TRACKING AND POSE ESTIMATION OF NON-COOPERATIVE SATELLITE FOR ON-ORBIT SERVICING

Nassir W. Oumer and Giorgio Panin

German Aerospace Center (DLR), Münchner straße 20, D-82234 Weßling, Germany

Email: {nassir.oumer, giorgio.panin}@dlr.de

ABSTRACT

We present a camera-based 3D feature tracking method, integrated into a bank of iterated, extended Kalman filters (IEKF), associated to each visual feature. The approach exploits motion field to estimate velocities of the rigid body. The depths are required only for initialization, and can be obtained either from joint estimation of depth and motion or from stereo correspondence. The motion of each 3D point cloud is predicted under a common rigid velocity constraint. A robust pose estimator, based on dual-quaternions and median statistics, is further applied to the estimated points. In case of temporarily missing measurements, the last estimated body velocity is used to predict the next poses. Results are shown on images of a satellite-mockup, to demonstrate performances for on-orbit servicing in space environment.

Key words: 3D tracking; Pose estimation; On-orbit servicing; EKF.

1. INTRODUCTION

Relative state estimation between space vehicles has been addressed in the context of formation flying, rendezvous and docking ([5], [3]), where the vehicles cooperatively exchange state information. In such cooperative scenario, vision-based pose tracking can be suppported with special markers on the target vehicle, simplifying the pose estimation problem.

Visual tracking of a non-cooperative space object, such as a malfunctioning satellite, is a current important research topic. One of the applications foreseen is on-orbit servicing: there exist plenty of defective satellites in space, occupying precious orbits: the ultimate goal would be either to repair, or to deorbit them. This servicing or de-orbiting task may be achieved through a robot mounted on the servicer satellite. For that purpose, the relative pose of the defective client satellite with respect to the servicer needs to be tracked and predicted over time, so that the servicer may approach and capture the client for visual servoing. One example of such a mission under development, is the German on-orbit servicing mission (DEOS).

There exist few works in the literature that address the problems of visual pose tracking of non-cooperative satellites. Most of them assume existence of a CAD model of the defective satellite: for example, [11] implemented an iterative closest point (ICP) algorithm to register model and range data, obtained from stereo cameras. Similarly, a wireframe model was used in [4] for camera-based tracking of Orbital Life Extension Vehicle (OLEV). Joint estimation of pose and dynamic parameters, for guiding a robotic manipulator to capture a tumbling satellite, based on Neptec's laser-camera system and a CAD model, is experimentally demonstrated in [1]. The aforementioned approaches are suitable, if apriori knowledge about the geometry of the satellite and the initial pose are given. Without such knowledge, the problem becomes more difficult.

To address this issue, Hillenbrand and Lampariello [7] developed a least-squares method for pose estimation using various simulated range data, that also identifies the six inertial parameters and center of mass of the target. A model-free, stereo-camera based approach using an extended Kalman filter was also used to estimate structure, relative pose and motion of non-cooperative targets [9]. In this scheme, inertial parameters are identified through a hypothesis-based likelihood score, over a finite set of possibilities, and numerical simulations have been presented.

On the other hand, traditional tracking approaches using feature points can be applied on relatively variable lighting conditions. Corner-like features [20] can be selected for sparse optical flow, with a multiresolution implementation satisfying most of the required real-time efficiency and accuracy. Since the above mentioned work, many extensions appeared. A few of them incorporated photometric and geometric cues of a scene [19], improved computational effeciency through an inverse compositional approach [2, 14]. Gouiffés et al. [6] also considered specular highlights and brightness variations. Such approaches require feature correspondences across stereo images, to reconstruct 3D points which, in turn, are registered in order to estimate rigid motion. For this purpose, vari-

ous pose estimation algorithms, based on the SVD [17] and [10], quaternions [8] or dual quaternions [13] can be used.

Our work is based on the existing relationship between motion field on the image plane and camera motion ([12], [18]), to recover the velocity of a rigid body. We propose a method to recover and track 3D sparse structure, exploiting specular keypoints of the rigid target, with the capability of predicting features in absence of image measurements. The approach is based on a bank of iterated, extended Kalman filters (IEKF), using pointwise kinematic models, which feeds the filtered state (depth) to the velocity estimator, and vice-versa. Thus, unlike [21] it does not require stereo matching throughout the sequence, nor any global model of target geometry or dynamics. This is particularly relevant for on-orbit servicing, where direct sunlight saturates the CCD sensor of the cameras, and a geometrical model of the client may not be available.

The paper is structured as follows: in section 2, we present the details of the proposed approach for tracking a predominantly specular and non-cooperative client satellite, in six degrees of freedom. In section 3 results on terresterial simulations are presented, and section 4 summarizes and concludes the paper.

2. THE 3D TRACKING APPROACH

During motion of a free-floating rigid body with negligible external torque or force, it is reasonable to assume a constant linear and angular velocity between two image sampling times. With this assumption, the last velocity from previous frame can be used to propagate the 3D points to the current frame. Hence the current a priori state, predicted from the last state using the estimated rigid body velocity, is combined with current measurement (tracked features) to refine the posterior state estimate. This recursive probabilistic refinement is the well known Kalman filter. The necessary models for the Kalman filter is briefly explained in the next subsections.

2.1. State and Observation Models

The motion of 3D surface points $\mathbf{p} = [X, Y, Z]^T$ of a rigid body, translating and rotating at velocity $\mathbf{V}, \boldsymbol{\omega}$ in camera frame, respectively, is expressed by

$$\dot{\mathbf{p}} = -\mathbf{V} - \boldsymbol{\omega} \times \mathbf{p} \tag{1}$$

The single-point dynamics at discrete time k can be easily derived from Equation 1, and is expressed as

$$\mathbf{p}_k = \mathbf{F}_k \mathbf{p}_{k-1} + \mathbf{G}\mathbf{V}_{k-1} + \mathbf{w}_k \tag{2}$$

where,

$$\mathbf{F}_k = \begin{bmatrix} 1 & \Delta t\omega_{zk} & -\Delta t\omega_{yk} \\ -\Delta t\omega_{zk} & 1 & \Delta t\omega_{xk} \\ \Delta t\omega_{yk} & -\Delta t\omega_{xk} & 1 \end{bmatrix} \tag{3}$$

$$\mathbf{G} = \begin{bmatrix} -\Delta t & 0 & 0 \\ 0 & -\Delta t & 0 \\ 0 & 0 & -\Delta t \end{bmatrix} \tag{4}$$

$$\mathbf{V}_k = \begin{bmatrix} V_{xk} & V_{yk} & V_{zk} \end{bmatrix}^T \tag{5}$$

and the measurement model is

$$\mathbf{z}_k = h(\mathbf{p}_{k/k-1}) + \mathbf{v}_k \tag{6}$$

where, Δt is time interval between two consecutive images and $\mathbf{z}_k$ are feature points, either from monocular or stereo camera, tracked over the sequence and re-projected under a pinhole camera model. Measurement errors $\mathbf{v}_k$ and error in motion model $\mathbf{w}_k$ are assumed Gaussian, zero-mean additive noise.

For rectified stereo cameras, with baseline T_x, we have

$$\mathbf{h} = \begin{bmatrix} f_{xl}X/Z + c_{xl} \\ f_{yl}Y/Z + c_{yl} \\ f_{xr}(X + T_x)/Z + c_{xr} \\ f_{yr}Y/Z + c_{yr} \end{bmatrix} \tag{7}$$

The y coordinate of the rectified system is common, hence one element of $\mathbf{h}$ is redundant and can be discarded. From Equation 7, the observation matrix H_k is computed as

$$H_k = \left.\frac{\partial \mathbf{h}}{\partial \mathbf{p}}\right|_{\mathbf{p}=\mathbf{p}_{k|k-1}} \tag{8}$$

$$\frac{\partial \mathbf{h}}{\partial \mathbf{p}} = \begin{bmatrix} f_{xl}/Z & 0 & -f_{xl}X/Z^2 \\ 0 & f_{yl}/Z & -f_{yl}Y/Z^2 \\ f_{xr}/Z & 0 & -f_{xr}(X + T_x)/Z^2 \\ 0 & f_{yr} & -f_{yr}Y/Z^2 \end{bmatrix} \tag{9}$$

where, l and r indicate left and right stereo camera, respectively.

In the case of monocular camera, only the first two rows of Equation 7 and 9 involve the observation model. To cope with its nonlinearity, after prediction the filter is iterated for each point, until convergence.

The main advantage of Bayesian filtering here is that, when a complete occlusion occurs, or the target moves momentarily out of the field of view, motion can be still predicted and used later on to re-initialize tracking.

2.2. Velocity Estimation and Depth Initialization

The translational and rotational velocity of the client satellite, which is used in the prediction stage of Bayesian

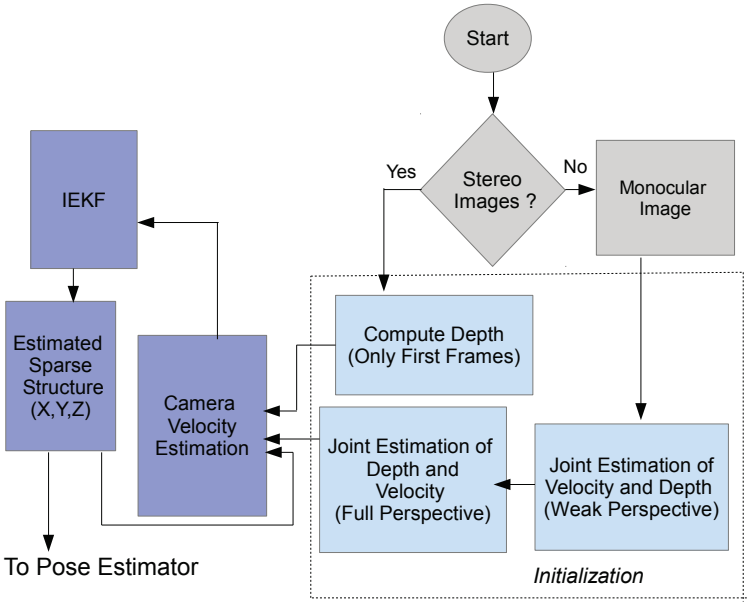

Figure 1. Initialization of rigid body velocity estimate, and interaction with Extended Kalman filter during tracking.

filter, can be estimated from the equation of motion field and camera motion, by

$$\underset{\mathbf{b}}{\operatorname{argmin}} \left\| \mathbf{A}^T \mathbf{b} - \mathbf{u} \right\|^2 \tag{10}$$

where

$$\mathbf{A} = \begin{bmatrix} -f_x/Z & 0 \\ 0 & -f_y/Z \\ (x-c_x)/Z & (y-c_y)/Z \\ (x-c_x)(y-c_y)/f_y & f_y(1+((y-c_y)/f_y)^2) \\ -f_x(1+((x-c_x)/f_x)^2) & -(y-c_y)(x-c_x)/f_x \\ f_x(y-c_y)/f_y & -f_y(x-c_x)/f_x \end{bmatrix}$$

$$\mathbf{b} = \begin{bmatrix} V_x & V_y & V_z & \omega_x & \omega_y & \omega_z \end{bmatrix}^T \tag{11}$$

is the velocity components to be estimated, and $\mathbf{u} = \begin{bmatrix} u & v \end{bmatrix}^T$ is feature velocity in the image

The depth, which is required to estimate velocites, is obtained from the update stage of the filter. Thus, velocity estimator and 3D feature tracker exchange their outputs, in order to estimate the required states (Fig. 1). However, depth needs to be initialzed externally at the onset of tracking, in order to estimate the velocities. If stereo camera is used, correpondence is searched in the first stereo pair, to compute depth for the initialization. In the case of monocular camera, joint estimation of depth and velocitiy is achieved by minimizing the error

$$\mathbf{e}_i = \mathbf{u}_i - \frac{1}{Z_i}\mathbf{A}_1(\mathbf{x}_i)\mathbf{V} - \mathbf{A}_2(\mathbf{x}_i)\boldsymbol{\omega}. \tag{12}$$

where $\mathbf{A}_1$ and $\mathbf{A}_2$ are 2×3 matrices formed from the transposes of the first three and the last three rows of matrix $\mathbf{A}$ respectively.

By concatenating all N residuals, $\mathbf{e} = [\mathbf{e}_1, ..., \mathbf{e}_N]^T$ as well as structure and motion parameters $\boldsymbol{\theta} = [\mathbf{V}, \boldsymbol{\omega}, Z_i]^T$ where $i = 1, ...N$, with proper initialization $\boldsymbol{\theta}$ are estimated by iterating

$$\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k + \delta\boldsymbol{\theta}_k \tag{13}$$

k times until convergence. The Jacobian of size $(2N) \times (6+N)$ is computed as

$$\mathbf{J} = \begin{bmatrix} \mathbf{J}_1 & \mathbf{J}_{1z} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{J}_2 & \mathbf{0} & \mathbf{J}_{2z} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{J}_3 & \mathbf{0} & \mathbf{0} & \mathbf{J}_{3z} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \mathbf{0} \\ \mathbf{J}_N & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{J}_{Nz} \end{bmatrix} \tag{14}$$

where $\mathbf{J}_i = -A^T$ at point locations $i = 1, ..., N$, and

$$\mathbf{J}_{iz} = \begin{bmatrix} (-f_x V_x + (x-c_x)V_z)/Z^2 \\ (-f_y V_y + (y-c_y)V_z)/Z^2 \end{bmatrix}. \tag{15}$$

Thus, the Gauss-Newton method is iteratively employed, by solving the normal equations

$$\mathbf{J}^T\mathbf{J}\delta\boldsymbol{\theta}_k = -\mathbf{J}^T\mathbf{e}. \tag{16}$$

Non-linear optimization over relatively large parameter space ($6 + N$ parameters, $N \geq 6$) could be prone to local optima, and good initialization is required to find the global minimum. Zucchelli et. al [15] used the separability of the differential epipolar constraint equation

and a subspace method was employed in [16] to initialize it. We propose an alternative method to initialize the iterative nonlinear optimization, that exploits the previous Gauss-Newton formulation but with less parameters.

We assume a weak perspective camera model, that considers relative variations of depth of the client to be small compared to the average depth. Consequently, we estimate only 7 parameters (6 velocites and a depth) using the same iterative method described above, but with Jacobian $\boldsymbol{J}_w$

$$\boldsymbol{J}_w = \begin{bmatrix} \boldsymbol{J}_1 & \boldsymbol{J}_{1z} \\ \vdots & \vdots \\ \boldsymbol{J}_N & \boldsymbol{J}_{Nz} \end{bmatrix} \tag{17}$$

A dual quaternion-based 3D point registration [13] is lastly utilized to estimate the pose relative to inital pose. A robust median statistics is incorporated into the pose estimator to prune outliers.

3. EXPERIMENTAL RESULTS

In this section, experimental results are presented to validate the performance of our tracking approach under space lighting conditions, based on a terrestrial simulation mockup. For this purpose, several ground-truth trajectories, consisting of roto-translation at the close range of a few meters, are generated according to the motion of a free-floating rigid body around its principal axes. The trajectories include a maximum spinning rate of 4

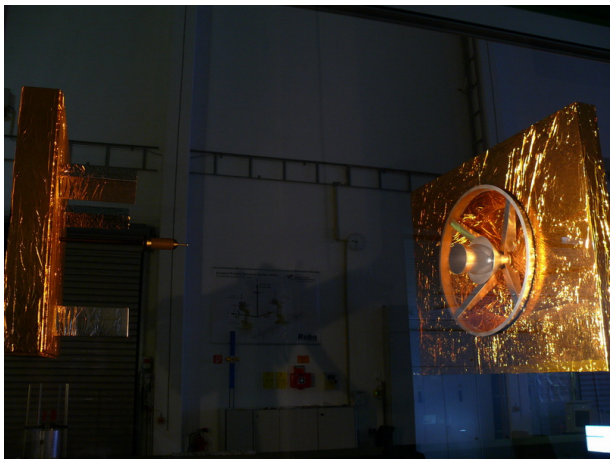

Figure 2. The DLR satellite mockup, for rendezvous and capture.

deg/s, nutation angle of 10 deg, and translational velocity 2 cm/s. Stereo cameras with a small baseline (12 cm), are mounted on the servicer, to capture images at 1Hz. The proposed method is realized with the European Proximity Operations Simulator (EPOS), using two 6 degree of freedom (dof) Kuka robots that simulate a servicer and client satellite, respectively (Fig. 2). The client satellite consists of a rigid structure, covered with a reflective golden foil. A direct high-power floodlight, simulating the spectrum of the sun, illuminates the surface of the client at different incident angles. In addition to stable features located on the surface irregularities, the overall specularity and illumination conditions produce predominatly virtual features.

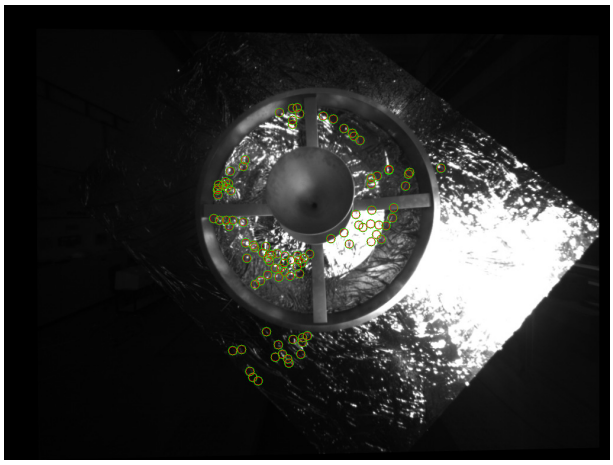

Figure 3. Reprojection of the recovered 3D points: correctly reprojected points (green) coincide with tracked features (red). Circles are drawn around the pixels, with one pixel difference in radii for visualization purpose.

.

Key-points are detected on the first frame, and tracked over the sequence [20]. No feature rejection and replacement are performed during the 2D tracking, as we need to estimate all client poses relative to its inital position, where the first frame has been acquired. However, spurious features are effectively rejected at higher level, by the robust pose estimator. When the pose estimator runs out of the minimum required number of points, a new reference frame is taken, its keypoints are detected, and the resulting pose estimations are done relatively to the new frame.

Kalman filter is usually initialized by guessing an initial state, in absence of covariance estimates. A larger initial covariance ensures that the estimate converges quickly, and the influence of the initial state will soon become negligible. In this experiment, we initialized all the states with approximately zero values, and identity covariance. The limitation of the approach is that monocular camera tracking requires slightly accurate initialization of the velocity. However, it is not difficult to predict the range of probable velocity of a free-floating space object.

The reprojection of reconstructed sparse 3D points, on the 20^{th} image, is shown in Fig. 3, where measurements and reprojections are marked with circles of different colors. Some of the results of absolute pose estimation, obtained by registering the sparse 3D cloud from the first and current frame, for four different motion trajectories are presented.

Fig. 4 displays estimated and actual roto-translation about

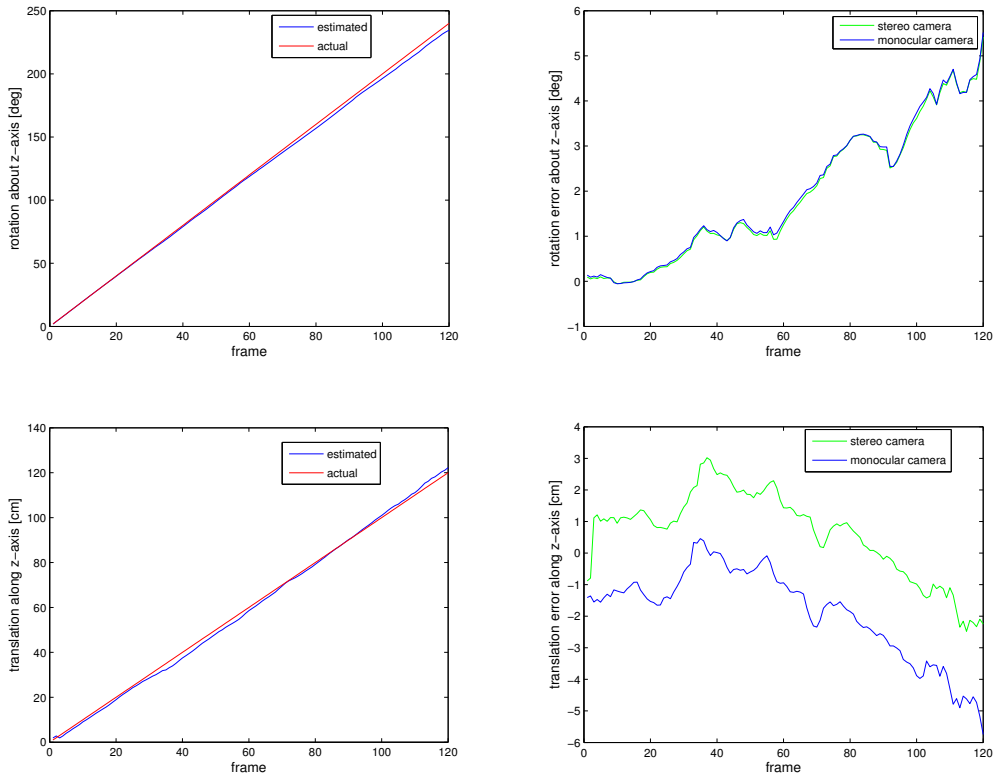

Figure 4. Estimated absolute Z-axis roto-translation with monocular camera, referred to first frame (left column), and absolute roto-translation errors with monocular and stereo camera. The client spins at 2 deg/s with a nutation angle of 5 deg, and translates at 1cm/s.

the spinning axis, for a typical spinning motion at 2 deg/s rate, with 5 deg nutation angle and 1 cm/s translational velocity. The absolute roto-translation errors in this sequence do not exceed 5 deg and 2 cm, for a period of 2 minutes respectively. Fig. 4 shows a similar trajectory that differs only in nutation angle (5 deg). Another motion trajectory, with a spinning rate 4 deg/s and translational velocity 1 cm/s and without nutation, is considered in Fig. 6. In this scenario, translation errors do not exceed 2 cm, and the worst-case rotation error is 10 deg at 300 deg rotation about Z-axis (which is 3% relative error). Related to this motion trajectory, which differs only in translational velocity (2 cm/s) is shown in Fig. 7. We notice that the error in this case is slightly higher than the former, whereby key-features could not be tracked after frame 60, indicating that a new reference frame should be taken. In general, with higher velocity and rotation angles, most feature points are quickly lost because of sensor saturation and strong light reflections.

We notice that no loop closure, such as bundle adjustment, is done to refine the pose estimation. Moreover, motion trajectories which contain both rotational and translational accelerations have not been tested, and will be included in future experiments. However, the assumption of constant velocity between two sampling times has shown to be a good approximation for predicting feature points during free-floating motion.

4. CONCLUSIONS

In this paper, we presented a system for feature-based 3D tracking of a non-cooperative satellite. The translational and rotaional velocities of the client satellite are estimated in camera frame, using image-plane motion fields. The estimated rigid motion is used by a bank of iterated Kalman filters, to propagate the point cloud to the next frame, where predicted points are updated with tracked image features. The tracked 3D points, in turn, are used to estimate the velocity of the rigid client; thus, depth estimates from the filter are passed on to the velocity estimator, and vice-versa. The monocular camera-based tracking is initialized by a joint estimation of depth and velocity, assuming a weak perspective camera model. By registering point clouds, the 6 dof motion is estimated through a robust dual-quaternion method.

Preliminary experimental results have been obtained from a simulation mockup, demonstrating robustness and accuracy properties of the proposed method. Future system improvements will be the integration of additional visual features (points, edges, lines, etc) into the filter, possibly including loop-closure mechanisms to compensate drift, incorporating a photometric model that compensates reflections and variation of illumination, and using more accurate models of tumbling rigid-body dynam-

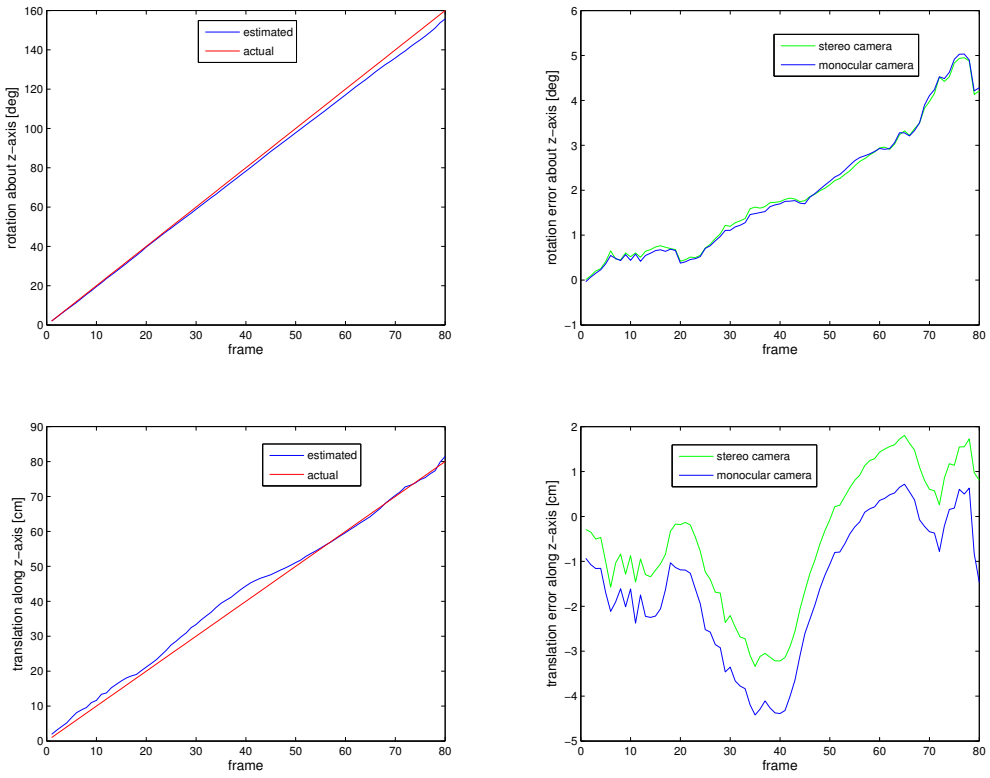

Figure 5. Estimated absolute Z-axis roto-translation with monocular camera, referred to first frame (left column), and absolute roto-translation errors with monocular and stereo camera. The client spins at 2 deg/s and translates at 1cm/s.

ics.

ACKNOWLEGMENTS

We would like to gratefully acknowelge Dr. Toralf Boge and Tilman Wimmer for their valuable suppport in realizing the experimental motion trajectories at the EPOS facility. Also we would like to express our gratitude to Dr. Ulrich Hillenbrand and Dr. Wolfgang Sepp, for their valuable contribution during the generation of ground-truth trajectories.

REFERENCES

[1] Farhad Aghili and Kourosh Parsa. An adaptive vision system for guidance of a robotic manipulator to capture a tumbling satellite with unknown dynamics. IEEE/RSJ International Conference on Intelligence Robots and Systems, pages 3064–3071, 2008.

[2] S. Baker and I. Mathews. Lucas-kanade 20 years on: A unifying framework. Int. Journal of Computer Vision, 56(3):221–225, 2004.

[3] P. Bodin, R. Noteborn, R. Larsson, S. DAmico, J.S. Ardaens, M. Delpech, and J.C. Berges. The prisma formation flying demonstrator: Overview and conclusion from the nominal mission. In proceedings of the 35th annual AAS Guidance and Control Conference, Colorado, USA, Febraury 2012.

[4] C. Miravet, L. Pascual, E. Krouch, and J.M. Delcura. An image-based sensor system for autonmous rendezvous with uncooperative satellites. In 7th International ESA Conference on Guidance, Navigation and Control Systems, June 2008.

[5] Robert B. Friend. Orbital express program summary and mission overview. In proceedings of the SPIE: Sensors and Systems for Space Applications II, volume 6958, Orlando, Florida, USA, march 2008.

[6] Michèle Gouiffès, Christophe Collewet, Christine Fernandez-Maloigne, and Alain Trémeau. Feature points tracking: Robustness to specular highlights and lighting changes. In 9th European Conference on Computer Vision, ECCV (4) 2006, pages 82–93, 2006.

[7] Hillenbrand and Roberto Lampariello. Motion and parameter estimation of a free-floating space object from range data for motion prediction. 8th International Symposium on Artifical Intelligence, Robotics and Automation in Space: i-SAIRAS 2005.

[8] Berthold K.P. Horn. A Closed-form Solution of Absolute Orientation Using Unit Quaternions. Journal of the Optical Society of America, 4(4):629–642, 1987.

[9] Shai Segal, Avishy Carmi, and Pini Gurfil. Vision-based relative state estimation of non-cooperative

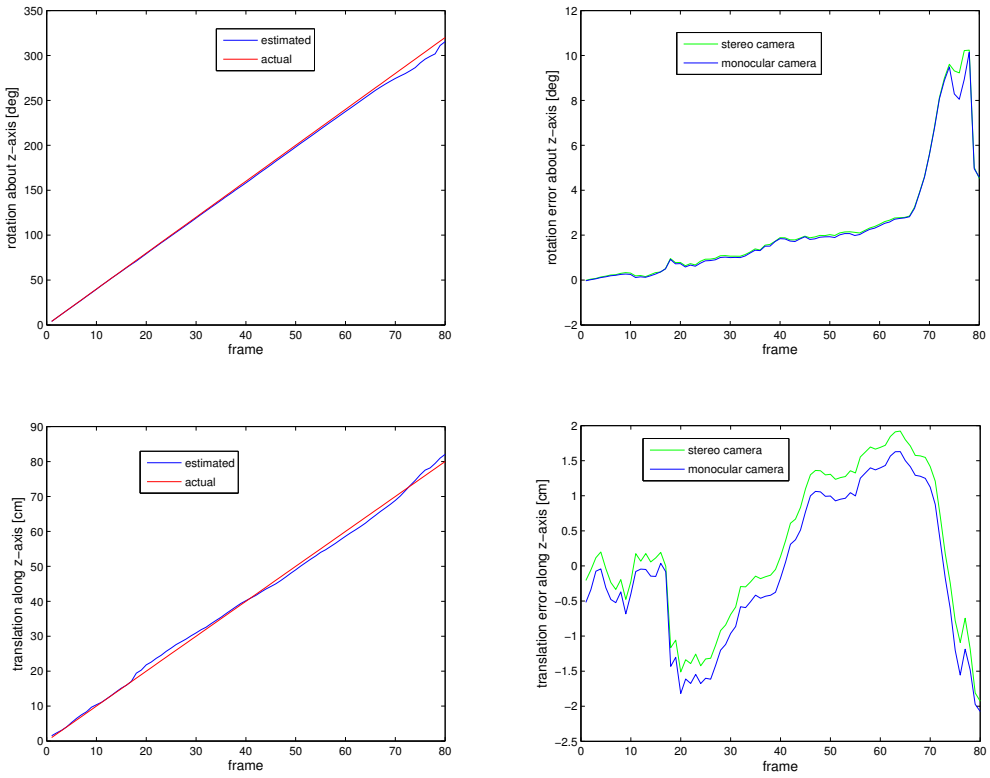

Figure 6. Estimated absolute Z-axis roto-translation with monocular camera, referred to the first frame (left column), and absolute roto-translation errors with monocular and stereo camera. The client spins at 4 deg/s and translates at 1cm/s.

spacecraft under modeling uncertainity. IEEE Aerospace Conference, 1:1–8, 2011.

[10] S. Umeyama. Least-squares estimation of transformation parameters between two point patterns. IEEE Trans. Pattern Analysis and Machine Intelligence, 13:376–380, 1991.

[11] Fuyuto Terui. Model based visual relative motion estimation and control of a spacecraft utilizing computer graphics. In 21st International Symposium on Space Flight dynamics, Tolouse, France, 2009.

[12] Allen M. WaxMan and James H. Duncan. Binocular image flows: Steps toward stereo-motion fusion. IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI 8(6), 1986.

[13] Michael W. Walker and Lejun Shao. Estimating 3-d location parameters using dual number quaternions. CVGIP : Image Understanding,54(3):358–367, 1991.

[14] Timo Zinsser, Christoph Grl, and Heinrich Niemann. Effecient feature tracking for long video sequences. Pattern Recognition, 26th DAGM Symposium, 2008.

[15] Marco Zucchelli, Josè Santos-Victor, and Henrik Christensen. Constrained structure and motion estimation from optical flow. In Proc. of the 2002 International Conference on Pattern Recognition, ICPR,Quebec, Canada, August 2002.

[16] David J. Heeger and Allan D. Jepson. Subspace methods for recovering rigid motion: Algorithm and implementation. In International Journal of Computer Vision, volume 7, pages 95–117, 1992.

[17] K.S. Arun, T.S. Huang, and S.D. Blostein. Least-squares fitting of two 3d point sets. IEEE Trans. Pattern Analysis and Machine Intelligence, 9:698–700, 1987.

[18] Shahriar Negahdaripour. Motion recovery from images sequences using only first order optical flow information. International Journal of Computer Vision, pages 163–184, 1992.

[19] Shahriar Negahdaripour. Revised definition of optical flow: Integration of radiometric and geometric cues for dynamic scene analysis. IEEE Trans. Pattern Anal. Mach. Intell., 20(9), 1998.

[20] Jianbo Shi and Carlo Tomasi. Good features to track. Proc. IEEE Conference on Computer Vision and Pattern Recognition, pages 593–600, 1994.

[21] Nassir W. Oumer and Giorgio Panin. 3d point tracking and pose estimation of a space object using stereo images. In 21th International Conference on Pattern Recognition, ICPR 2012, Tsukuba Science City, Japan, 2012 (to appear).

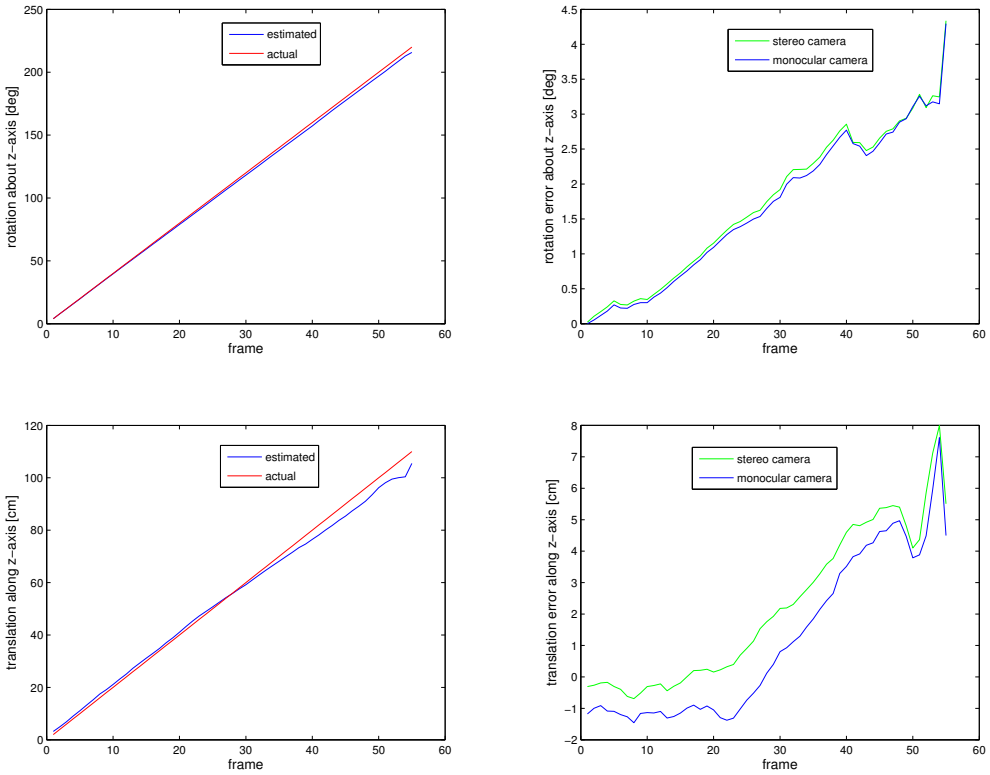

Figure 7. Estimated absolute Z-axis roto-translation with monocular camera, referred to first frame (left column), and absolute roto-translation errors with monocular and stereo camera. The client spins at 4 deg/s and translates at 2cm/s.