INSTANTANEOUS POSE ESTIMATION USING ROTATION VECTORS

Frederik Beutler, Marco F. Huber, Uwe D. Hanebeck

Intelligent Sensor-Actuator-Systems Laboratory, Institute of Computer Science and Engineering,
Universität Karlsruhe (TH), Germany
`{beutler|marco.huber|uwe.hanebeck}@ieee.org`

ABSTRACT

An algorithm for estimating the pose, i.e., translation and rotation, of an extended target object is introduced. Compared to conventional methods, where pose estimation is performed on the basis of time-of-flight (TOF) measurements between external sources and sensors attached to the object, the proposed approach directly uses the amplitude values measured at the sensors for estimation purposes without an intermediate TOF estimation step. This is achieved by modeling the wave propagation by a nonlinear dynamic system comprising a system and a measurement equation. The nonlinear system equation includes a model of the time-variant structure of the object rotation based on rotation vectors. As a result, the measured amplitude values at the sensors can be processed instantaneously in a recursive fashion. Uncertainties in the measurement process are systematically considered by employing a stochastic filter for estimating the pose, i.e., the state of the nonlinear dynamic system.

Index Terms— Acoustic signal processing, acoustic tracking, inertial navigation, model-based signal processing, state estimation

1. INTRODUCTION

This paper is concerned with estimating the pose, i.e., translation and rotation, of a target object with respect to an external coordinate system. This kind of problem typically arises in indoor tracking systems, where sources placed in the environment emit signals that are picked up by sensors attached to the target [1]. Based on the emitted and the received signals, ranges are calculated that are used in a second step to estimate the translation and the rotation. Translation and rotation describe the relationship between the static global coordinate system and the coordinate system of the moving target object. Indoor tracking systems can be used for example for operator tracking in a telepresence scenario [2]. Here, the position and orientation of a head-mounted display is estimated, that is attached to the operator's head.

The problem of localization or tracking is often approached by a multi-step procedure. Conventional approaches first calculate the range between sources and sensors by means of the well-known matched filter where the data is processed block-wise. By doing so, it is implicitly assumed that the TOF is constant over the data window and thus, a smearing effect of the estimate occurs if the object is moving. An extended matched filter, which is more robust against reflections, can be found in [3]. The ranges obtained can then be used to estimate the pose of the object by employing a Kalman filter or closed-from solutions [4].

A typical application for those algorithms is described in [1]. A tracking system named WHISPER is introduced to track the hands of a user. Thanks to wide-band signals, the system is much more robust to shadowing effects. The algorithm utilizes these signals to estimate the ranges by means of a matched filter. In a second step, these ranges are used in a Kalman filter to estimate the position of the hands. Based on the resulting estimate, range bounds are calculated to minimize the search space for the matched filter. Again, the considered problem is split into two parts and uncertainties are not considered systematically.

The approach proposed in this paper comprises two main contributions. The first contribution is to model the wave propagation by a nonlinear dynamic system (see Section 2 and Section 3, respectively), where uncertainties are systematically considered. Similar to [1], the proposed approach relies on wide-band signals in order to obtain a distinct mapping to the sources, but instead of ranges, the amplitude values at the sensors are directly exploited for estimation purposes. Thus, the intermediate TOF estimation step can be omitted, since the whole considered problem is mapped to the system and measurement equation of the nonlinear dynamic system model. By employing an appropriate state estimator, the state is estimated on the basis of every received measurement, i.e., the proposed approach estimates the pose in an instantaneous recursive manner. Compared to standard approaches, where the processing is performed block-wise, the assumption that the state is stationary over the block length is not longer required.

The second contribution, presented in Section 3.2, is that rotation vectors are used instead of quaternions or Euler angles. The rotation matrix, which describes the relationship between two coordinate systems, can be parameterized with quaternions [5], Euler angles, or rotation vectors [6], [7]. For example in [5], quaternions are used for orientation tracking. An unscented Kalman filter is used, where the mathematical operations are adapted to quaternions space. The main problem with quaternions results from the fact that the parameter vector is not minimal and the unit quaternions constraint can lead to inaccurate state estimates. Rotation vectors instead are a minimal state representation. A further advantage of rotation vectors is that the dynamic behavior can be described by means of a nonlinear differential equation [6] and thus, it can be applied in the system model of a Kalman filter [7].

2. MEASUREMENT MODEL

The proposed approach estimates the pose based on the amplitude values measured at the sensors. In doing so, the wave propagation from the sources to the sensors (e.g., microphones) has to be described by a model.

2.1. Sensor - Microphone

To derive a measurement equation for the microphones, the wave propagation between N sources and the sensors is considered. A

time signal emitted by a source $i = 1, \ldots, N$ at position $\underline{x}^i$ is described by the discrete-time sequence

$$s_d^i(t) = \sum_{k=-\infty}^{+\infty} s_k^i \cdot \delta\left(t - k \cdot T_s\right) ,$$

where s_k^i is the k-th data symbol and T_s is the sampling time interval. The data sequence is weighted with a source impulse filter, which results in a continuous-time signal

$$s^i(t) = \sum_{k=-\infty}^{+\infty} s_k^i \cdot h_I\left(t - k \cdot T_s\right) .$$

This continuous-time signal is transmitted through a partially unknown time-variant channel

$$h^{i,j}(t,\tau) = \frac{1}{4\pi\left\|\underline{p}^j(t)-\underline{x}^i\right\|}\delta\left(t - \tau - \frac{\left\|\underline{p}^j(t)-\underline{x}^i\right\|}{c}\right)$$

to sensor j at the position $\underline{p}^j(t)$. The channel model just delays and attenuates the emitted signal. If only the path of one source and sensor is considered, the received signal from one source is

$$y^{i,j}(t) = \int_0^t s^i(\tau)\, h^{i,j}(t,\tau)\mathrm{d}\tau .$$

The received signal at sensor j consists of the overlapping signals from the N sources. The sensor discretizes the signal in such a way that the amplitude value at each time step is given by

$$y^j(n \cdot T_e) = \sum_{i=1}^{N} \frac{1}{4\pi\left\|\underline{p}^j(n\cdot T_e)-\underline{x}^i\right\|} s^i\left(n \cdot T_e - \frac{\left\|\underline{p}^j(n\cdot T_e)-\underline{x}^i\right\|}{c}\right)$$
$$+ v^j(n \cdot T_e) , \tag{1}$$

where $v^j(n \cdot T_e)$ is the measurement noise. The sensor position $\underline{p}^j(n \cdot T_e)$ with respect to the global coordinate system is unknown. However, the sensor position $\tilde{\underline{p}}^j$ with respect to the target coordinate system is known. To convert the position from the local coordinate system to the global coordinate system, the relationship

$$\underline{p}^j(n \cdot T_e) = \mathbf{D}\left(\underline{r}_n\right)\tilde{\underline{p}}^j + \underline{T}_n \tag{2}$$

is used. $\underline{T}_n$ is the translation vector and $\underline{r}_n$ is the parameter vector, which describes the rotation. $\mathbf{D}$ is the rotation matrix. In this paper, the rotation matrix is described by the rotation vector

$$\mathbf{D}\left(\underline{r}_n\right)\tilde{\underline{p}}^j = \cos\left(\left\|\underline{r}_n\right\|\right)\tilde{\underline{p}}^j + \frac{\sin\left(\left\|\underline{r}_n\right\|\right)}{\left\|\underline{r}_n\right\|}\underline{r}_n \times \tilde{\underline{p}}^j$$
$$+ \frac{1-\cos\left(\left\|\underline{r}_n\right\|\right)}{\left\|\underline{r}_n\right\|^2}\left(\underline{r}_n^{\mathrm{T}}\tilde{\underline{p}}^j\right)\underline{r}_n . \tag{3}$$

The direction of the rotation vector is the rotation axis and the length of the vector is the angle of rotation.

The resulting measurement equation for the sensors is obtained by combining the equations (1) to (3). Furthermore, an ideal low-pass filter for the sources is assumed. Hence, the source impulse filter is given by a sinc-function. If the sampling intervals of the sources and the sensors are equal to T, the measurement equation is given by

$$y_n^j = y^j(n \cdot T) = \sum_{i=1}^{N} \frac{1}{4\pi\left\|\mathbf{D}\left(\underline{r}_n\right)\tilde{\underline{p}}^j+\underline{T}_n-\underline{x}^i\right\|}$$
$$\sum_{k=0}^{n} s_k^i \cdot \text{sinc}\left((n-k)\cdot\pi - \frac{\left\|\mathbf{D}\left(\underline{r}_n\right)\tilde{\underline{p}}^j+\underline{T}_n-\underline{x}^i\right\|}{cT}\right) + v_n^j .$$

2.2. Sensor - Gyroscope

The gyroscopes measures the angular velocity $\underline{\omega}_n$ of the object. The measurement equation is given by

$$\underline{y}_n^\omega = k_G\left(\mathbf{A}\left(\underline{\omega}_n + \underline{b}\right)\right) + \underline{v}_n^\omega , \tag{4}$$

where k_G is a sensor-specific coefficient, $\mathbf{A}$ is a misalignment matrix, $\underline{b}$ is the offset and $\underline{v}_n^\omega$ is the measurement noise. The distribution of the measurement noise is assumed to be white, zero-mean, and Gaussian.

3. SYSTEM MODEL

3.1. Translation

For the translation, a constant velocity model is assumed [8]. This model is given by a linear differential equation, which is represented in discrete time as

$$\underline{z}_{n+1} = \mathbf{A}\underline{z}_n + \underline{w}_n^z , \tag{5}$$

where $\mathbf{A}$ is given by

$$\mathbf{A} = \begin{bmatrix} \mathbf{I}_{3,3} & T\mathbf{I}_{3,3} \\ \mathbf{0} & \mathbf{I}_{3,3} \end{bmatrix}$$

and $\underline{z}_n$ consists of the translation and the velocity. The process noise is assumed to be white, zero-mean and Gaussian with the covariance

$$\mathbf{Q}_{\underline{w}_n^z} = \begin{bmatrix} \frac{T^3}{3}\mathbf{q} & \frac{T^2}{2}\mathbf{q} \\ \frac{T^2}{2}\mathbf{q} & T\mathbf{q} \end{bmatrix}, \quad \mathbf{q} = \text{diag}\left(\begin{bmatrix} q_{w_{V,x}}^2 & q_{w_{V,y}}^2 & q_{w_{V,z}}^2 \end{bmatrix}^{\mathrm{T}}\right) .$$

3.2. Rotation

In a tracking scenario the rotation vector is typically time variant. This dynamic behavior can be described by a nonlinear differential equation

$$\dot{\underline{r}}(t) = \begin{cases} a(\underline{\omega}(t), \underline{r}(t)) & \text{for } \|\underline{r}(t)\| \in\,]0, \pi] \\ \underline{\omega}(t) & \text{for } \|\underline{r}(t)\| = 0 \end{cases} , \tag{6}$$

which depends on the angular velocity. The nonlinear function $a(\cdot)$ is given by

$$a(\underline{\omega}(t), \underline{r}(t)) = \tfrac{1}{2}\|\underline{r}(t)\|\cot\left(\frac{\|\underline{r}(t)\|}{2}\right)\underline{\omega}(t) + \tfrac{1}{2}\underline{r}(t)\times\underline{\omega}(t)$$
$$+ \frac{1-\tfrac{1}{2}\|\underline{r}(t)\|\cot\left(\frac{\|\underline{r}(t)\|}{2}\right)}{\|\underline{r}(t)\|^2}\left(\underline{r}(t)^{\mathrm{T}}\underline{\omega}(t)\right)\underline{r}(t) , \tag{7}$$

where $\cot(\cdot)$ is the cotangent. In (6), the range of the norm of the rotation vector lies in the interval zero to π. In this case, we assume that the rotation vector moves in the open ball with radius π and if the norm of the rotation vector is higher than π a forwards inference is done by

$$\underline{r}_{new}(t) = \underline{r}(t)\left(1 - \frac{2\pi}{\|\underline{r}(t)\|}\right) . \tag{8}$$

The forward inference can be calculated by a prediction step, where no process noise is assumed. The nonlinear differential equation (6) is discretized by the Euler formula

$$\dot{\underline{r}}(t) \approx \frac{\underline{r}_{n+1}-\underline{r}_n}{T}$$

and the resulting discrete-time differential equation is

$$\underline{r}_{n+1} = \underline{r}_n + T\, a(\underline{\omega}_n, \underline{r}_n) . \tag{9}$$
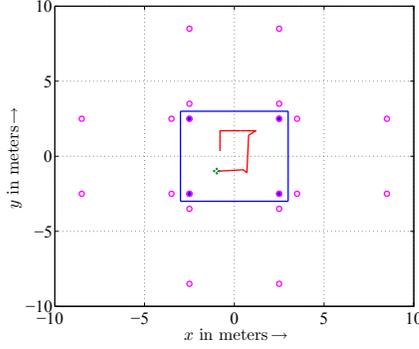
Fig. 1. Simulation setup: The blue stars are the sources, the red circles are the mirror image sources, the green point is the starting point of the trajectory, the blue lines are the walls, and the red line is the trajectory of the target object.

The angular velocity is modeled as a random walk according to

$$\underline{\omega}_{n+1} = \underline{\omega}_n + \underline{w}_n^{\omega} \ , \tag{10}$$

where the process noise is white, zero-mean, and Gaussian with covariance

$$\mathbf{Q}_{\underline{w}^{\omega}} = \mathrm{diag}\left(\begin{bmatrix} q_{w_{\omega,x}}^2 & q_{w_{\omega,y}}^2 & q_{w_{\omega,z}}^2 \end{bmatrix}^{\mathrm{T}}\right) \ .$$

4. RESULTS

The approach is evaluated in both simulations and by means of a real experiment. For estimating the state, the Gaussian estimator (GE) introduced in [9] is applied. The estimator represents the state estimate by means of a multivariate Gaussian density. The parameters of this density are calculated on the basis of a deterministic sampling approach, where the sample points are represented by a Dirac mixture density, whose deviation to the true density is minimized. Compared to the popular unscented Kalman filter (UKF) [10], the sample points of the GE are placed more systematically. Furthermore, the number of sample points is adjustable, which in turn allows approximating higher-order moments more accurately.

4.1. Simulation

In the simulation setup, a moving target object is simulated. Four microphones, which are attached to the target, receive the signals from four loudspeakers. The velocity and the angular velocity of the object are assumed to be piecewise constant. Furthermore, an inertial measurement unit consisting of three gyroscopes measures the angular velocity with respect to the target coordinate system. The measurement rate of the microphones and the gyroscopes are 48000 Hz and 480 Hz, respectively. The signals received at the microphones are corrupted with noise. This noise is generated by mirror image sources in order to model reverberations. The signals emitted by the four loudspeakers are reflected at walls of the room, which is modeled by 24 mirror image sources. The attenuation factor of the walls is set to be 0.5, which results in an SNR of 4.247 dB. The simulation setup is shown in Fig. 1.
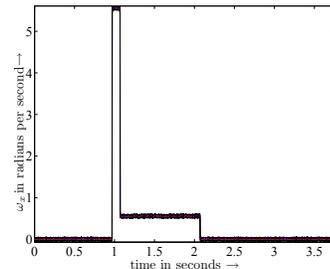
The results of the simulation are shown in Fig. 2, where every hundredth estimate is plotted. Only the translation in z, the rotation vector in x, and the angular velocity related to the x-axis of the gyroscope coordinate system are shown. In Fig. 3 and Fig. 4, the
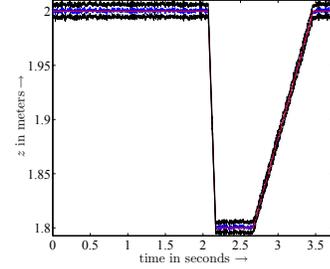
(a) Translation in z.
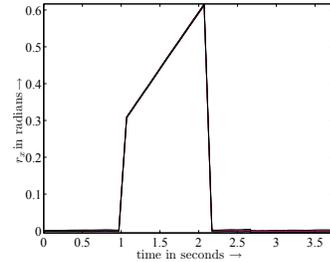
(b) Rotation vector in x.

(c) Angular velocity in x.

Fig. 2. The red line is the true trajectory of the target object, the blue line is the estimated trajectory, and the black corresponds to the three-σ-bound.

root mean square error of translation and rotation are shown. The mean error in the translation is 2 mm. If the object moves very fast, the error increases to 8 mm. The rotation error lies in the interval from 0.0014 radians to 0.016 radians. The error peaks arise from the simulation setup, where a constant angular velocity is simulated, which exhibits some discontinuities. Of course, this piecewise constant simulation model for generating the measurements is not very realistic. Real scenarios are even simpler to handle as the translational and angular velocities do not have points of discontinuities. The approach provides accurate estimates in this colored noise scenario, where the colored noise is generated by the mirror image sources.

4.2. Experiment

The experimental setup is similar to the simulation setup. In the experiment, five loudspeakers are used for periodically emitting MC-CDMA signals to achieve a distinct mapping to the sources. The measurement frequency of the gyroscopes is 200 Hz. In the experiment, the object is moved ten times on a given trajectory describing a half circle. The start and the end point are measured with $\begin{bmatrix} -0.05 & -0.75 & 1.16 \end{bmatrix}$ meters and $\begin{bmatrix} -0.04 & 1.33 & 1.16 \end{bmatrix}$ meters, respectively. The distance between start and end point was
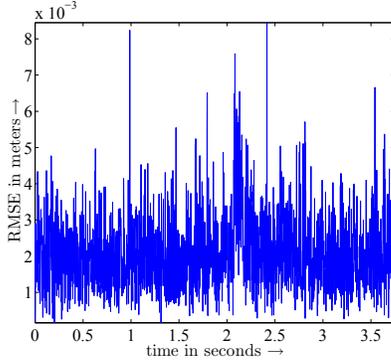
Fig. 3. RMSE in translation.

Fig. 4. RMSE in rotation.

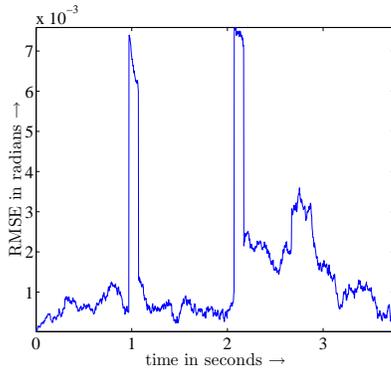**Fig. 5**. Trajectories of the ten test runs.

2.08 meters. In Fig. 5, the trajectory in xy view is shown. After a transition time, the mean of the estimates of all test runs was $\begin{bmatrix} -0.0653 & -0.7653 & 1.1593 \end{bmatrix}$ meters with a standard deviation of $\begin{bmatrix} 0.0482 & 0.0442 & 0.0259 \end{bmatrix}$ meters. After arrival at the end point, the mean of the estimates was $\begin{bmatrix} 0.0253 & 1.3194 & 1.0557 \end{bmatrix}$ meters with a standard deviation of $\begin{bmatrix} 0.0207 & 0.0634 & 0.0902 \end{bmatrix}$ meters. The distance based on the estimate was 2.0910 meters with a standard deviation of 0.0599 meters.

5. CONCLUSIONS

This paper solves the problem of pose estimation based on raw sensor data without explicit range calculation. The proposed approach employs an end-to-end problem formulation in order to obtain an integrated solution for estimating the pose of a target object from delay and attenuation of time signals under consideration of both model and measurement uncertainties, which are characterized in a stochastic manner. This model-based approach considers the physical behavior of the wave propagation and the dynamics of the object. Furthermore, the rotational motion of the object is described by a nonlinear differential equation, which allows the systematic incorporation of the angular velocity measured by a gyroscope. In doing so, the state estimator can process the measurements instantaneously and thus, batch processing as in conventional methods employing matched filters is not required.
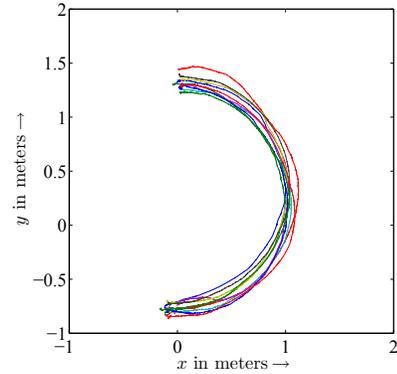
6. REFERENCES

[1] N. M Vallidis, *WHISPER: A Spread Spectrum Approach to Occlusion in Acoustic Tracking*, Ph.D. thesis, University of North Carolina at Chapel Hill, 2002.

[2] P. Rößler, F. Beutler, U. D. Hanebeck, and N. Nitzsche, "Motion Compression Applied to Guidance of a Mobile Teleoperator," in *Proceedings of the 2005 IEEE International Conference on Intelligent Robots and Systems (IROS 2005)*, 2005, pp. 2495–2500.

[3] F. van der Heijden, S. van Koningsveld, and P.P.L. Regtien, "Time-of-Flight Estimation using Extended Matched Filtering," in *Proceedings of IEEE Sensors 2004*, 2004, pp. 1460–1463.

[4] F. Beutler and U. D. Hanebeck, "Closed-Form Range-Based Posture Estimation Based on Decoupling Translation and Orientation," in *Proceedings of the IEEE Intl. Conference on Acoustics, Speech, and Signal Processing (ICASSP 05)*, 2005, pp. 989–992.

[5] E. Kraft, "A Quaternion-Based Unscented Kalman Filter for Orientation Tracking," in *Proceedings of the Sixth International Conference of Information Fusion, 2003*, 2003, vol. 1, pp. 47–54.

[6] J. E. Bortz, "A New Mathematical Formulation for Strapdown Inertial Navigation," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-7, no. 1, pp. 61–66, Januar 1971.

[7] H.-S. Ahn and C.-H. Won, "Fast Alignment Using Rotation Vector and Adaptive Kalman Filter," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 42, no. 1, pp. 70–83, Januar 2006.

[8] G. Welch, B. D. Allen, A. Ilie, and G. Bishop, "Measurement Sample Time Optimization for Human Motion Tracking/Capture Systems," in *Proceedings of Trends and Issues in Tracking for Virtual Environments, Workshop at the IEEE Virtual Reality 2007 Conference*, 2007.

[9] M. F. Huber and U. D. Hanebeck, "Gaussian Filter based on Deterministic Sampling for High Quality Nonlinear Estimation," in *Proceedings of the 17th IFAC World Congress (IFAC 2008)*, 2008.

[10] S. J. Julier and J. K. Uhlmann, "Unscented Filtering and Nonlinear Estimation," *Proceedings of the IEEE*, vol. 92, no. 3, pp. 401–422, 2004.