Estimation of the Absolute Camera Pose for Environment Recognition of Industrial Robotics

R. Schmitt*, Y. Cai and S. Fang

Laboratory for Machine Tools and Production Engineering (WZL), RWTH Aachen University, Germany
*E-Mail: R.Schmitt@wzl.rwth-aachen.de

Summary: The problem of estimating and predicting the absolute camera pose (position and orientation with respect to the world coordinate system) is approached by fusion of measurements from inertial sensors (accelerometers and gyroscopes) and robot control system. The sensor fusion approach described in this paper is based on non-linear filtering of multi-rate extended Kalman filter (EKF). In this way, accurate camera pose estimates, with improved sampling rate and robustness as well as reduced computation complexity, are available. Experiments that an industrial robot moves the sensors (camera and inertial measurement unit) in an indoor-GPS-based global referencing system are presented. The ground truth of the absolute camera pose, provided by indoor-GPS (global positioning system), allows for an objective performance evaluation. The results of environment recognition applications of industrial robotics confirm also the accuracy and robustness improvement of the estimated absolute camera pose.

Keywords: Pose Estimation, Environment Recognition, Industrial Robot.

1. Introduction to the camera pose estimation

The improvement of the efficiency of production processes is one of the key enablers for market competition [1]. Because of the speed, repeatability and efficiency of the industrial robot, its application in the production process allows manufacturing companies to increase productivity and reduce costs.

For the realization of dynamic tasks in unstructured production environment, the autonomous mobile systems need sequential and fast knowledge of the manufacturing surroundings. In recent years, the use of vision as a primary perception sensor for providing information to control autonomous systems, such as industrial robots, has grown significantly [2]. Only on the basis of the reliably 3D viewing can these systems adapt autonomously and intelligently to unknown surroundings and infer the immediate and future actions of the robot [2].

The common approaches to visual 3D perception, or environment recognition, are based on the triangulation principle [3]. Based on the known geometrical relations between the two stereo cameras, by calculating the relative perspective projection of object points in both images, their 3D world coordinates can be reconstructed up to certain accuracy [4]. In the case of a monocular moving camera, by analysing the perspective views between the acquired images, the relative camera pose with respect to its previous poses has to be estimated and 3D visual information can be extracted subsequently [5].

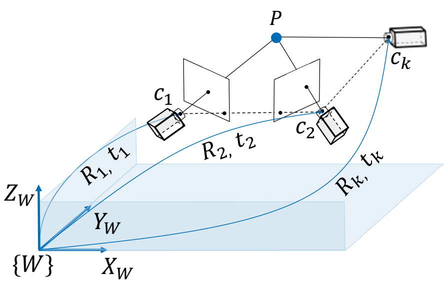

Figure 1. Camera pose and scene reconstruction problem.

The problem of the estimation of the absolute camera pose (namely its position and orientation) is illustrated in Figure 1. As the camera C moves through the Euclidean space, the distances to the imaged objects, as well as the translation t_i and rotation R_i of C_i with respect to the world coordinate system $\{W\}$ should be determined [2, 3]. The 3D position of the target point P is inherently obtained once the absolute pose of the camera has been calculated.

2. Related work to the camera pose estimation

In the computer vision area, the relative camera pose for a given image is usually estimated solely using a set of pixel correspondences. With a sufficient number of correspondences the relative camera pose can be determined uniquely [3]. The correspondence-based algorithm utilises the vision information only and does not provide an absolute camera position [4].

The problem of absolute pose estimation is approached using the combination of a camera and an inertial measurement unit (IMU) to obtain a robust system [6, 7]. These methods could give a good absolute accuracy of a few centimeters in position and a few degrees in orientation [6]. However, camera system will experience problems during periods with uninformative or no vision data. This will occur, typically due to occlusion or fast motion. The high computational complexity for image processing has to be strongly reduced to provide an accurate short time pose estimates.

The industrial robots are very flexible and are able to move with a high absolute precision (absolute accuracy of millimetre and a repeatability of tenth of a millimetre) today [1, 6]. By utilising the true position and orientation of robot movement, accurate pose estimation could be obtained, as the camera is mounted onto the robot arm (Figure 2). The relative position and orientation between robot and camera coordinate frames is determined by the hand-eye calibration [8].

This paper deals with estimating the absolute position and orientation of a camera, using measurements from inertial sensors (accelerometers and gyroscopes) and robot control data (3D position and 3D orientation). In the proposed approach, camera pose estimation is achieved by fusion of inertial and robot measurements using the framework of nonlinear state estimation (Extended Kalman Filter, EKF). The most important reasons combing inertial measurement unit (IMU) with robot are:

- Producing more robust estimates. The inertial sensor is able to measure high-dynamic motion for a short period accurately. An IMU will help to bridge gaps between the robot measurements.
- Determining the absolute position and orientation of the camera with high sampling rate and low latency. The fused camera pose has the same sampling rate as the IMU.
- Reducing computational demands. Accurate short time and long-term pose estimates are available from the IMU and robot measurement system respectively.
-

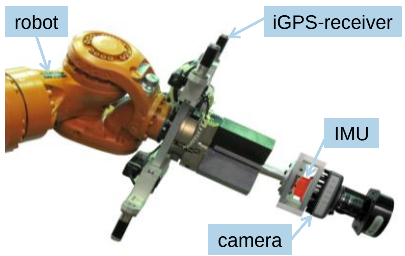

Figure 2. The camera and the IMU are mounted onto an robot.

3. System modelling and sensor fusion

As shown in Figure 2, a flange is connected at the last robot arm and four iGPS receivers, an IMU and a camera are mounted rigidly on this flange. Several different coordinate systems have to be considered to represent geometric quantities appropriately (Figure 3).

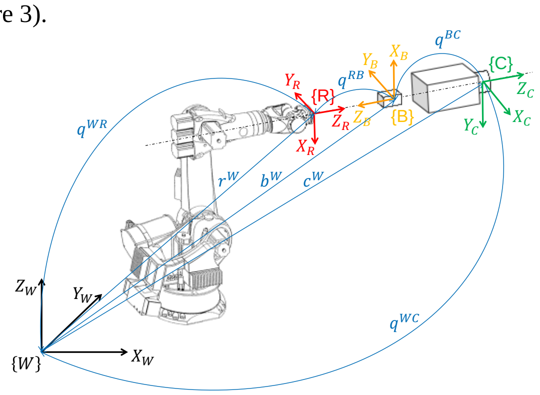

Figure 3. Coordinate systems: the world frame {W}, the robot frame {R}, the body frame {B} and the camera frame {C}.

To use the Kalman filter to estimate the camera pose recursively in time, two dynamic models of the pose estimation system are required. The process model, describes the evolution of the states (camera pose etc.) over time [6, 9].

$$\boldsymbol{b}_{k+1}^{W} = \boldsymbol{b}_{k}^{W} + T\dot{\boldsymbol{b}}_{k}^{W} + \frac{T^2}{2}\ddot{\boldsymbol{b}}_{k}^{W} \tag{1}$$

$$\dot{\boldsymbol{b}}_{k+1}^{W} = \dot{\boldsymbol{b}}_{k}^{W} + T\ddot{\boldsymbol{b}}_{k}^{W} \tag{2}$$

$$\boldsymbol{q}_{k+1}^{WB} = \boldsymbol{q}_{k}^{WB} \odot \exp\left(\begin{bmatrix} 0 \\ \frac{T}{2}\boldsymbol{\omega}_{WB,k}^{B} \end{bmatrix}\right) \tag{3}$$

$$\boldsymbol{\delta}_{a,k+1}^{B} = \boldsymbol{\delta}_{a,k}^{B} + \boldsymbol{v}_{\delta_a,k}^{B} \tag{4}$$

$$\boldsymbol{\delta}_{\omega,k+1}^{B} = \boldsymbol{\delta}_{\omega,k}^{B} + \boldsymbol{v}_{\delta_\omega,k}^{B} \tag{5}$$

The measurement model infers all the information from the sensor measurements onto the state and forms how the measurements from the IMU and the robot relate to the state [10, 11].

$$\boldsymbol{y}_{a,k} = (\boldsymbol{R}_{k}^{WB})'\left(\ddot{\boldsymbol{b}}_{k}^{W} - \boldsymbol{g}^{W}\right) + \boldsymbol{\delta}_{a,k}^{B} + \boldsymbol{e}_{a,k}^{B} \tag{6}$$

$$\boldsymbol{y}_{\omega,k} = \boldsymbol{\omega}_{WB,k}^{B} + \boldsymbol{\delta}_{\omega,k}^{B} + \boldsymbol{e}_{\omega,k}^{B} \tag{7}$$

$$\boldsymbol{y}_{p,k} = \boldsymbol{b}_{k}^{W} - \boldsymbol{R}_{k}^{WB}\boldsymbol{R}^{BR}\boldsymbol{b}^{R} + \boldsymbol{e}_{p,k}^{W} \tag{8}$$

$$\boldsymbol{y}_{o,k} = \exp\left(\begin{bmatrix} 0 \\ \frac{1}{2}\boldsymbol{e}_{o,k}^{W} \end{bmatrix}\right) \odot (\boldsymbol{q}_{k}^{WB} \odot \boldsymbol{q}^{BR}) \tag{9}$$

Since the process and the measurement models are both non-linear, the extended Kalman filter (EKF, [12]) is used for data fusion and state estimation. This iterative scheme of EKF update proceeds as follows:

1. Set initial state estimate and covariance.
2. Apply the time update and with the process model with the inertial measurements as input signals.
3. Perform measurement update with the IMU measurements and, when available, the robot measurement.
4. For the next time instant iterate from step 2.

4. Indoor-GPS as global referencing system

The indoor-Global Positioning System (iGPS) is a relatively new measurement technology for use in large-scale metrology and tracking applications [13]. The system is able to measure both static and dynamic performances of the target objects with six degree of freedom (DoF) measurements.

The iGPS system operates on the same general principle as traditional GPS and determines the position of sensors within a measurement volume encompassed by a network of transmitters (Figure 4). The transmitters send one-way signals to the sensors. Similar to spherical coordinate system, the ray between a sensor and a transmitter can be defined uniquely by the combination of azimuth and elevation angles (Figure 4). Under the assumption of calibrated relative position and orientation of the transmitters, the intersecting rays from multiple transmitters at a sensor enable the triangulation for position calculation (Figure 4).

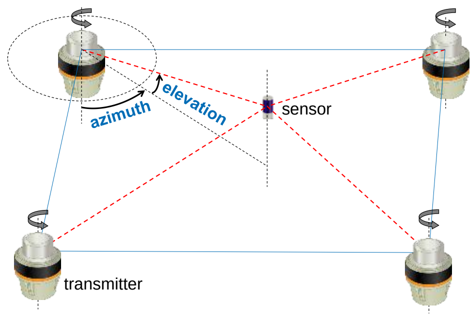

Figure 4. The iGPS network und sensor triangulation.

5. Experimental result

A number of experiments on the camera pose estimation are conducted at the indoor-GPS based robot cell at WZL. Based on the true position and orientation provided by iGPS, an absolute pose error can be calculated and this enables an objective performance evaluation.

The camera pose estimates are compared with the ground truth data and the position errors are shown separately for each axis in Figure 5a. The absolute accuracy for position is below 0.5 cm for x-axis, 1 cm for y-axis and 2 cm for z-axis.

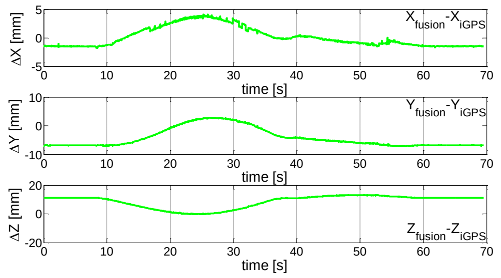

Figure 5a. Position errors of the semi-circle-shaped trajectory.

The orientations exhibit similar behaviour (Figure 5b). The absolute accuracy is below 1° for roll (rotation around x-axis), 0.5° for pitch (rotation around y-axis) and 2° for yaw (rotation around z-axis).

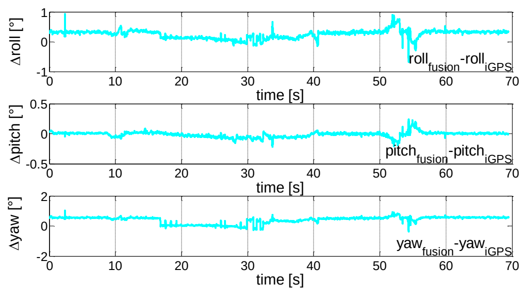

Figure 5b. Orientation errors of the semi-circle-shaped trajectory.

6. Robot vision application

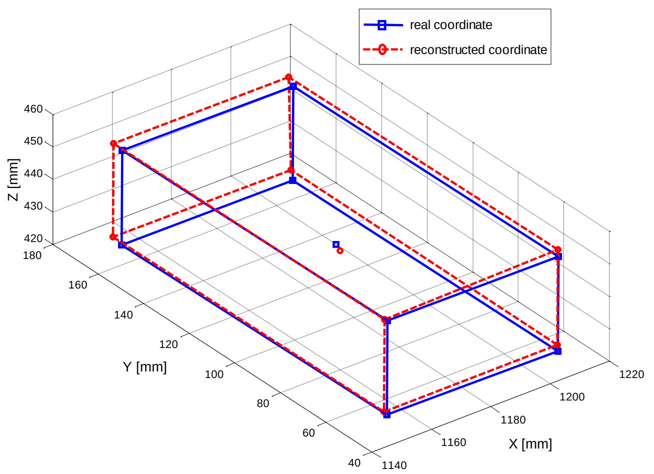

Figure 6. 3D reconstruction of test block

The proposed method to estimate the absolute camera pose is also applied to real 3D Euclidean reconstruction problem. A sequence of images of is acquired, while the camera is moving with the robot arm around a static cuboid. For the reconstruction of each corner point, the camera pose is estimated from the algorithm presented in this paper. From the reconstructed corner points the coordinate of the block center is calculated, whose deviation to the real one accounts for about 2.0269 mm (Figure 6).

7. Conclusion

Based on a nonlinear state estimation, a system has been developed to obtain 250 Hz camera pose estimates by fusion of inertial measurements and robot measurement data using an EKF. Experimental applications in industrial robots show that this setup is able to estimate the absolute camera pose with high accuracy, improved robustness and reduced computational complexity.

Acknowledgements

The authors are very grateful to the DFG (Deutsche Forschungsgemeinschaft, German Research Foundation) for financial supporting this work as part of the research project (SCMH-1856/23-1) “Scene recognition with a monocular, accelerated moving camera in industrial robotics”.

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions to improve the quality of the paper.

References

[1] Schmitt, R., Schoenberg, A., Damm, B., 2010, Indoor-GPS based robots as a key technology for versatile production, Proceedings for the joint conference of ISR 2010 (41st International Symposium on Robotics) and ROBOTIK 2010 (6th German Conference on Robotics):199 -205.

[2] Kragic, D., Christensen, H.-I., 2005, Advances in robot vision, Robotics and Autonomous Systems, Elesvier: 52:1-3.

[3] Hartley, R., Zisserman, A., 2004, Multiple View Geometry in Computer Vision, 2. Edition, Cambridge University Press.

[4] Ma, Y., Soatto, S., Kosecka, J., Sastry, S., 2003, An Invitation to 3-D Vision, From Images to Geometric Models, Springer.

[5] Moons, T., Gool, L. V., Vergauwen, M., 2009, 3D Reconstruction from Multiple Images, Part 1: Principles, Now Publishers.

[6] Hol, J. D., Schoen, T. B., Gustafsson, F., 2010, Modeling and calibration of inertial and vision sensors, international Journal of Robotics Research, 29/2:231-244.

[7] Corke, P., Lobo, J., Dias, J., 2007, An introduction to inertial and visual sensing, International Journal of Robotics Research, 26/6:519-535.

[8] Tsai, R. Y., Lenz, R. K., 1989, A new technique for fully autonomous and efficient 3D robotics hand/eye calibration, IEEE Transactions on Robotics and Automation, 5/3: 345-358.

[9] Titterton, D. H., Weston, J. L., 1997, Strapdown inertial navigation technology, IEE radar, sonar, navigation and avionics series, Peter Peregrinus Ltd..

[10] Hol, J. D., 2011, Sensor Fusion and Calibration of Inertial Sensors, Vision, Ultra-Wideband and GPS, dissertations, Linköping University.

[11] Kuipers, J. B., 1999, Quaternions and Rotation Sequences, Princeton University Press.
[12] Maybeck, P. S., 1979, Stochastic models, estimation, and control: Volume 1, Academic Press.
[13] Norman, A., Schoenberg, A., Gorlach, I., Schmitt, R., 2010, Cooperation of Industrial Robots with Indoor-GPS, Proceedings of the International Conference on Competitive Manufacturing: 215-224.