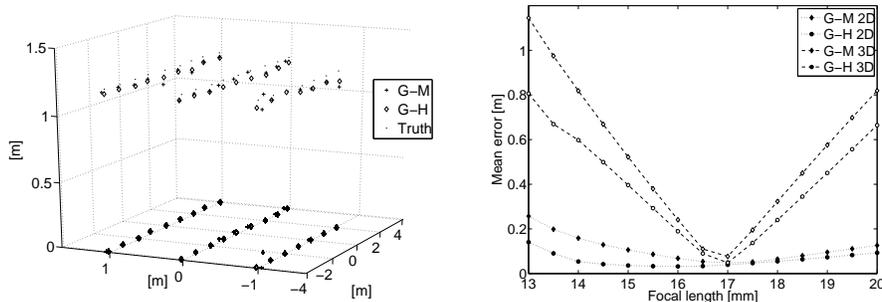


Fig. 3. **Left:** navigation results. The 3D positions are also projected to plane for clarity. **Right:** focal length vs. mean error.

[1]; he obtained an average planar error of 2.8 cm within a room of 5 meters diameter using a triangulation method and with exact calibration of the system. Cauchois et al [4] reached about 1 cm accuracy in 2D using an image database method with a conical mirror and a room of $2 \times 3 \text{ m}^2$.

4 Conclusions

The objective of this work was to realize 2D-3D pose estimation for omnidirectional vision using line-plane correspondences. The pose was computed by a stochastic estimation method, which accounts for uncertainties in the image data.

The experimental results clearly demonstrate that our combination of 2D-3D pose estimation with omnidirectional vision does provide exact results for navigation within relatively big environments. The results of our house experiments show that we still obtain good results, if we utilize our method for conventional 2D-3D object pose estimation.

Especially the 2D-navigation was found out to be very robust in respect to changes of focal length. The change of the focal length scales the image radially. Since the images of vertical world lines are radial lines in the image they are

invariant in this scaling. On the other hand the positions of image points on the radial lines are not invariant. This motivates studies on the differences in the robustness of point-line and line-plane pose estimation in 2D-navigation.

In the future we would like to automate the point extraction from the image in order to construct a ready to use method for robotics. This is plausible as the calculation time for the pose estimation (including 3D-visualization) is under 1 second using a scripting language (CLUCalc, see [11]) on a 3 GHz Pentium 4 computer.

References

1. Daniel 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. Pierre 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. Adnan Ansar and Konstantinos Daniilidis. Linear pose estimation from points or lines. In *7th European Conference on Computer Vision (ECCV), Copenhagen, Denmark*, pages 282–296, 2002.
4. Cyril Cauchois, Eric Brassart, Laurent Delahoche, and Cyril Drocourt. Spatial localization method with omnidirectional vision. In *11th IEEE International Conference on Advanced Robotics (ICAR), Coimbra, Portugal*, pages 287–292, 2003.
5. Olivier Faugeras. *Three-Dimensional Computer Vision*. MIT Press, 1993.
6. Jos Gaspar and Jos Santos-Victor. Vision-based navigation and environmental representations with an omni-directional camera. *IEEE Transactions on Robotics and Automation*, 16(6):890–898, 2000.
7. Christopher Geyer and Kostas Daniilidis. Catadioptric projective geometry. *International Journal of Computer Vision*, 45(3):223–243, 2001.
8. D. Hestenes and G. Sobczyk. *Clifford Algebra to Geometric Calculus: A Unified Language for Mathematics and Physics*. Reidel, Dordrecht, 1984.
9. K.-R. Koch. *Parameter Estimation and Hypothesis Testing in Linear Models*. Springer, 1997.
10. Shree K. Nayar and Venkata Peri. Folded catadioptric cameras. In *Conference on Computer Vision and Pattern Recognition (CVPR), Ft. Collins, CO, USA*, pages 2217–, 1999.
11. C. Perwass, C. Gebken, and D. Grest. CLUCalc. <http://www.clucalc.info/>, 2006.
12. 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.
13. 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.
14. Antti Tolvanen, Christian Perwass, and Gerald 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.