Filtering in a unit quaternion space for model-based object tracking

Aleš Ude [1]

Kawato Dynamic Brain Project, Exploratory Research for Advanced Technology, Japan Science and Technology Corporation,
2-2 Hikaridai Seika-cho, Soraku-gun, Kyoto 619-0288, Japan

Abstract

The main idea in object tracking is to support the processing of incoming images by predicting future object's poses and features using the knowledge about the object's previous motion. In this paper we present a new method for the prediction and adjustment of motion parameters to the current measurements using the quaternion representation for the orientation and the Gauss–Newton iteration on the unit sphere S^3. Unlike other trackers, our tracker searches for estimates directly in the space of rigid body motions (represented by $\mathbb{R}^3 \times S^3$) and takes into account the differential and geometric properties of this space. We present some results showing the successful tracking in synthetic and real image sequences. © 1999 Elsevier Science B.V. All rights reserved.

Keywords: Object tracking; Quaternion; Rigid body motion

1. Introduction

Object tracking in long image sequences is an important issue in many robotic applications and has been studied extensively in the past. Most of the proposed approaches consist of the initialisation and the tracking phase. In the initialisation phase, the object of interest must be found and its initial motion parameters must be determined. In this paper we deal only with the tracking phase, in which the information about the object's previous motion is used to calculate the next set of motion parameters. This is usually divided into four steps: prediction, projection, measurement and adjustment. Tracking systems usually use a variant of Kalman filtering to govern this process [3,6,10,11,13].

The kinematics of any point P belonging to a rigid body is completely characterised by $v(t)$, the velocity of the point of the body coinciding with the origin O of the body-fixed coordinate system, and $\omega(t)$, the angular velocity of the point P around the origin of the reference coordinate system

$$v_P(t) = v(t) + \omega(t) \times \overrightarrow{OP}. \tag{1}$$

A closed form solution for this general motion is not known. A common approach in tracking is to assume that the acceleration of the origin of the body coordinate system and the angular acceleration are both random with expected value zero and with constant power spectra a and α, respectively,

$$\mathrm{E}(\dot{v}) = 0, \qquad \mathrm{E}(\dot{v}\dot{v}^T) = a\mathbf{I}, \qquad \mathrm{E}(\dot{\omega}) = 0, \qquad \mathrm{E}(\dot{\omega}\dot{\omega}^T) = \alpha\mathbf{I}, \tag{2}$$

where $\mathbf{I}$ denotes the identity matrix. In this paper we show how to implement the prediction and the adjustment phase based on the above motion model.

The velocity of the origin of the body coordinate system v should not be mixed up with the spatial velocity of a rigid body v^s, which is often used in kinematics and describes the velocity of a (possibly imaginary) point of the body coinciding with the origin of the reference coordinate system. They are related by $v^s = -\omega \times \overrightarrow{OP} + v$. We prefer to use the velocity of the origin of the body coordinate system to parameterise the rigid body motion because the mapping describing the position of a rigid body undergoing motion with constant spatial velocity is nonlinear – it is given by screw actions of a constant twist multiplied by the time parameter [8,13].

2. Parameterisation of the orientation and the exponential map

While the position of an object can be represented by a 3-D vector, this is not the case for the orientation because the set of all orientations does not form a vector space. The set of all orientations, however, is a group and a real three-dimensional manifold [8]. It is well known that there exists no minimal, i.e. three-dimensional, representation of the orientation with the following two properties: (a) the representation contains no singularities, i.e., a continuous motion of the object whose orientation is represented always produces a continuous trajectory in the parameter space, and (b) the partial derivatives of the parameters with respect to any differential rotation, at any orientation, are finite. On the other hand, there are many non-minimal representations which fulfil these two requirements [3]. The most noticeable examples among them are quaternions of unit norm and rotation matrices. Minimal representations comprise Euler's (or RPY) angles [7,10,11] and rotation vectors [13].

Despite all the deficiencies, minimal representations are currently a method of choice rather than an exception in object tracking. The reason for this is that a non-minimal parameter system fulfilling the two conditions above forms a three-dimensional manifold in a higher-dimensional Euclidian space. Such parameters must thus fulfil additional conditions, such as the unity of norm, to be guaranteed to lie on the constraint manifold. This causes problems when an optimisation criterion involving non-minimal parameters representing the orientation should be minimised because optimisation methods are usually based on the linearisation of the criterion function. However, linear corrections applied to such a set of parameters can cause the parameters to depart from the constraint manifold. This difficulty can be resolved by stating the problem as a constrained optimisation problem, but such problems are significantly harder to solve than their unconstrained counterparts. In this paper we propose a better approach based on the properties of the manifold of rigid body motions.

A quaternion $q = (w, u)$ is a four-dimensional vector which can be considered to consist of a real part w and a 3-D vector u. The quaternion multiplication is defined by

$$q * q' = (ww' - u^T u', wu' + w'u + u \times u'). \tag{3}$$

Quaternions form a non-commutative group under the above multiplication. The magnitude of a quaternion is defined as follows

$$|q| = \sqrt{q * \bar{q}} = \sqrt{w^2 + u^T u}, \qquad \bar{q} = (w, -u). \tag{4}$$

A quaternion with a zero vector part is identified with a real number in (4). A vector $v' \in \mathbb{R}^3$ rotated from a vector $v \in \mathbb{R}^3$ by a rotation represented by a unit quaternion q (see Eq. (7)) can be calculated by a simple quaternion multiplication

$$v' = q * v * \bar{q}. \tag{5}$$

In this multiplication the 3-D vectors v and v' are treated as quaternions with a zero scalar component.

The set of unit quaternions can be viewed as a sphere S^3 in $\mathbb{R}^4$. We denote the tangent space of S^3 at q by $T_q(S^3)$. In the following we shall need the exponential map $\exp : T_1(S^3) \equiv \mathbb{R}^3 \to S^3$, which is defined by [4]

$$\exp(r) = \begin{cases} \left(\cos(\|r\|), \sin(\|r\|) \dfrac{r}{\|r\|} \right), & \|r\| \neq 0, \\ 1 = (1, 0, 0, 0), & r = 0. \end{cases} \tag{6}$$

To give a geometric meaning to the exponential map, we make use of geodesic curves which are defined as the shortest paths connecting any two points on the manifold. The exponential map transforms each tangent vector $r \in T_1(S^3)$ into a point $q \in S^3$, where q is the point at distance $\|r\|$ from 1 along the geodesic curve starting from 1 in the direction of r [5]. Such a geodesic curve can be written as $\exp(tr)$. The exponential map can be used to calculate a unit quaternion representing the rotation by ϑ about a unit axis vector n:

$$q(\vartheta, n) = \left(\cos\left(\frac{\vartheta}{2} \right), \sin\left(\frac{\vartheta}{2} \right) n \right) = \exp\left(\frac{\vartheta}{2} n \right). \tag{7}$$

It turns out that for any quaternion $q \in S^3$ and for any tangent vector $r_q \in T_q(S^3)$, we have $r_q * \bar{q} \in T_1(S^3) \equiv \mathbb{R}^3$ and the exponential map at q, $\exp_q : T_q(S^3) \to S^3$, is given by [5]

$$\exp_q(r_q) = \exp(r_q * \bar{q}) * q. \tag{8}$$

Geodesic curves on the manifold of rigid body motions depend on a choice of a metric. The metric giving rise to geodesic curves, which can be expressed by the exponential map (6), does not take into account the dynamic properties of the observed body. Differential equations defining the shortest paths with respect to a metric that does take into account the dynamic properties of a rigid body can be found in [12]. Unfortunately, a closed form solution for these differential equations is not known. The absence of a closed form expression makes this kind of geodesic curves much more difficult to compute and therefore not suitable for our approach to tracking.

The exponential map plays a crucial role in our approach. We relied on the parameterisation of the constraint manifold by unit quaternions to define it. Nevertheless, it is possible to define the exponential map using only the properties of the underlying manifold [8]. Thus the exponential map does not depend on the parameterisation of the manifold. An equivalent approach to tracking could be developed using rotation matrices and the exponential map given by the Rodrigues' formula (see Section 3). We prefer the unit quaternion representation because of its intuitive interpretation as a sphere in $\mathbb{R}^4$.

If we limit the domain of the exponential map to $\|r\| < \pi$, it becomes a one to one mapping whose inverse function, which is the quaternion logarithm, $\log : S^3 / (-1, 0, 0, 0) \to \mathbb{R}^3 \equiv T_1(S^3)$, can be calculated by

$$\log(q) = \begin{cases} \arccos(w) \dfrac{u}{\|u\|}, & u \neq 0 \\ (0, 0, 0), & u = 0 \end{cases} = \begin{cases} \arctan\left(\dfrac{\|u\|}{w} \right) \dfrac{u}{\|u\|}, & w \neq 0, \\ \dfrac{\pi}{2} \dfrac{u}{\|u\|}, & w = 0, \end{cases} \tag{9}$$

where $q = (w, u)$. The second definition is valid only for $\|r\| = \|\log(q)\| \leqslant \pi/2$. The logarithmic map can be used to define the angular metric on S^3

$$d(q_1, q_2) = 2 \begin{cases} \|\log(q_1 * \bar{q}_2)\|, & q_1 * \bar{q}_2 \neq (-1, 0, 0, 0), \\ \pi, & \text{otherwise}, \end{cases} \tag{10}$$

which gives us the length of a geodesic curve connecting the points q_1 and q_2.

3. Prediction of the state vector and error modelling

We turn now to the problem of object tracking. Our aim is to estimate at every measurement time instant t_i the state vector $(p_i, q_i, v_i, \omega_i)$, which consists of the object's pose and velocity, and the covariance matrix S_i of the state vector using the measurements provided by a vision system and the motion model (2). Let us assume that after processing the ith set of measurements, the state vector and its covariance matrix are known. Since the expected values of the acceleration of the origin of the body coordinate system and of the angular acceleration are equal to zero, we can predict the pose and the velocity of the object at the next time instant t_{i+1} as follows:

$$\hat{p}_{i+1} = p_i + (t_{i+1} - t_i)v_i, \tag{11}$$

$$\hat{q}_{i+1} = q_{i,i+1} * q_i, \tag{12}$$

$$\hat{v}_{i+1} = v_i, \tag{13}$$

$$\hat{\omega}_{i+1} = \omega_i, \tag{14}$$

where $q_{i,i+1} = \exp((t_{i+1} - t_i)\omega_i/2)$. Predictions (11)–(14) would be exact if the acceleration was equal to zero on the time interval $[t_i, t_{i+1}]$.

While the differences between the measured and the true values for position, velocity of the body coordinate system and angular velocity can be modelled as additive, i.e.

$$p_i = p_i' + u_i^p, \qquad v_i = v_i' + u_i^v, \qquad \omega_i = \omega_i' + u_i^\omega, \tag{15}$$

where u_i^p, u_i^v, u_i^ω are the corresponding errors, this is not the case for the difference between the measured orientation q_i and the true orientation q_i'. This error can be written as

$$q_i = \exp(u_i^q) * q_i', \tag{16}$$

where $u_i^q \in \mathbb{R}^3$ is the error in the orientation. Error in the orientation is a vector only for differential errors because the equation $\exp(r_1 + r_2) = \exp(r_1) * \exp(r_2)$ is true only for differential values of r_1 and r_2. Note that for a differential rotation error $\mathbf{d}r$ we have $\exp(\mathbf{d}r) \approx (1, \mathbf{d}r)$, thus our model can be viewed as a generalisation of a more common orientation error model which assumes small orientation errors. To find the relationship between the orientation errors at the time instants t_i and t_{i+1}, respectively, we make the following crucial observation

$$q_{i+1} = q_{i,i+1} * q_i = q_{i,i+1} * \exp(u_i^q) * q_i' = q_{i,i+1} * \exp(u_i^q) * \overline{q_{i,i+1}} * q_{i,i+1} * q_i'$$
$$= \exp(q_{i,i+1} * u_i^q * \overline{q_{i,i+1}}) * q_{i+1}'. \tag{17}$$

Hence the error propagated from t_i to t_{i+1} is given by

$$q_{i,i+1} * u_i^q * \overline{q_{i,i+1}} = R(\Delta t_i \omega_i) u_i^q, \tag{18}$$

where $\Delta t_i = t_{i+1} - t_i$ and the rotation matrix $R(\Delta t_i \omega_i)$ is given by the Rodrigues' formula

$$R(\Delta t_i \omega_i) = I + \sin(\Delta t_i \|\omega_i\|) X\left(\frac{\omega_i}{\|\omega_i\|}\right) + (1 - \cos(\Delta t_i \|\omega_i\|))X\left(\frac{\omega_i}{\|\omega_i\|}\right)^2,$$

$$X(r) = X(r_x, r_y, r_z) = \begin{bmatrix} 0 & -r_z & r_y \\ r_z & 0 & -r_x \\ -r_y & r_x & 0 \end{bmatrix}.$$

Errors in the predicted orientation are caused also by errors in the estimated angular velocity. Let ω_t be the error in the angular velocity at the time instant t, $t \in [t_i, t_{i+1}]$. This error has the following effect on the error in the orientation at the time instant t_{i+1} (compare with (18))

$$R((t_{i+1} - t)\omega_i)\omega_t \, dt. \tag{19}$$

Using the fact that differential rotation vectors may be summated and assuming that the overall error is small and that the angular velocity error is constant over the time interval $[t_i, t_{i+1}]$, i.e. $\omega_t = u_i^\omega$, we obtain the following approximation for the error in the orientation caused by the error in the angular velocity

$$\int_{t_i}^{t_{i+1}} R((t_{i+1} - t)\omega_i)u_i^\omega \, dt = \left(\int_0^{\Delta t_i} R(t\omega_i) \, dt \right) u_i^\omega = G_i u_i^\omega. \tag{20}$$

The integration of (20) results in (see also [3])

$$G_i = \Delta t_i \begin{bmatrix} (1-c)n_1^2 + c & (1-c)n_1 n_2 - sn_3 & (1-c)n_1 n_3 + sn_2 \\ (1-c)n_1 n_2 + sn_3 & (1-c)n_2^2 + c & (1-c)n_2 n_3 - sn_1 \\ (1-c)n_1 n_3 - sn_2 & (1-c)n_2 n_3 + sn_1 & (1-c)n_3^2 + c \end{bmatrix}, \tag{21}$$

where

$$n = \frac{\omega_i}{\|\omega_i\|}, \qquad s = \frac{2}{\Delta t_i \|\omega_i\|} \sin^2 \left(\frac{\Delta t_i \|\omega_i\|}{2} \right),$$

$$c = \frac{2}{\Delta t_i \|\omega_i\|} \sin \left(\frac{\Delta t_i \|\omega_i\|}{2} \right) \cos \left(\frac{\Delta t_i \|\omega_i\|}{2} \right).$$

Assuming that errors (18) and (20) are small, they may be considered as physical vectors and the overall orientation error is given as their sum.

The calculation of the combined error in position and in both velocities is much simpler because, unlike the orientation error, these errors are physical vectors. The predicted error in all parameters can be written as

$$\begin{bmatrix} \widehat{u}_{i+1}^p \\ \widehat{u}_{i+1}^q \\ \widehat{u}_{i+1}^v \\ \widehat{u}_{i+1}^\omega \end{bmatrix} = J_i \begin{bmatrix} u_i^p \\ u_i^q \\ u_i^v \\ u_i^\omega \end{bmatrix}, \qquad J_i = \begin{bmatrix} I & 0 & \Delta t_i I & 0 \\ 0 & R(\Delta t_i \omega_i) & 0 & G_i \\ 0 & 0 & I & 0 \\ 0 & 0 & 0 & I \end{bmatrix}. \tag{22}$$

The values of u_i^p, u_i^q, u_i^v, and u_i^ω are of course unknown in practice, but the above transformation is important for the calculation of the covariance matrix of predictions (11)–(14).

In (2) we have introduced the plant noise in form of random acceleration in order to account for deviations of the actual object motion from the assumed motion model. While the expected value of the plant noise is zero and thus does not affect predictions (11)–(14), it does have an impact on the covariance matrix of these predictions. Let S_i be the covariance matrix of the error vector combining the errors in the pose and velocity of the observed object at the time instant t_i and let $\widehat{S}_{i+1}$ denote the covariance matrix of the error in the predicted values (11)–(14). Using the usual covariance propagation rules we obtain the following relationship

$$\widehat{S}_{i+1} = J_i S_i J_i^T + \begin{bmatrix} \frac{1}{3} a \Delta t_i^3 I & 0 & \frac{1}{2} a \Delta t_i^2 I & 0 \\ 0 & \frac{1}{3} \alpha \Delta t_i^3 I & 0 & \frac{1}{2} \alpha \Delta t_i^2 I \\ \frac{1}{2} a \Delta t_i^2 I & 0 & a \Delta t_i I & 0 \\ 0 & \frac{1}{2} \alpha \Delta t_i^2 I & 0 & \alpha \Delta t_i I \end{bmatrix}, \tag{23}$$

where J_i is as in (22). The second term in the above equation describes the effect of the random acceleration on the predicted covariance matrix. Its derivation can be found in [1]. It should be noted at this point that predictions (11)–(14) and (23) are the same as the ones proposed by Gennery [3]. However, our treatment of errors in the orientation differs from Gennery's. This is important for the adjustment of the predicted values to the new measurements, which is presented in the next section.

4. Adjustment of the predicted state vector to the measurements

Lets assume now that the correspondences between the measured image features y_{jk}^{i+1} and the model features m_k, $k = 1, \ldots, N$, at the time instant t_{i+1} are given. In our experiments we used line segments as features and represented them by a reduced mid-point parameter set which consists of the middle point and the angle of a segment. A good estimate for the current state vector, which is consistent with both the measurements and the predicted state vector, can be calculated by minimising the following nonlinear criterion

$$
\frac{1}{2} \sum_{j=1}^{M} \sum_{k=1}^{N} (y_{jk}^{i+1} - g_{jk}(p, q))^{\mathsf{T}} \Sigma_{jk}^{i+1^{-1}} (y_{jk}^{i+1} - g_{jk}(p, q)) + \frac{1}{2} \begin{bmatrix} p - \widehat{p}_{i+1} \\ \log(q * \overline{\widehat{q}_{i+1}}) \\ v - \widehat{v}_{i+1} \\ \omega - \widehat{\omega}_{i+1} \end{bmatrix}^{\mathsf{T}} \widehat{S}_{i+1}^{-1} \begin{bmatrix} p - \widehat{p}_{i+1} \\ \log(q * \overline{\widehat{q}_{i+1}}) \\ v - \widehat{v}_{i+1} \\ \omega - \widehat{\omega}_{i+1} \end{bmatrix},
$$

$$(24)$$

where g_{jk} is the map which projects the kth model feature onto the jth image plane and thus depends among others on the camera calibration. Σ_{jk}^{i+1} are the positive definite covariance matrices of the measured image features at the time instant t_{i+1}. In our experiments we made use of the formula given in [2] to calculate the covariance matrices for the parameters representing the line segments. The summation over j accounts for the fact that more than one camera can be used. Obviously, the criterion (24) can be rewritten as

$$
\frac{1}{2} \sum_{k=1}^{3MN} h_{i+1}^{k}(p, q, v, \omega)^2 = \frac{1}{2} h_{i+1}(p, q, v, \omega)^{\mathsf{T}} h_{i+1}(p, q, v, \omega).
$$

$$(25)$$

The minimisation of (25) over (p, q, v, ω) would be a classic nonlinear least squares optimisation problem if we could treat 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 $|q| = 1$ to the above criterion. In the following we propose a better solution.

To define an iteration on the unit sphere we note that the neighbouring points of a unit quaternion q on the sphere S^3 are given by $\exp(r) * q$, $r \in \mathbb{R}^3$ [5]. Eq. (8) guarantees us that in this way we cover all directions starting from q. Let $(p_{i+1}^{l}, q_{i+1}^{l}, v_{i+1}^{l}, \omega_{i+1}^{l})$ be the approximation for the state vector at the time instant t_{i+1} on the lth step of the iteration. Then we can write the optimisation criterion (24) or, equivalently, (25) as

$$
F_{i+1}^{l}(p, r, v, \omega) = \frac{1}{2} \sum_{k=1}^{3MN} h_{i+1}^{k}(p_{i+1}^{l} + p, \exp(r) * q_{i+1}^{l}, v_{i+1}^{l} + v, \omega_{i+1}^{l} + \omega)^2
$$

$$
= g_{i+1}^{l}(p, r, v, \omega)^{\mathsf{T}} g_{i+1}^{l}(p, r, v, \omega) / 2.
$$

$$(26)$$

F_{i+1}^{l} can be viewed as a map from $\mathbb{R}^{12}$ to $\mathbb{R}$. Taking $(0, 0, 0, 0)$ as a current approximation for the minimum of (26) and denoting the $(3MN + 12) \times 12$ Jacobian matrix of g_{i+1}^{l} at $(0, 0, 0, 0)$ by D_{i+1}^{l}, we employ the Gauss–Newton iteration [9] to calculate a new estimate for the minimum of criterion (26)

$$
[\Delta p_{i+1}^{l\,\mathsf{T}} \Delta r_{i+1}^{l\,\mathsf{T}} \Delta v_{i+1}^{l\,\mathsf{T}} \Delta \omega_{i-1}^{l\,\mathsf{T}}]^{\mathsf{T}} = -(D_{i+1}^{l\,\mathsf{T}} D_{i+1}^{l})^{-1} D_{i+1}^{l\,\mathsf{T}} g_{i+1}^{l}(0, 0, 0, 0).
$$

$$(27)$$

Using these corrections we can calculate the next estimate for the minimum of criterion (25)

$$
\begin{bmatrix} p_{i+1}^{l+1} \\ q_{i+1}^{l+1} \\ v_{i+1}^{l+1} \\ \omega_{i+1}^{l+1} \end{bmatrix} = \begin{bmatrix} p_{i+1}^{l} + \Delta p_{i+1}^{l} \\ \exp(\Delta r_{i+1}^{l}) * q_{i+1}^{l} \\ v_{i+1}^{l} + \Delta v_{i+1}^{l} \\ \omega_{i+1}^{l} + \Delta \omega_{i+1}^{l} \end{bmatrix}.
$$

$$(28)$$

While the corrections in the position and both velocities are linear, this is not the case for the correction in the orientation; the next approximation for the orientation lies along the geodesic curve starting at the current approximate q^l_{i+1}. The direction of this curve and the length of the step are determined by the Gauss–Newton method.

The iteration can be initialised by setting $p^0_{i+1} = \widehat{p}_{i+1}$, $q^0_{i+1} = \widehat{q}_{i+1}$, $v^0_{i+1} = \widehat{v}_{i-1}$, $\omega^0_{i+1} = \widehat{\omega}_{i+1}$. The iteration is stopped when the norm of the gradient is small enough

$$\|\nabla F^l_{i+1}(0, 0, 0, 0)\| = \|D^{l \ \mathrm{T}}_{i+1} g^l_{i+1}(0, 0, 0, 0)\| < \varepsilon. \tag{29}$$

The last estimate is taken to be a new state vector. The covariance matrix of the new state vector is given by

$$S_{i+1} = (D^{l \ \mathrm{T}}_{i+1} D^l_{i-1})^{-1}, \tag{30}$$

where l is the step at which iteration (28) has been stopped.

5. Experimental results and conclusions

We developed a new method for object tracking. Using the properties of the exponential map and the unit quaternion representation of the orientation, we showed how to model the errors in the orientation and how to propagate them between different measurement time instants. Based on these results we formulated the adjustment of the predictions to the measurements as an unconstrained nonlinear least squares optimisation problem in $\mathbb{R}^3 \times S^3 \times \mathbb{R}^3 \times \mathbb{R}^3$. An important contribution of this paper is a new method for the calculation of the adjustment based on the Gauss–Newton iteration in $\mathbb{R}^3 \times S^3 \times \mathbb{R}^3 \times \mathbb{R}^3$. The main advantage of our approach over other approaches, like for instance the one proposed by Gennery [3], who developed a technique similar to ours, is that our method fully takes into account the nonlinearity of the optimisation problem and searches for the optimal position and orientation directly in the space of rigid body motions. Hence no normalisation step to force the resulting parameters representing the orientation towards the constraint manifold is needed with our approach.

To test the performance of the tracker we carried out several experiments with both synthetic and real data. Here we present the performance of the tracker in ideal conditions, i.e. when the motion model (2) is correct. We simulated a camera (only one camera was used in this experiment) with 480 lines by 640 pixels per line, with the image size 6.6 mm by 8.8 mm and with the focal length 10 mm. The imaginary object used in this experiment was a cube with 500 mm long edges, which were used as features. The synthetic images of this object moving along trajectory given by $p(t) = [-300\sqrt{3}, 300\sqrt{3}, 2800] + t 100\sqrt{3}[1, -1, 1]^{\mathrm{T}}$ in millimeters and $q(t) = \exp(t\pi/(12\sqrt{3})[1, 1, -1]^{\mathrm{T}}/2) * (0, 0.8, -0.6, 0)$ were generated for 6 seconds with a frequency of 30 images per second. This process simulated a constant velocity motion with $\|v\| = 300$ mm/s and $\|\omega\| = 15$ deg/s. Additive normally distributed Gaussian noise with standard deviation 7 pixels in the direction of a line segment and 1 pixel in the direction perpendicular to a line segment was added to the two end-points of each line segment. These noisy end-points were then digitised and the resulting line segments were used as measurements. The smoothing parameters a and α from (2) were set to 2.0 and 0.0002, respectively. The filter (state vector and its covariance matrix) was initialised by calculating the pose of the object and its covariance matrix in the first image of a sequence using a pose estimation algorithm and by setting the linear and angular velocity to zero and the corresponding diagonal elements of the state covariance matrix to large values (10^8).

Since the initial values for the velocity were quite far from the true values, the tracker needed some time to adapt the motion parameters to the true values (see Fig. 1). The computed accuracy of the tracking (mean values of the error norm) for the second part of the image sequence $(3 \leqslant t \leqslant 6)$, when the tracker was already well adapted to the underlying object's motion, were 1.65 mm in position, 0.41 degree in orientation, 0.86 mm/s in linear velocity and 0.37 deg/s in angular velocity. The average errors in the predicted values for position and orientation were 1.66 mm and 0.42 degree, respectively. The predicted values were very accurate and differed only little from the adjusted ones. The convergence of the method is shown in Table 1.
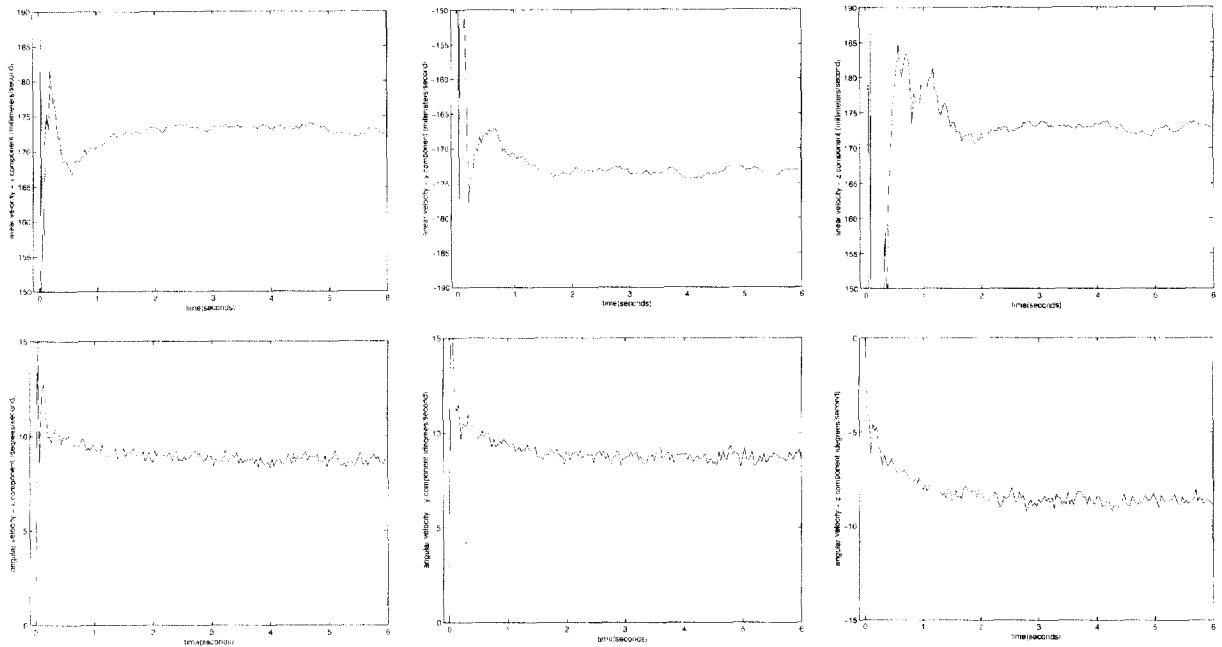
Fig. 1. Linear (first row) and angular (second row) velocity estimated by the proposed filter (simulation results).

Table 1
Convergence of the method (see Eqs. (28) and (29) for definitions of parameters)

$\|\Delta p^j_{i+1}\|$	$2\|\Delta r^j_{i+1}\|$	$\|\Delta v^j_{i+1}\|$	$\|\Delta \omega^j_{i+1}\|$	$\|\nabla F^j_{i+1}(0,0,0,0)\|^2$
				2.658889e+05
6.635401e−01	1.259752e−02	1.610356e+0	1.065318e+0	3.440169e+01
6.678278e−02	8.998415e−05	1.724678e−01	1.687343e−01	3.695909e−04
1.409111e−04	8.215723e−07	3.660243e−04	3.603225e−04	2.299708e−08
1.780821e−06	6.842665e−09	4.644085e−06	4.551428e−06	1.551377e−12
1.389029e−08	5.793137e−11	3.620987e−08	3.551121e−08	1.103704e−16

We used the constant acceleration model for the motion of the origin of the body coordinate system. It is possible to use higher order polynomial models for this motion without making the tracker more complicated. This is not the case for the rotational portion of motion because closed form solutions for higher order rotational motions are not known. The amount of smoothing applied to the parameters describing the motion model is governed by the amount of system noise, which is modelled by parameters a and α from (2), and by the amount of measurement noise, which is modelled by covariance matrices Σ^{i+1}_{jk} from (24) (see [2] for the case of line segments). The smaller the system noise is, the more weight the predictions (11)–(14) get and the larger the amount of smoothing is. The converse relationship holds for the measurement noise. The parameters specifying the amount of system and measurement noise must be determined experimentally in the current version of the tracker.

The presented method was employed successfully for the analysis of human demonstrations recorded by a stereo camera. A linear procedure was used for the calibration of both cameras. A special real-time image processing board was utilised to extract line segments. The resulting edges were sent to a workstation for further processing. Object models were made available by a CAD system. Using the predicted state vector, the model edges were matched
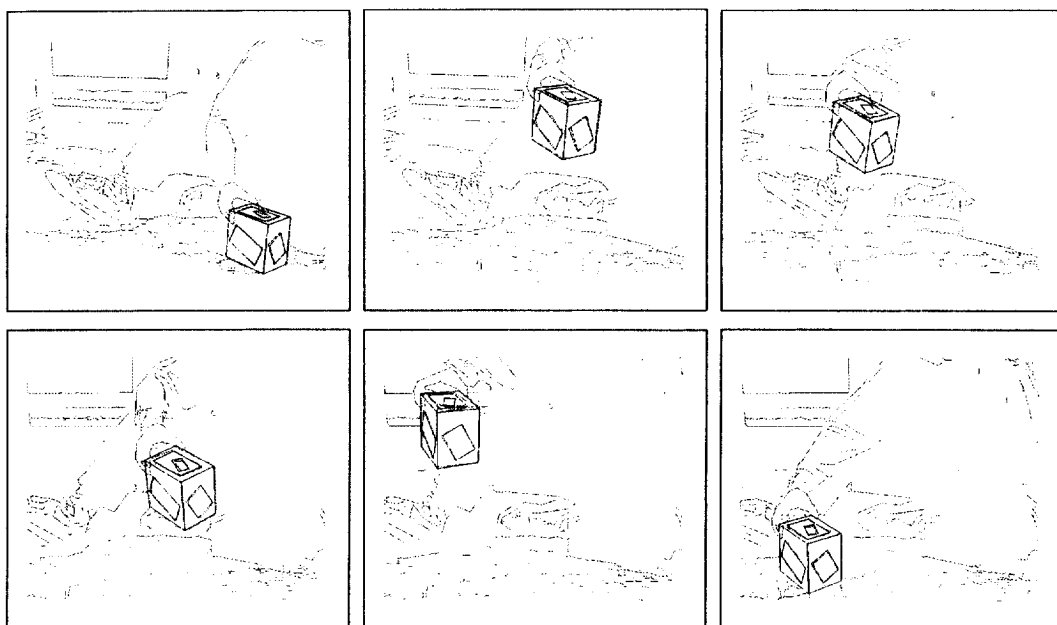
Fig. 2. Six images from a stereo image sequence consisting of 250 image pairs taken at different time instants. There are three images taken with the same camera in each row. The edges of the tracked object were projected from the estimated poses onto the detected edges.

with the detected edges at each measurement time instant. Some of the images recorded in one of our experiments and the projections of the tracked object from the calculated poses are shown in Fig. 2. Even with a fairly bad initial approximation, the method still converged. This shows that there is no need to use more globally convergent techniques like the Levenberg–Marquardt method to solve the problem of model-based object tracking. However, the presented iteration can be easily modified into a Levenberg–Marquardt type iteration if necessary. This might be necessary in the presence of significant measurement errors because the Gauss–Newton iteration often does not work well when the criterion function does not tend to zero. We observed in our simulations that the convergence of the method deteriorates when the amount of measurement noise increases.

Acknowledgements

Appendix A. Calculation of the Jacobian

The Jacobian matrix needed to calculate the adjustment (27) can be calculated by a successive application of the chain rule: $d(F \circ G)_{|x} = dF_{|G(x)} \cdot dG_{|x}$. Problematic is the calculation of $d \exp$ at $r = 0$ and $d \log$ at $q = 1$ because of the singularities in the analytical forms describing these maps. The calculation of $d \exp$ is described in [4]. The Jacobian of the logarithm is given by

$$\mathrm{d}\log_{|q=(w,u)} = \begin{bmatrix} -x & \dfrac{x^2}{\|u\|^2}(w-a)+a & \dfrac{xy}{\|u\|^2}(w-a) & \dfrac{xz}{\|u\|^2}(w-a) \\[3mm] -y & \dfrac{xy}{\|u\|^2}(w-a) & \dfrac{y^2}{\|u\|^2}(w-a)+a & \dfrac{yz}{\|u\|^2}(w-a) \\[3mm] -z & \dfrac{xz}{\|u\|^2}(w-a) & \dfrac{yz}{\|u\|^2}(w-a) & \dfrac{z^2}{\|u\|^2}(w-a)+a \end{bmatrix}, \tag{A.1}$$

where $u = (x, y, z)$ and

$$a = \frac{1}{\|u\|}\arctan\left(\frac{\|u\|}{w}\right).$$

The application of the l'Hopital's rule yields the Jacobian of the logarithm at $q = 1 = (1, 0, 0, 0)$

$$\mathrm{d}\log_{|(1,0,0,0)} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{A.2}$$

References

[1] Y. Bar-Shalom, T.E. Fortmann, Tracking and Data Association, Academic Press, Orlando, FL, 1988.

[2] R. Deriche, O. Faugeras, Tracking line segments, in: O. Faugeras (Ed.), Computer Vision – ECCV'90: First European Conference on Computer Vision, Antibes, France, Lecture Notes in Computer Science, Vol. 427, Springer, Berlin, 1990, pp. 259–268.

[3] D.B. Gennery, Visual tracking of known three-dimensional objects, Intern. J. Comput. Vis. 7 (3) (1992) 243–270.

[4] M.-J. Kim, M.-S. Kim, S.Y. Shin, A compact differential formula for the first derivative of a unit quaternion curve, J. Vis. Comput. Animat. 7 (1) (1996) 43–57.

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

[6] S. Lee, Y. Kay, A Kalman filter approach for accurate 3-D motion estimation from a sequence of stereo images, CVGIP: Image Understanding 54 (2) (1991) 244–258.

[7] D.G. Lowe, Robust model-based motion tracking through the integration of search and estimation, Internat. J. Comput. Vis. 8 (2) (1992) 113–122.

[8] R.M. Murray, Z. Li, S.S. Sastry, A Mathematical Introduction to Robotic Manipulation, CRC Press, Boca Raton, LA, 1994.

[9] L.E. Scales, Introduction to Non-Linear Optimization, Macmillan, London, 1985.

[10] J. Weng, P. Cohen, N. Rebibo, Motion and structure estimation from stereo image sequences, IEEE Trans. Robotics Automat. 8 (3) (1992) 362–382.

[11] J.J. Wu, R.E. Rink, T.M. Caelli, V.G. Gourishankar, Recovery of the 3-D location and motion of a rigid object through camera image (an extended Kalman filter approach), Internat. J. Comput. Vis. 3 (1988) 373–394.

[12] M. Žefran, V. Kumar, C.B. Croke, On the generation of smooth three-dimensional rigid body motions, IEEE Trans. Robotics Automat. 14 (4) (1998) 576–589.

[13] Z. Zhang, O. Faugeras, 3D Dynamic Scene Analysis, Springer, Berlin, 1992.

Aleš Ude studied applied mathematics at the University of Ljubljana, Slovenia. In 1990 he joined the Department of Automatics, Biocybernetics and Robotics, Jožef Stefan Institute, Ljubljana. From 1991 until 1996 he was with the Institute for Process Control and Robotics, University of Karlsruhe, Germany, where he received a doctoral degree from the Department of Computer Science. Since March 1998 he is a postdoctoral fellow in the Kawato Dynamic Brain Project, Exploratory Research for Advanced Technology, Japan Science and Technology Corporation, Kyoto. His main research interests are in various fields of computer vision and intelligent robot systems.