Nonlinear Least Squares Optimisation of Unit Quaternion Functions for Pose Estimation from Corresponding Features

Aleš Ude*

Jožef Stefan Institute, Department of Automatics, Biocybernetics and Robotics
Jamova 39, 1000 Ljubljana, Slovenia, E-mail: ales.ude@ijs.si

Abstract

Pose estimation from an arbitrary number of 2-D to 3-D feature correspondences is often done by minimising a nonlinear criterion function using one of the minimal representations for the orientation. However, there are many advantages in using unit quaternions to represent the orientation. Unfortunately, a straightforward formulation of the pose estimation problem based on quaternions results in a constrained optimisation problem. In this paper we propose a new method for solving general nonlinear least squares optimisation problems involving unit quaternion functions based on unconstrained optimisation techniques. We demonstrate the effectiveness of our approach for pose estimation from 2-D to 3-D line segment correspondences.

1. Introduction

The object pose is defined as the displacement of the co-ordinate frame rigidly attached to the object from its initial position, where it is aligned with the world coordinate frame, to its current position. There exist analytical and linear solutions to the problem of pose estimation from 2-D to 3-D feature correspondences [1, 2], but they are sensitive to noise. In the presence of noise, which is unavoidable in real-world applications, algorithms based on nonlinear optimisation methods give more accurate results.

Nonlinear optimisation techniques have been used for pose estimation by many researchers in the past. A good overview is given in [1]. In most of these approaches, Euler's angles were used to parameterise the group of rotations $\mathrm{SO}(3)$ of the Euclidean space. However, it is well known that $\mathrm{SO}(3)$, which is a three dimensional manifold, cannot be globally embedded in the three dimensional Euclidean space. It follows that if the rotation group is represented by three real parameters, the Euclidean metric topology in

*Currently, the author is with the Kawato Dynamic Brain Project, ERATO, Japan Science and Technology Corporation, 2-2 Hikaridai Seika-cho, Soraku-gun, Kyoto 619-0288, Japan, e-mail: ude@erato.atr.co.jp.

$\mathbb{R}^3$ does not induce a global topology and metric structure in $\mathrm{SO}(3)$. This suggests that common solutions using minimal representations of the rotation group are not ideal.

The representation of the rotation group by *unit quaternions,* which form a sphere S^3 in $\mathbb{R}^4$, has many advantages over minimal representations. Methods for pose estimation based on the quaternion representation of the orientation have been proposed in the literature before [1], but the pose estimation problem has been formulated as an optimisation problem in $\mathbb{R}^4$ rather than on S^3 in these approaches.

2. Preliminaries

In the following we shall need the exponential map $\exp : \mathbb{R}^3 \to \mathrm{S}^3$, which is given by

$$\exp(\boldsymbol{r}) = \begin{cases} \left(\cos(\|\boldsymbol{r}\|), \sin(\|\boldsymbol{r}\|)\dfrac{\boldsymbol{r}}{\|\boldsymbol{r}\|} \right), \ \boldsymbol{r} \neq 0 \\ (1, 0, 0, 0), \ \boldsymbol{r} = 0 \end{cases}. \quad (1)$$

The exponential map transforms a tangent vector $\boldsymbol{r} \in \mathbb{R}^3 \equiv \mathrm{T_1}(\mathrm{S}^3)$ into $\boldsymbol{q} \in \mathrm{S}^3$, where $\boldsymbol{q}$ is a point at distance $\|\boldsymbol{r}\|$ from $\mathbf{1}$ along a geodesic curve starting from $\mathbf{1}$ in the direction of $\boldsymbol{r}$ [3]. Geodesics are defined as shortest paths connecting any two points on a manifold (sphere S^3). It turns out that for any other point $\boldsymbol{q} \in \mathrm{S}^3$ and for any $\boldsymbol{r_q} \in \mathrm{T_q}(\mathrm{S}^3)$, $\boldsymbol{r_q} * \overline{\boldsymbol{q}} \in \mathrm{T_1}(\mathrm{S}^3) \equiv \mathbb{R}^3$ and the exponential map at $\boldsymbol{q}$, $\exp_{\boldsymbol{q}} : \mathrm{T_q}(\mathrm{S}^3) \to \mathrm{S}^3$, having the above properties is given by

$$\exp_{\boldsymbol{q}}(\boldsymbol{r_q}) = \exp(\boldsymbol{r_q} * \overline{\boldsymbol{q}}) * \boldsymbol{q}, \quad (2)$$

where $*$ denotes the quaternion multiplication.

Lets consider the problem of pose estimation from 2-D to 3-D line segment correspondences. Let $\boldsymbol{m}_k = (\boldsymbol{x}_k^1, \boldsymbol{x}_k^2)$, $k = 1, \ldots, N$, be the end-points of the k-th 3-D line segment belonging to the object's model and let $\boldsymbol{A}_j$, $j = 1, \ldots, M$, be the projective mapping onto the j-th image plane. These mappings should be made available by a camera calibration procedure. Let $\boldsymbol{f}$ denote the mapping which transforms the end-point representation of a 2-D line segment into its mid-point representation

$$\boldsymbol{f}(\boldsymbol{v}_1, \boldsymbol{v}_2) = \left[\frac{\boldsymbol{v}_1^T + \boldsymbol{v}_2^T}{2}, \ \arctan\left(\frac{y_2 - y_1}{x_2 - x_1} \right) \right]^T, \quad (3)$$

where $\boldsymbol{v}_l = [x_l, y_l]^T$. The length of a line segment is excluded from this mapping because it is more noisy than other mid-point parameters. We encode the pose $\boldsymbol{p}$ by a 3-D vector $\boldsymbol{t} \in \mathbb{R}^3$ and by a quaternion $\boldsymbol{q} \in \mathrm{S}^3$. Let $\boldsymbol{g}_{jk} : \mathbb{R}^3 \times \mathrm{S}^3 \to \mathbb{R}^3$ be the following mappings

$$\boldsymbol{g}_{jk}(\boldsymbol{p}) = \boldsymbol{f}(\boldsymbol{A}_j(\boldsymbol{q} * \boldsymbol{x}_k^1 * \overline{\boldsymbol{q}} + \boldsymbol{t}), \boldsymbol{A}_j(\boldsymbol{q} * \boldsymbol{x}_k^2 * \overline{\boldsymbol{q}} + \boldsymbol{t})). \quad (4)$$

Each $\boldsymbol{g}_{jk}$ maps the k-th 3-D model segment onto the 2-D image segment obtained by moving the model segment by $\boldsymbol{p} = (\boldsymbol{q}, \boldsymbol{t})$ and by projecting the resulting segment onto the j-th image plane.

We assume now that the correspondences between the measured image segments $\boldsymbol{y}_{jk} = [\boldsymbol{u}_{jk}^T, \vartheta_{jk}]^T$ and the model segments $\boldsymbol{m}_k$ are given. The optimal estimate for the current object pose can be calculated by minimising the following nonlinear criterion function

$$\frac{1}{2} \sum_{j=1}^{M} \sum_{k=1}^{N} \left(\boldsymbol{y}_{jk} - \boldsymbol{g}_{jk}(\boldsymbol{t}, \boldsymbol{q})\right)^\top \boldsymbol{\Sigma}_{jk}^{-1} \left(\boldsymbol{y}_{jk} - \boldsymbol{g}_{jk}(\boldsymbol{t}, \boldsymbol{q})\right)$$
$$= \frac{1}{2} \sum_{k=1}^{3MN} h_k(\boldsymbol{t}, \boldsymbol{q})^2, \quad (5)$$

where $\boldsymbol{\Sigma}_{jk}$ are the covariances of the measured segments.

The minimisation of (5) over $\boldsymbol{t}$ and $\boldsymbol{q}$ would be a classic nonlinear least squares optimisation problem if we could treat $\boldsymbol{q}$ as an element of $\mathbb{R}^4$ and not of S^3. Since this is not the case, the classic approach would be to add the constraint $|\boldsymbol{q}| = 1$ to the above criterion. In the next section we propose a better solution.

3. Least squares optimisation on unit sphere

Lets consider the minimisation of sum of squares of general unit quaternion functions

$$\min_{\boldsymbol{q} \in \mathrm{S}^3} \left\{ F(\boldsymbol{q}) = \frac{1}{2} \sum_{k=1}^{n} f_k(\boldsymbol{q})^2 = \frac{1}{2} \boldsymbol{f}(\boldsymbol{q})^T \boldsymbol{f}(\boldsymbol{q}) \right\}. \quad (6)$$

Denoting the $n \times 4$ Jacobian matrix of $\boldsymbol{f}(\boldsymbol{q})$ as $\boldsymbol{J}(\boldsymbol{q})$, the gradient $\nabla F(\boldsymbol{q})$ and the Hessian $\nabla^2 F(\boldsymbol{q})$ are given by

$$\nabla F(\boldsymbol{q}) = \boldsymbol{J}(\boldsymbol{q})^T \boldsymbol{f}(\boldsymbol{q}), \quad (7)$$
$$\nabla^2 F(\boldsymbol{q}) = \boldsymbol{J}(\boldsymbol{q})^T \boldsymbol{J}(\boldsymbol{q}) + \sum_{k=1}^{n} f_k(\boldsymbol{q}) \nabla^2 f_k(\boldsymbol{q}). \quad (8)$$

Let $\boldsymbol{q}_i$ be the current approximation for the minimum of (6). The Taylor series expansion for the vector function $\nabla F(\boldsymbol{q})$ around $\nabla F(\boldsymbol{q}_i)$ is given by

$$\nabla F(\boldsymbol{q}) \approx \nabla F(\boldsymbol{q}_i) + \nabla^2 F(\boldsymbol{q}_i)(\boldsymbol{q} - \boldsymbol{q}_i) \quad (9)$$

Using the above expansion and the fact that $\nabla F(\boldsymbol{q}) = 0$ at the minimum of F, the next approximation for the minimum can be calculated as follows

$$\boldsymbol{q}_{i+1} = \boldsymbol{q}_i - (\nabla^2 F(\boldsymbol{q}_i))^{-1} \nabla F(\boldsymbol{q}_i). \quad (10)$$

If we assume that the value of F is small for all $\boldsymbol{q}$ belonging to the neighbourhood of the solution, i.e. $f_k(\boldsymbol{q}) \approx 0$ for all k, we obtain the following approximation for the Hessian in the neighbourhood of the solution based on Eq. (8)

$$\nabla^2 F(\boldsymbol{q}) \approx \boldsymbol{J}(\boldsymbol{q})^T \boldsymbol{J}(\boldsymbol{q}). \quad (11)$$

Writing this approximation for the Hessian in (10), we arrive to the *Gauss-Newton* iteration, which is very effective provided a good starting point is known

$$\boldsymbol{q}_{i+1} = \boldsymbol{q}_i - (\boldsymbol{J}(\boldsymbol{q}_i)^T \boldsymbol{J}(\boldsymbol{q}_i))^{-1} \boldsymbol{J}^T(\boldsymbol{q}_i) \boldsymbol{f}(\boldsymbol{q}_i). \quad (12)$$

Unfortunately, $\boldsymbol{q}_{i+1}$ given by iteration (12) does not necessarily lie on the unit sphere. This is due to the fact that this iteration searches for the minimum of (6) in the $\mathbb{R}^4$-neighbourhood and not in the S^3-neighbourhood of $\boldsymbol{q}_i$. To define an iteration on the unit sphere we observe that the neighbouring points of $\boldsymbol{q}_i$ in S^3 are given by $\exp(\boldsymbol{\omega}) * \boldsymbol{q}_i$, $\boldsymbol{\omega} \in \mathbb{R}^3$. Hence we can write criterion (6) as

$$G_i(\boldsymbol{\omega}) = \frac{1}{2} \sum_{k=1}^{n} f_k(\exp(\boldsymbol{\omega}) * \boldsymbol{q}_i)^2 = \frac{1}{2} \sum_{k=1}^{n} g_k^i(\boldsymbol{\omega})^2$$
$$= \frac{1}{2} \boldsymbol{g}^i(\boldsymbol{\omega})^T \boldsymbol{g}^i(\boldsymbol{\omega}). \quad (13)$$

G_i can be viewed as a mapping from $\mathbb{R}^3$ to $\mathbb{R}$. Eq. (2) guarantees us that in this way we cover the whole tangent space $\mathrm{T}_{\mathbf{q}_i}(\mathrm{S}^3)$ or, in other words, all directions starting from $\boldsymbol{q}_i$. We want to calculate the next approximate $\boldsymbol{q}_{i+1}$ as $\boldsymbol{q}_{i+1} = \exp(\boldsymbol{\omega}_i) * \boldsymbol{q}_i$. Because of the properties of the exponential map, such $\boldsymbol{q}_{i+1}$ lies along the geodesic curve starting at the current approximate $\boldsymbol{q}_i$ in the direction of $\boldsymbol{\omega}_i * \boldsymbol{q}_i$. Since the exponential map preserves distances, the length of the step on the sphere is equal to the norm of $\boldsymbol{\omega}_i$.

To determine $\boldsymbol{\omega}_i$, we take zero as an initial approximation for the minimum of criterion (13). We denote the $n \times 3$ Jacobian matrix of $\boldsymbol{g}^i$ at $\boldsymbol{\omega} = 0$ as $\boldsymbol{J}_i$. In these circumstances, one step of the Gauss-Newton iteration (12) results in the following approximation for the minimum of criterion (13)

$$\boldsymbol{\omega}_i = -(\boldsymbol{J}_i^T \boldsymbol{J}_i)^{-1} \boldsymbol{J}_i^T \boldsymbol{g}^i(0). \quad (14)$$

Thus the Gauss-Newton iteration on the unit sphere can be summarised as follows:

(i) Initialisation:

$$\boldsymbol{q}_0 \leftarrow \text{initial approximation.}$$

(ii) Loop:

$$\boldsymbol{q}_{i+1} = \exp\left(-(\boldsymbol{J}_i^T \boldsymbol{J}_i)^{-1} \boldsymbol{J}_i^T \boldsymbol{g}^i(0)\right) * \boldsymbol{q}_i.$$

(iii) Convergence test:

$$\|\nabla G_i(0)\| = \|\boldsymbol{J}_i^T \boldsymbol{g}^i(0)\| < \varepsilon.$$
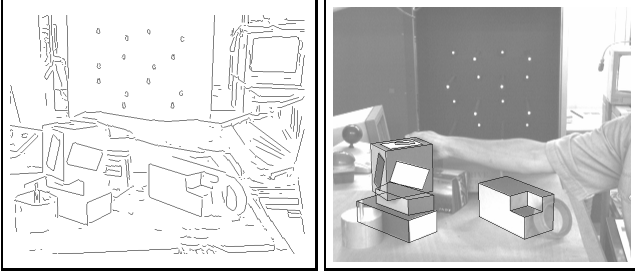
Figure 1. Stereo image pair showing the line segments extracted from one image and the edges of the localised objects projected onto the other image

Our goal is to develop a method for the minimisation of (5) over $(\boldsymbol{t}, \boldsymbol{q}) \in \mathbb{R}^3 \times \mathrm{S}^3$. To achieve this we write (5) as

$$\frac{1}{2} \sum_{k=1}^{3MN} h_k(\boldsymbol{t}_i + \boldsymbol{d}, \exp(\boldsymbol{\omega}) * \boldsymbol{q}_i)^2 = \frac{1}{2} \boldsymbol{g}^i(\boldsymbol{d}, \boldsymbol{\omega})^T \boldsymbol{g}^i(\boldsymbol{d}, \boldsymbol{\omega}).$$
(15)

Using the same approach as in the case of pure unit quaternion functions, the Gauss-Newton iteration on $\mathbb{R}^3 \times \mathrm{S}^3$ can be formulated as follows

$$\begin{aligned}
\boldsymbol{t}_{i+1} &= \boldsymbol{t}_i + \boldsymbol{d}_i, \\
\boldsymbol{q}_{i+1} &= \exp(\boldsymbol{\omega}_i) * \boldsymbol{q}_i, \\
\begin{bmatrix} \boldsymbol{d}_i^T & \boldsymbol{\omega}_i^T \end{bmatrix}^T &= -(\boldsymbol{J}_i^T \boldsymbol{J}_i)^{-1} \boldsymbol{J}_i^T \boldsymbol{g}^i(0,0),
\end{aligned}$$
(16)

where $\boldsymbol{J}_i$ is the Jacobian of $\boldsymbol{g}^i$ at $(0,0)$. This iteration generates a sequence which is guaranteed to lie in the search space $\mathbb{R}^3 \times \mathrm{S}^3$. Since there are no conversions of orientation in some foreign form, such as Euler's angles, to a quaternion form, the non-uniqueness of the quaternion representation does not cause any problems. Since the metric structure of $\mathrm{SO}(3)$ is the same as the one of S^3, the above iteration may be viewed as an iteration in $\mathbb{R}^3 \times \mathrm{SO}(3)$.

4. Experimental results and conclusions

To validate the proposed method experimentally, we used it for the calculation of objects' poses from line segment correspondences in a system for object recognition and localisation (see Fig. 1). The convergence of the method is shown in Tab. 1. Even when the starting point was very inaccurate, the Gauss-Newton method converged to the true object pose provided the feature correspondences were correct. Hence there is no need to use more robust techniques like the Levenberg-Marquardt method for pose verification in calibrated stereo images. Assuming that feature correspondences are correct and the measurement noise is moderate, the criterion function (15) tends to zero in the neighbourhood of the solution and the convergence of the method is nearly as good as the convergence of the Newton-Raphson iteration.

Table 1. Convergence of the method. The Euclidean and the angular metric were used to measure the change in the position and orientation, respectively.

Step (trans.)	Step (orien.)
4.962116e+01	2.827351e-01
1.095884e+01	7.604011e-02
2.851983e-01	2.920441e-03
7.462760e-04	1.835331e-05
2.017427e-06	1.241607e-07

Further experiments should be carried out to test the method for the case when only one camera is available. The presented method is general and works with any kind of features. The usage of line or point correspondences requires only a redefinition of function $\boldsymbol{f}$, which is defined for the case of line segments in Eq. (3).

Comparing our approach with the one of Phong et al. [4], who also used quaternions to represent the orientation, our iteration has the advantage that it searches for the optimal orientation directly in S^3 and not in $\mathbb{R}^4$ as the method of Phong et al. Phong et al. had to introduce a penalty term to force the minimum of their criterion to tend towards S^3. However, this penalty term requires the setting of a user-defined parameter which is at best arbitrary and can cause problems with the convergence of the iteration. Our method does not suffer from this problem. Moreover, it is possible to determine the search direction using the trust region approach of Phong et al. in our iteration and thus make it less sensitive to the quality of a starting point and wrong correspondences. This was not necessary for our application.

References

[1] R. L. Carceroni and C. M. Brown. Numerical methods for model-based pose recovery. Technical Report 659, Computer Science Department, The University of Rochester, Rochester, New York, August 1997.

[2] B. K. P. Horn. Closed-form solution of absolute orientation using unit quaternions. *J. Opt. Soc. Am. A*, 4(4):629–642, 1987.

[3] M.-H. Kyung, M.-S. Kim, and S.-J. Hong. A new approach to through-the-lens camera control. *Graphical Models and Image Processing*, 58(3):262–285, May 1996.

[4] T. Q. Phong, R. Horaud, A. Yassine, and D. T. Pham. Object pose from 2-d to 3-d point and line correspondences. *Intern. J. Comput. Vis.*, 15:225–243, 1995.