Perspective Pose Estimation with Geometric Algebra

Christian Gebken and Gerald Sommer

Institute of Computer Science, Chair of Cognitive Systems, Christian-Albrechts-University of Kiel, Germany

Abstract. A novel method which entirely resides inside conformal geometric algebra (CGA) is presented estimating the pose of a camera from one image of a known object. At first, subproblems covering only three feature points are solved and globally assessed. The object model is accordingly pruned and rigidly fitted to corresponding projection rays by evaluating a succinct CGA expression which emerged from a purely geometric approach. It results a set of 3-point poses each given by a motor. These spinor elements of CGA embody rigid body motions from the manifold $SE(3)$. The poses are then to be averaged according to their quality. This is the second aspect of this work as the respective motors do not come from a linear space and averaging must be carried out appropriately. For this purpose, a technique called weighted intrinsic mean is used.

Keywords: geometric (Clifford) algebra, conformal space, geometric calculus, perspective projection, intrinsic mean, Lie groups
PACS: 02.10.Xm, 02.20.Sv, 02.40.Dr

INTRODUCTION

In pose estimation the orientation and position of an internally calibrated camera is recovered from its images. For this purpose, the 3D point-model of at least one pictured object is assumed to be known together with a set of correspondences, which interrelate model points and image points. This kind of pose estimation is often referred to as 'perspective *n*-point problem' (PnP).

The classic but challenging task of pose estimation is from the field of computer vision. Most approaches to that subject are iterative, highly nonlinear or require an initialization. Closed form solutions to the 3-point problem (P3P), where the number of correspondences is three, exist [1] but may result in up to four distinct solutions because P3P is not necessarily unique. As extension to P3P it is also possible to consider four points. Fischler and Bolles [2], for example, take subsets and perform consistency checks to eliminate the P3P ambiguity for most point configurations. In [3] Quan and Lan present an algorithm capable of finding the unique solution to PnP. They first generate a global system of linear equations based on all correspondences. Next, the exact 3D-vectors to the object points w.r.t. the camera coordinate system are estimated. Finally, camera orientation and position are evaluated one after another. But this class of techniques is shown in [4] to improperly model the physical imaging, i.e. a perspective projection must be considered. Rosenhahn and Sommer [5] formulate algebraic constraints with CGA. They obtain a hybrid system of linear equations based on correspondences between points, lines and between point and line. Starting from an initialization the pose is iteratively estimated in 3D. It is to mention that such global PnP approaches are not able to spot and disregard false or noisy correspondences.

In this text we derive a vivid geometric formulation of P3P with CGA. At the same time, this motivates a sound selection strategy for point triplets, i.e. not all possible 3-combinations in the correspondences must be considered. Solutions of PnP are rigid body motions (RBM) from the manifold $SE(3)$. We show that the respective P3P-solution is fully determined by a certain angle θ^*. Our geometric approach further leads to an algebraic function $h(\theta) \in \mathbb{R}$, with θ^* being a root of which. For each root the corresponding RBM is globally assessed regarding its effect on the entire *n*-point scenario. The set of 3-point candidate solutions can then be reduced by solutions from obviously false correspondences. The remaining RBMs, at most one for every triplet considered, are finally averaged by means of the weighted intrinsic mean, which is tailored to elements of $SE(3)$.

Geometry with Geometric Algebra

For a detailed introduction to geometric algebra (GA) see e.g. [6, 7]. Here we only convey a minimal framework. We consider the geometric algebra $\mathbb{G}_{4,1} = \mathscr{C}\ell(\mathbb{R}^{4,1}) \supset \mathbb{R}^{4,1}$ of the 5D conformal space (CGA), cf. [8, 9]. Let $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_+, \mathbf{e}_-\}$ denote the basis of the underlying Minkowski space $\mathbb{R}^{4,1}$, where $\mathbf{e}_1^2 = \mathbf{e}_2^2 = \mathbf{e}_3^2 = \mathbf{e}_+^2 = 1 = -\mathbf{e}_-^2$. We use the common abbreviations $\mathbf{e}_\infty = \mathbf{e}_- + \mathbf{e}_+$ and $\mathbf{e}_o = \frac{1}{2}(\mathbf{e}_- - \mathbf{e}_+)$ for the point at infinity and the origin, respectively. The algebra elements are termed multivectors, which we symbolize by capital bold letters, e.g. $\boldsymbol{A}$. A juxtaposition of algebra elements, like $\boldsymbol{CR}$, indicates their geometric product, being the associative, distributive and noncommutative algebra product. We denote '$\cdot$' and '$\wedge$' the inner and outer (exterior) product, respectively. A point $\boldsymbol{x} \in \mathbb{R}^3 \subset \mathbb{G}_{4,1}$ of the 3D Euclidian space is mapped to a conformal point (null vector) $\boldsymbol{X}$, with $\boldsymbol{X}^2 = 0$, by $\boldsymbol{X} = \boldsymbol{x} + \frac{1}{2}\boldsymbol{x}^2\mathbf{e}_\infty + \mathbf{e}_o$. A geometric object, say $\boldsymbol{O}$, can now be defined in terms of its inner product null space $\mathbb{X}(\boldsymbol{O}) = \{\boldsymbol{X} \in \mathbb{R}^{4,1} | \boldsymbol{X} \cdot \boldsymbol{O} = 0\}$. This implies an invariance to scalar multiples $\mathbb{X}(\boldsymbol{O}) = \mathbb{X}(\lambda\, \boldsymbol{O})$, $\lambda \in \mathbb{R}\backslash\{0\}$. Verify that $\boldsymbol{S} = \boldsymbol{s} + \frac{1}{2}(\boldsymbol{s}^2 - r^2)\mathbf{e}_\infty + \mathbf{e}_o$ represents a 2-sphere centered at $\boldsymbol{s} \in \mathbb{R}^3$ with radius r. Similarly, if $\boldsymbol{n} \in \mathbb{R}^3$, with $\boldsymbol{n}^2 = 1$, then $\boldsymbol{P} = \boldsymbol{n} + d\,\mathbf{e}_\infty$ represents a plane at a distance d from the origin with orientation $\boldsymbol{n}$. Using this vector valued geometric primitives higher order entities can be build: given two objects, e.g. the planes $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$, their line of intersection $\boldsymbol{L}$ is simply the outer product $\boldsymbol{L} = \boldsymbol{P}_1 \wedge \boldsymbol{P}_2$. An important GA operation is the reflection. The reflection of $\boldsymbol{A}$ in $\boldsymbol{O}$ is given by the sandwiching product $\boldsymbol{B} = \boldsymbol{OAO}$. The most general case is the reflection in a sphere, which actually represents an inversion. Note that any RBM can be represented by consecutive reflections in planes. In CGA the resulting elements, for example $\boldsymbol{M} = \boldsymbol{P}_1\boldsymbol{P}_2$, are referred to as motors, cf. [5]. Some object $\boldsymbol{C}$ would then be subjected to $\boldsymbol{MC}\widetilde{\boldsymbol{M}}$, whereby the symbol '$\widetilde{\ }$' stands for an order reversion, i.e. the reverse of $\boldsymbol{M}$ is $\widetilde{\boldsymbol{M}} = \boldsymbol{P}_2\boldsymbol{P}_1$. Since reflection is an involution, a double reflection $\boldsymbol{O}(\boldsymbol{OAO})\boldsymbol{O}$ must be the identity w.r.t $\mathbb{X}(\boldsymbol{A})$, therefore $\boldsymbol{O}^2 \in \mathbb{R}$ by associativity. It looks like a conclusion, but in GA every vector valued element $\boldsymbol{X} \in \mathbb{R}^{4,1}$ squares to a scalar by definition $\boldsymbol{X}^2 := \boldsymbol{X} \cdot \boldsymbol{X} \in \mathbb{R}$. Using the above definitions, we have $\boldsymbol{S}^2 = r^2$ and $\boldsymbol{P}^2 = 1$.

THALES' THEOREM REVISITED

In this section we demonstrate how a simple geometric theorem motivates a solution to P3P. Our considerations refer to the left and middle part of figure 1.

The generalization of Thales' theorem states that, given a circle $\boldsymbol{K}$, the centric angle $\angle(\boldsymbol{x}_1', \boldsymbol{m}, \boldsymbol{x}_1)$ at $\boldsymbol{m}$ is twice the peripheral angle $\angle(\boldsymbol{x}_1', \boldsymbol{O}, \boldsymbol{x}_1)$ at $\boldsymbol{O}$. We use this fact and define two successive rotations: the first rotates $\boldsymbol{x}_1$ to $\boldsymbol{x}_1'$ and the second rotates $\boldsymbol{x}_1'$ back onto the straight line connecting $\boldsymbol{O}$ and $\boldsymbol{x}_1$. We obtain point $\boldsymbol{x}_1''$. It is crucial, that any second point $\boldsymbol{x}_2$ on $\boldsymbol{K}$ will also move towards $\boldsymbol{O}$. Moreover, the distance from $\boldsymbol{x}_1$ to $\boldsymbol{x}_2$ stays constant since rotations are distance preserving.

Before we enlighten the value of this observation we work out the general transformation, that we denote $\boldsymbol{R}_\theta$, in terms of CGA. We therefore replace each rotation with two reflections in suitable planes. We have to take care that the dihedral angle between the planes equals half of the rotation angle. Further, the planes' line of intersection must coincide with the rotation axis. The two rotations can ultimately be realized by four reflections in the three planes $\boldsymbol{P}_1$, $\boldsymbol{P}_2$ and $\boldsymbol{P}_3$. The order of application must be $\boldsymbol{P}_1$, $\boldsymbol{P}_2$, $\boldsymbol{P}_3$ and $\boldsymbol{P}_1$ again. We obtain the motor $\boldsymbol{R}_\theta = \boldsymbol{P}_1\boldsymbol{P}_3\boldsymbol{P}_2\boldsymbol{P}_1$. For the derivation we take a canonical coordinate system as a basis. The whole setup is depicted in figure 1. We define $\boldsymbol{P}_1 = \mathbf{e}_1$, $\boldsymbol{P}_2 = \cos(\theta)\,\mathbf{e}_1 + \sin(\theta)\,(\mathbf{e}_2 + r\,\mathbf{e}_\infty)$ and $\boldsymbol{P}_3 = \cos(\theta/2)\,\mathbf{e}_1 + \sin(\theta/2)\,\mathbf{e}_2$, whereby r denotes the radius of circle $\boldsymbol{K}$. After some algebra we obtain

$$\boldsymbol{R}_\theta = \cos(\theta/2) + \sin(\theta/2)\Big[\underbrace{\mathbf{e}_1\mathbf{e}_2 + r\Big((\cos(\theta)+1)\mathbf{e}_1\mathbf{e}_\infty + \sin(\theta)\mathbf{e}_2\mathbf{e}_\infty\Big)}_{\boldsymbol{L}_\theta}\Big] = \exp\Big(\theta/2\,\boldsymbol{L}_\theta\Big). \tag{1}$$

The element $\boldsymbol{L}_\theta$ is a line representing the rotation axis of $\boldsymbol{R}_\theta$ and plays the role of the imaginary unit i of complex numbers, for $\boldsymbol{L}_\theta^2 = -1$. By noting that $\boldsymbol{R}_\theta = \boldsymbol{P}_1(\boldsymbol{P}_3\boldsymbol{P}_2)\boldsymbol{P}_1$ it can be recognized that $\boldsymbol{R}_\theta$ is the reflection of $\boldsymbol{R}_\theta' := \boldsymbol{P}_3\boldsymbol{P}_2$ in plane $\boldsymbol{P}_1$. Hence $\boldsymbol{R}_\theta$ rotates by an angle θ being twice the dihedral angle between $\boldsymbol{P}_3$ and $\boldsymbol{P}_2$.

Regarding pose estimation this result can be interpreted as a way to rigidly move two 3D model points on their respective projection rays. The latter can be computed from the internal calibration and the corresponding image points. The point $\boldsymbol{O}$ represents the optical center of the camera. Unfortunately, it is not sufficient to consider two points, but the above result leads to a 3-point approach.

THE PERSPECTIVE 3-POINT PROBLEM

In this section, we concentrate on the subproblem of determining the pose of point triplets. We thus consider three 3D model points $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3\}$ and their corresponding image points. From these we compute the respective projection rays

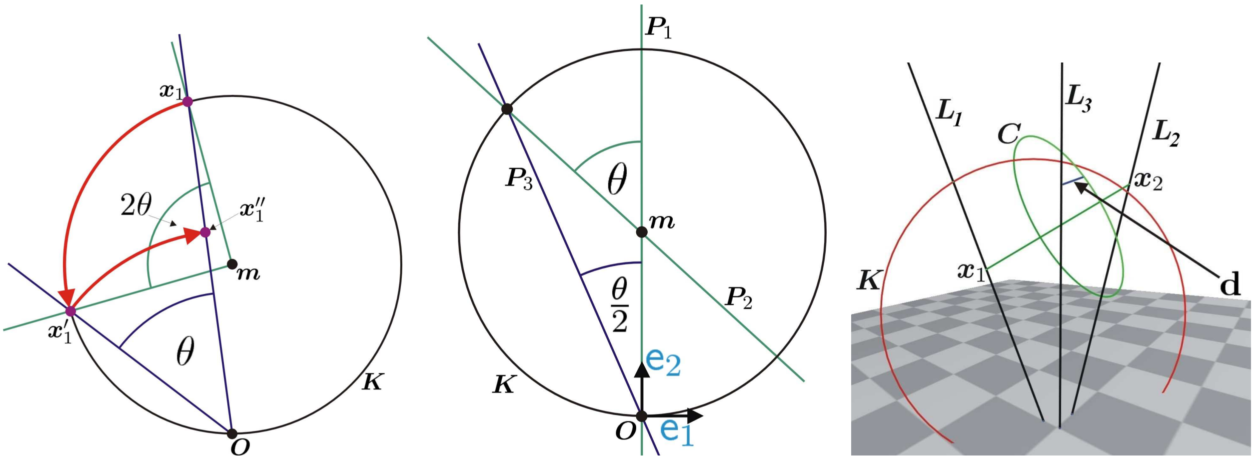

FIGURE 1. Left: the generalization of Thales' theorem explains why two successive rotations form a translation. **Middle:** the transformation can be realized by a sequence of four reflections in three well-chosen planes $\boldsymbol{P}_1$, $\boldsymbol{P}_2$ and $\boldsymbol{P}_3$. The overall transformation is then $\boldsymbol{R}_\theta = \boldsymbol{P}_1\boldsymbol{P}_3\boldsymbol{P}_2\boldsymbol{P}_1$. **Right:** Distance **d** between circle $\boldsymbol{C}_\theta$ and the third projection ray $\boldsymbol{L}_3$.

$\{\boldsymbol{L}_1, \boldsymbol{L}_2, \boldsymbol{L}_3\}$.

The aim is to find a position and orientation for the model such that each model point coincides with its respective projection ray. This problem is referred to as fitting. It enables the computation of an RBM that interrelates the external (world) coordinate system (WCS) of the 3D model and the camera coordinate system (CCS). Note that the RMB, which we encode in a motor, constitutes the camera pose.

The solution arises from the question where the third point $\boldsymbol{x}_3$ might be while we use $\boldsymbol{R}_\theta$ to move $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ along their projection rays. Clearly the mutual distances between the three points must me retained. The locus of the third point must therefore be a circle $\boldsymbol{C}$ which has to be subjected to $\boldsymbol{R}_\theta$ along with $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, see figure 1. Note[1] that we have reached our aim if $\boldsymbol{C}_\theta = \boldsymbol{R}_\theta \boldsymbol{C} \widetilde{\boldsymbol{R}_\theta}$ intersects ray $\boldsymbol{L}_3$. The CGA expression $\boldsymbol{C}_\theta \wedge \boldsymbol{L}$ represents a sphere which degenerates to the point of intersection in case $\boldsymbol{L}_3$ hits $\boldsymbol{C}_\theta$. Hence a suitable function is $h(\theta) = (\boldsymbol{C}_\theta \wedge \boldsymbol{L})^2 \in \mathbb{R}$ which has to attain zero.

Since the derivatives of $h(\theta)$ are analytically available, we employ the Newton-Raphson method for the computation of the at most four roots $\Theta = \{\theta_1, \theta_2, \ldots, \theta_k\}$, $k \leq 4$. The application of the respective motor $\boldsymbol{R}_{\theta\in\Theta}$ to the points $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3\}$ yields the fitted points $\{\boldsymbol{x}_1', \boldsymbol{x}_2', \boldsymbol{x}_3'\}$, being the connection between CCS and WCS. For each angle $\theta \in \Theta$ we estimate the interrelating motor $\boldsymbol{M}_\theta$ by means of standard methods, see [10]. The application of the $\boldsymbol{M}_{\theta\in\Theta}$ to the entire n-point problem assesses their quality and ultimately reveals inappropriate pose candidates.

It remains to specify a selection strategy for triplets in order to put a limit on the computational complexity $\binom{n}{3}$: we select those triplets the image points of which form maximum area triangles on the image plane. In this way the impact of noise in the coordinates of the image points is minimized, i.e. the fitting is more constrained as the noise induced jittering of the model does not carry much weight.

THE PERSPECTIVE n-POINT PROBLEM

The issue regarding the fusion of the P3P motors $\{\boldsymbol{M}_1, \boldsymbol{M}_2, \ldots, \boldsymbol{M}_N\}$ is the second key aspect of this work. Since any motor is from the Lie group $SE(3)$, being a manifold, the customary arithmetic mean may not be used: $A, B \in SE(3) \not\Rightarrow A + B \in SE(3)$. The Lie group $SE(3)$ is connected to its Lie algebra $se(3)$ (tangent space to the identity element of $SE(3)$) by the `exp`/`log` map. Note that in $se(3)$ any customary mean can be built as the algebra elements form a vector space. This is exploited by the 'weighted intrinsic mean', in which the `log` map is used to compute first-order mean approximations via the tangent space, see [11, 12]. The N motors are input to the outlined algorithm below, whereby

[1] The three original model points were initially subjected to an RBM such that $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ of the resulting points $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3\}$ already lie on their corresponding projection rays. We then use $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3\}$ to determine $\boldsymbol{K}$ and $\boldsymbol{C} = \boldsymbol{C}_0$.

the weights w_i, $1 \leq i \leq N$, reflect the motor assessments. Starting from the motor $\boldsymbol{M} = \texttt{identity}$ the subsequent three steps are repeated until $||\texttt{log}(\Delta\boldsymbol{M})||$ falls below a certain threshold ε.

$$\begin{array}{lrcll} 1. & \Delta\boldsymbol{A}_i & = & \log\left(\boldsymbol{M}^{-1}\boldsymbol{M}_i\right) & 1 \leq i \leq N \\ 2. & \Delta\boldsymbol{M} & = & \exp\left(\frac{1}{W}\sum_{i=1}^{N} w_i\,\Delta\boldsymbol{A}_i\right) & W = \sum_{i=1}^{N} w_i \\ 3. & \boldsymbol{M} & = & \boldsymbol{M}\,\Delta\boldsymbol{M} & \end{array} \quad (2)$$

Notice that the motor $\boldsymbol{M}$ is repeatedly updated by the residuals $\Delta\boldsymbol{M}$, which originate from the weighted averaging of algebra elements $\Delta\boldsymbol{A}_i$, $1 \leq i \leq N$. The term $\boldsymbol{M}^{-1}\boldsymbol{M}_i$ in step 1 moves the input closer to the identity element of $SE(3)$ in order to minimize the averaging error in step 2. A derivation of the algorithm and a uniqueness proof is given in [13].

Here we do not report experimental results but state that our pose estimation has already been vastly used. Besides, in synthetic experiments the method has proven to produce results competitive to the ground truth or to a stochastically optimal solution.

CONCLUSION

Starting from a clear geometric concept, we have proposed a solution to the perspective *n*-point problem which is free from initialization. Our method is consistent since it always yields the correct solution in the absence of noise. The involved 3-point problem has been streamlined to a scalar valued function solely depending on an angle. Our approach enables the rejection of solution candidates which stem from false correspondences. We have introduced a selection strategy for point triplets which reduces the computational effort and improves the accuracy at the same time. In the final weighted averaging of rigid body motions we have taken their algebraic nature into account using the intrinsic mean. Evidently, this work is also a valuable contribution to the field of (conformal) geometric algebra, which turned out to be the ideal framework to deal with geometric objects and transformations.

REFERENCES

1. R. M. Haralick, C.-N. Lee, K. Ottenberg, and M. Noelle, “Review and Analysis of Solutions of the Three Point Perspective Pose Estimation Problem,” in *International Journal on Computer Vision*, 1994, vol. 13, pp. 331–356.
2. M. A. Fischler, and R. C. Bolles, “Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography,” in *Commun. ACM*, ACM Press, New York, NY, USA, 1981, vol. 24, pp. 381–395.
3. L. Quan, and Z.-D. Lan, “Linear N-Point Camera Pose Determination,” in *PAMI*, 1999, vol. 21, pp. 774–780.
4. B. K. P. Horn, Projective Geometry Considered Harmful (1999), available at http://www.ai.mit.edu/people/bkph/papers/harumful.pdf.
5. B. Rosenhahn, and G. Sommer, “Pose estimation in conformal geometric algebra, Part I: the stratification of mathematical spaces,” in *Journal of Mathematical Imaging and Vision*, Springer Science + Business Media, Inc., 2005, vol. 22, pp. 27–48.
6. D. Hestenes, and G. Sobczyk, *Clifford Algebra to Geometric Calculus: A Unified Language for Mathematics and Physics*, Reidel, Dordrecht, 1984.
7. L. Dorst, D. Fontijne, and S. Mann, *Geometric Algebra for Computer Science: An Object-Oriented Approach to Geometry*, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2007.
8. P. Angles, “Construction de revêtements du groupe conforme d’un espace vectoriel muni d’une ≪métrique≫ de type (p,q),” in *Annales de l’institut Henri Poincaré*, 1980, vol. 33, pp. 33–51.
9. H. Li, D. Hestenes, and A. Rockwood, “Generalized homogeneous coordinates for computational geometry,” in *Geometric computing with Clifford algebras: theoretical foundations and applications in computer vision and robotics*, Springer-Verlag, London, UK, 2001, pp. 27–59.
10. K. S. Arun, T. S. Huang, and S. D. Blostein, “Least-Squares Fitting of Two 3-D Point Sets,” in *PAMI*, 1987, vol. 9, pp. 698–700.
11. P. T. Fletcher, C. Lu, and S. C. Joshi, “Statistics of Shape via Principal Geodesic Analysis on Lie Groups,” in *CVPR*, 2003, pp. 95–101.
12. S. Buchholz, and G. Sommer, “On Averaging in Clifford Groups,” in *Computer Algebra and Geometric Algebra with Applications*, edited by H. Li, P. Olver, and G. Sommer, 6th International Workshop IWMM 2004, Springer, 2005, vol. 3519 of *LNCS*, pp. 229–238.
13. S. R. Buss, and J. P. Fillmore, “Spherical averages and applications to spherical splines and interpolation,” in *Commun. ACM*, ACM Press, New York, NY, USA, 2001, vol. 20, pp. 95–126.