ISSN 2075-1087, Gyroscopy and Navigation, 2017, Vol. 8, No. 2, pp. 150–158. © Pleiades Publishing, Ltd., 2017.
Published in Russian in Giroskopiya i Navigatsiya, 2016, No. 4, pp. 98–111.

Monocular Vision-Based Range Estimation Supported by Proprioceptive Motion

P. Davidson*, J.-P. Raunio, and R. Piché

Tampere University of Technology, Tampere, Finland

**e-mail: pavel.davidson@tut.fi*

Received September 7, 2016

Abstract—This paper describes an approach for fusion of monocular vision measurements, camera motion, odometer and inertial rate sensor measurements. The motion of the camera between successive images generates a baseline for range computations by triangulation. The recursive estimation algorithm is based on extended Kalman filtering. The depth estimation accuracy is strongly affected by the mutual observer and feature point geometry, measurement accuracy of observer motion parameters and line of sight to a feature point. The simulation study investigates how the estimation accuracy is affected by the following parameters: linear and angular velocity measurement errors, camera noise, and observer path. These results impose requirements to the instrumentation and observation scenarios. It was found that under favorable conditions the error in distance estimation does not exceed 2% of the distance to a feature point.

DOI: 10.1134/S2075108717020043

INTRODUCTION

Reconstruction of three-dimensional geometry from two-dimensional images is one of the primary objectives of computer vision. The image sequence acquired by a moving monocular camera in a static environment contains detailed information about both the motion of the camera and the shape of the environment. This phenomenon is called motion parallax effect. Recovery of the structure of the scene using parallax effect is commonly referred to as structure-from-motion (SFM). Closely related to SFM is also relative position and depth estimation using monocular vision. Development of approaches to determine the relative position between a robot and objects in its environment is an active area of research. The purpose of relative position sensing includes obstacle avoidance, mapping (building a catalog of objects and their positions), localization (determining the robot position relative to mapped objects), and relative position control of the robot with the respect to an object to enable observation, modeling and manipulation.

The camera motion between successive images generates a baseline for the range computations by triangulation. However, estimating distance to the feature points is often quite challenging due to very small motion between frames, especially when a feature point is located close to the focus of expansion. Successful approaches usually integrate proprioceptive sensors to estimate the observer's ego-motion to produce a more robust sensing system than typical vision-only techniques. The fusion of a bearing measurement provided by a camera with ego-motion measurements leads to a nonlinear estimation problem, which can be solved by a nonlinear estimator like the EKF. Better results are obtained when a specific state representation is used that eliminates non-linearity in measurement model. These algorithms can be implemented on mobile robotic platforms equipped with inertial sensors, odometer and monocular cameras.

The proposed algorithm is motivated by the fact that humans can reliably estimate the scene structure without using binocular vision (with one eye only) by relying on motion parallax and on their vestibular system. In this case proprioception is supported by the brain, which uses information from the vestibular system in the head and motion sensing throughout the body to understand the body's translational and rotational kinematics. A similar approach inspired by nature can be applied to machine vision based on monocular camera and motion sensors. We present an algorithm that fuses information from a monocular camera with information from an IMU and an odometer to estimate the relative position between a camera mounted on a moving wheeled robot and a stationary landmark. This paper describes an extended Kalman filter (EKF) based algorithm that is uniquely adapted to this sensor fusion problem, and presents simulation results for different observation scenarios and instrumentation errors. The sensing strategy takes advantage of the complementary nature of monocular vision measurements, inertial rate sensor measurements, and camera motion.

RELATED WORK

Currently stereo vision is the main approach for the reconstruction of three-dimensional geometry from two-dimensional images [1, 2]. Stereo vision is usually implemented using a so-called stereo rig: two identical cameras with parallel optical axes that are separated by a known distance. The main difficulty in this approach is more in the hardware side: the stereo rig has to be rigid, the cameras must be identical, synchronized and have a wide viewing angle [3]. During fast vehicle motion no difference between capturing time is tolerated. In this approach the cameras have to be carefully calibrated by computing the rotation and the translation between the cameras. Another implementation problem is a point-by-point matching between the two images from the stereo setup to derive the disparity maps. This task is more difficult for stereo vision compared to a monocular camera because the corresponding point is not necessarily in the same location on image sensor in both images while in monocular vision the corresponding points between the successive images are almost in the same location provided that the frame-rate is high enough. Stereo vision can accurately recover the depth for near field objects, but the accuracy degrades with the distance and it becomes inaccurate for distant objects [4].

If two images relative to the same area of observation are acquired sequentially when the camera is moving the three-dimensional geometry can be also reconstructed. If the scene reconstruction is based on only one stereo pair at one time the data processing is similar to a stereo rig case from geometrical point of view [2]. However, the accuracy can be improved if a sequence of images is available for the scene reconstruction. The application of a sequence of stereo images assumes successive processing of a series of stereo pairs [5]. This approach can achieve better accuracy because of the potentially wide baseline and can be successfully used for computing the distance to remote objects.

In human vision the binocular disparity is the most important depth cue when the distance is less than 5 meters [4]. For larger distances the monocular depth cues, especially the motion parallax, play a more important role in sensing the depths. Perceptual psychologists have extensively studied motion parallax and have shown that it is of paramount importance for the spatial orientation of car drivers [6] and for landing an aircraft [7]. Regan et al. [8] described experiments in which pilots in the act of landing planes were deprived of the use of one eye. Their performance did not deteriorate significantly, so binocular cues cannot be important. Many authors agree with Gibson [7, 9] that the main cue in these cases is the so-called focus of expansion.

Wexler et al. [10] studied the importance of self-motion for perception of 3D structure from motion. They compared the non-moving, passive observer in an environment of moving rigid objects with the active observer moving in an environment of stationary objects. It was demonstrated that this substitution is not fully equivalent, because despite experiencing the same visual stimulus the active observers' perception of 3D structure depends on extra-visual self-motion information.

Longuet-Higgins et al. [11] used an equation that relates the coordinates of a texture element in the scene and its velocity with the observer's translational and angular velocities to derive an equation describing the kinematics of retinal position and velocity. They analyzed these equations and showed that from a monocular view of a rigid, textured surface it is possible to determine the motion of the eye relative to it from the velocity field of the changing retinal image. They also came to conclusion that the depth computed based on motion parallax is specified completely by the retinal velocity.

Motion parallax can be combined with other depth cues such as stereovision, kinetic depth effect, shading and occlusion to obtain a more stable estimate of viewing distance. Landy et al. [12] proposed to fuse motion parallax information with stereo disparity by minimizing the inconsistency between depth from disparity and depth from motion parallax.

Tkocz and Janschek [13] developed a method for the distance to landmark estimation that can be applied for an airborne camera mounted on UAV. In addition to the distance they also estimated the camera velocity. Their approach is based on EKF. The novelty was a batch algorithm for landmark initialization. Although the algorithm can track many points at a time the accuracy is not improved compared to individual landmark tracking. The algorithm was validated only by simulations under very optimistic assumptions.

Hustler and Rock [14, 15] proposed a system that fuses monocular vision with inertial measurement unit (IMU) measurements to estimate the distance between a moving observer and a stationary object. The approach is adapted to underwater robots, which have to operate in the presence of disturbances and uncertain dynamics. They observed that the combined IMU and vision sensing strategy is robust to vision drop-outs and is able to determine relative position with minimal requirements on the vision system. They also mentioned that the combination of limited observability and significant nonlinearities, which are inherent to this sensing strategy, creates an estimation problem that cannot be solved with a standard EKF. To overcome this difficulty they used a specific state representation that leads to a linear sensor model and transfers all of the nonlinearities into the state dynamics avoiding linearization in the measurement update. The state propagation is implemented with the unscented transform, which does not require linearization.

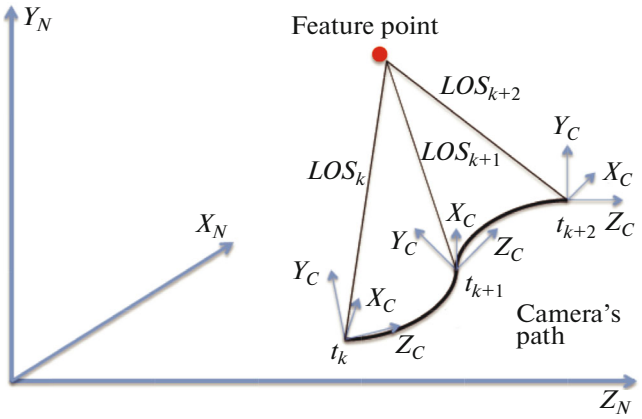

Fig. 1. Relative geometry of a camera (observer) and a feature point.

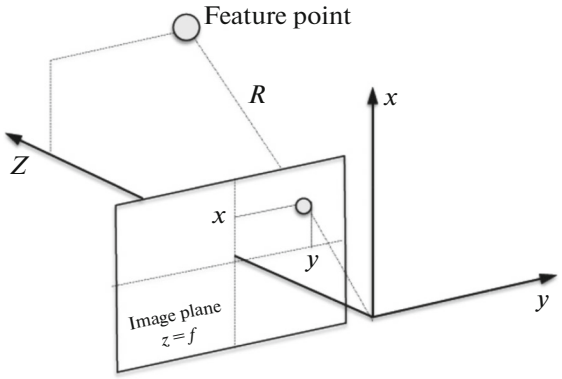

Fig. 2. Perspective projection camera model.

Our work is closely related to the relative position estimation approach proposed by Hustler and Rock [14, 15], and the algorithm for distance to landmark estimation proposed by Tkocz and Janschek [13]. The algorithm has the ability to determine relative position by tracking just one feature point. However, taking advantage of additional features, when available, should provide significant benefits, like improved accuracy and increased robustness. This multi-feature approach provides additional sensor measurements without adding additional hardware. It also adds flexibility to the sensing strategy, whereby a feature that is traveling out of the field of view of the camera can be augmented by a more appropriate feature before the first feature is lost.

RECURSIVE ESTIMATION OF DEPTH FROM SEQUENCE OF IMAGES

This section presents an approach for a depth (distance to a feature point) estimation from a sequence of images while tracking a feature point. The direction to the feature point is usually called Line-of-Sight (LOS) and it can be measured by a camera. The geometry of an observer and a feature point is shown in Fig. 1. The X_N, Y_N, and Z_N axes define the geographical coordinate system N. The X_C, Y_C, and Z_C axes define the camera frame C and the coordinate system of the camera where the Z_C axis corresponds to the camera's optical axis.

Many points can be tracked at the same time and distance to them can be estimated using independent estimators for each point. The accuracy of the estimate depends on the point observability: points which are close to the focus of expansion have poor observability and as a consequence, the distance to these points cannot be estimated very accurately. A cooperative estimation algorithm can be also considered in the future. It allows improving the estimation of points with poor observability by imposing additional constraints and thus estimation of distance to a certain point is now supported by other points.

The Estimation Algorithm

The estimator design is based on the Kalman Filter framework. We assumed a pinhole camera, which is represented by the standard perspective projection model shown in Fig. 2. The Cartesian components of the projection of a feature point on the image plane is denoted by x and y. The image plane is shown at $Z_C = f$. It is assumed, without loss of generality, that the camera measurements are scaled such that the effective focal length is 1. A feature point located at $R = [X_C, Y_C, Z_C]$ in the camera frame appears as an image plane feature at $x = X_C/Z_C$ and $y = Y_C/Z_C$. Therefore, the camera measurement z is the projection of a feature point onto the image plane, and is modeled as follows:

$$z = \begin{bmatrix} x \\ y \end{bmatrix} + n = \frac{1}{Z}\begin{bmatrix} X \\ Y \end{bmatrix} + n. \quad (1)$$

The camera measurement errors n are assumed to be zero-mean Gaussian noise. Feature tracking results in measurements of image location of the feature in subsequent frames. The kinematic relation between the feature point and its projection to the camera image plane is described by the following equations [11]

$$\begin{aligned} \dot{x} &= -v_x\zeta + xv_z\zeta + xy\omega_x - (1 + x^2)\omega_y + y\omega_z, \\ \dot{y} &= -v_y\zeta + yv_z\zeta + (1 + y^2)\omega_x - xy\omega_y - x\omega_z, \\ \dot{\zeta} &= v_z\zeta^2 + y\omega_x\zeta - x\omega_y\zeta. \end{aligned} \quad (2)$$

In these equations $\mathbf{v} = [v_x, v_y, v_z]^T$ is the observer translational velocity and $\boldsymbol{\omega} = [\omega_x, \omega_y, \omega_z]^T$ is the angular velocity. Both of these ego-motion parameters are assumed to be measured by the IMU and odometer with reasonably good accuracy. We used a specific representation of range, $\zeta = 1/Z$, which was introduced by Hustler and Rock in [14, 15]. This representation

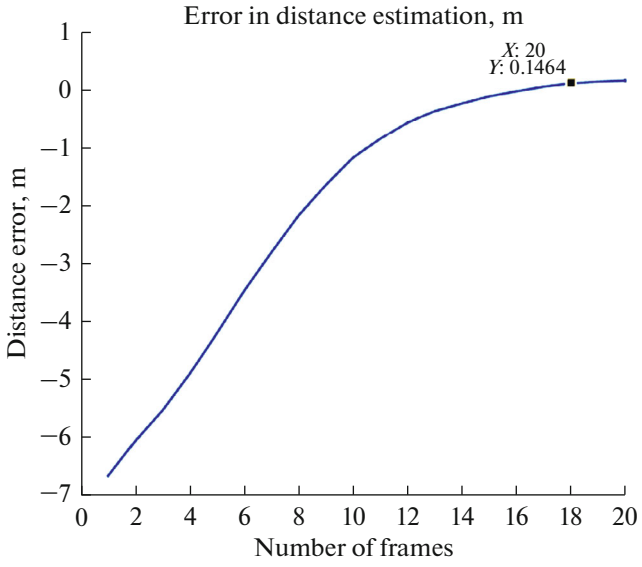

Fig. 3. Estimation error in distance to the feature point for a good observability without linear and angular velocity measurement errors. Camera measurement noise is 1 pixel (0.1 mm).

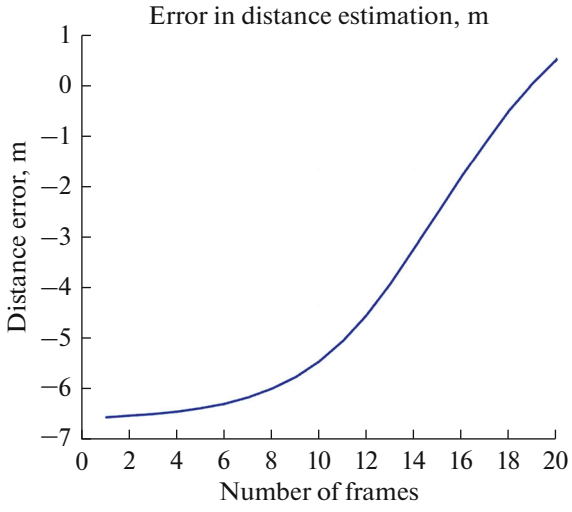

Fig. 4. Estimation error in distance to the feature point for a poor observability without linear and angular velocity measurement errors. Image noise is 1 pixel.

reduces the dominant nonlinearities in the state dynamics and results in more accurate estimator time-updates. The estimation algorithm is based on the extended Kalman filter (EKF) where for every feature point the measurements and the state dynamics are described by Eqs. 1 and 2 respectively. When several feature points are tracked several independent sets of equations have to be solved simultaneously. However, addition of new points does not improve the estimation accuracy if independent estimators for each point are used.

SIMULATION STUDY

The simulation results for estimating distance from the camera to a single point are presented below. The purpose of this simulation study is to understand how the estimation accuracy is affected by the following parameters: linear and angular velocity measurement errors, camera noise, and mutual observer and feature point geometry. In all cases it is assumed that the camera is moving forward with constant speed of 0.2 m/s during 20 s and covering a distance of 4 m. In most of the cases the feature point is located at $R = [5, 5, 10]^T$. In two cases with poor observability the feature points are located more closely to the focus of expansion at $R = [1, 1, 10]^T$ and at $R = [0.5, 0.5, 10]^T$ respectively. The estimation is based on a sequence of 20 images.

The Effect of Observer Trajectory

The distance estimation accuracy for the cases with good observability and without odometry and gyro errors is shown in Fig. 3. The only source of error is the random noise in estimated feature point coordinates on cameara's image plane. The error in distance estimation after 20 images is about 0.14 m. It takes about 15 s for the filter to converge.

The distance estimation accuracy for poor observability when a feature point is located close to the focus of expansion at $R = [1, 1, 10]^T$ is shown in Fig. 4. In this case the filter converges slowly and 20 s is not enough to obtain good accuracy. The error in distance estimation after 20 images is about 1.3 m.

The next example demonstrates the filter performance when the feature point is located very close to the focus of expansion ($R = [0.5, 0.5, 10]^T$). The angle between the camera's boresight and the LOS to a feature point is smaller than 5 deg. The estimation results are shown in Fig. 5. The convergence is very slow and the error in distance estimation after 20 images is about 2.1 m. From these results we can see that the camera's trajectory can make significant impact on the accuracy of distance estimation. Since this approach is based on motion parallax effect the accrued camera's path during estimation has to be at least 30% of the distance to a feature point and the angle between the camera's velocity and LOS to a feature point has to be larger than 10 deg. Currently only straight camera paths were considered. A curved path can improve the accuracy of distance estimation. The results are scalable and depth estimation accuracy depends on non-dimensional parameter similar to what was used in [16, 17]:

$$K = \frac{VT}{R_0}, \tag{3}$$

where V is average speed, T is observation time, R_0 is initial range to a feature point. The K parameter rep-

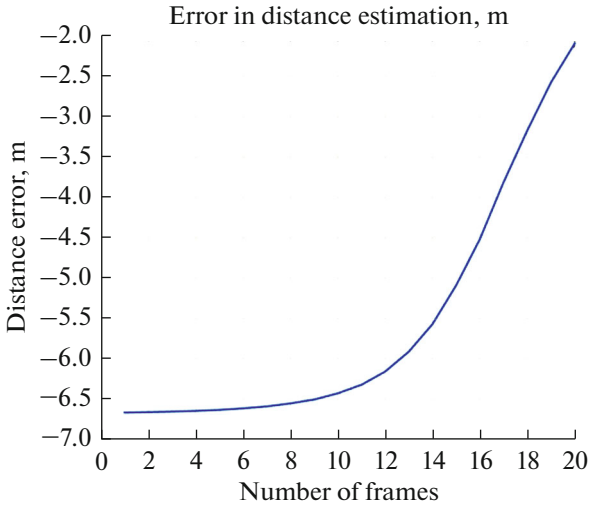

Fig. 5. Estimation error in distance to the feature point for a very poor observability when the feature point is very close to the focus of expansion. Image noise is 1 pixel.

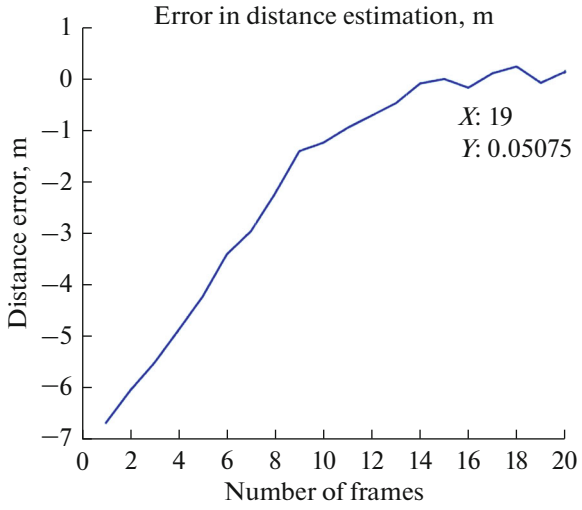

Fig. 6. Estimation error in distance to the feature point for a good observability without linear and angular velocity measurement errors. Image noise is 10 pixels.

resents the camera's (observer) ability to create maximum difference in the apparent position of a feature point viewed along two different lines of sights. Parameter K does not take into account the shape of the camera's path. So, for the same K the estimation accuracy will be different for different observation scenarios.

The Effect of Camera Noise

This parameter corresponds to camera measurement noise as well as random errors in feature point tracking. The camera's pixel size is 5.5 μm, therefore the additive Gaussian noise with standard deviation of 100 μm (0.1 mm) in Eq. 1 will be a good representation of this type of error in nominal case. In addition to this two other cases of very large measurement noise of 1 mm and 10 mm are considered to find out when the estimation starts to diverge. To investigate the effect of camera noise on estimation accuracy we can assume good observability without linear and angular velocity measurement errors. The nominal case is shown in Fig. 1. The depth estimation accuracy for 10 times larger camera noise of 1mm is shown in Fig. 6. The error in distance estimation after 20 images is about the same as in the case with nominal measurement noise.

When the camera noise is increased to 10 mm (not possible in practical applications) to show the effect of very large noise, the filter still can converge to a true distance as is shown in Fig. 7. These results show that the additive measurement noise in feature point coordinates estimation on the image plane does not have significant impact on the estimation accuracy.

The Effect of Angular Velocity Measurement Errors

The measurement of camera's angular velocity is required to calculate coordinates of a feature point projection on image plane as is described by Eqs. 2. This type of error can be caused by the measurement errors in angular velocity as well as unaccounted error in vehicle's orientation tracking and LOS trembling. These errors can be caused, for example, by vehicle's vibrations when IMU sampling rate is not sufficient to accurately track the vehicle's movement. This type of error can be modeled as additive systematic or random error to the nominal angular velocity.

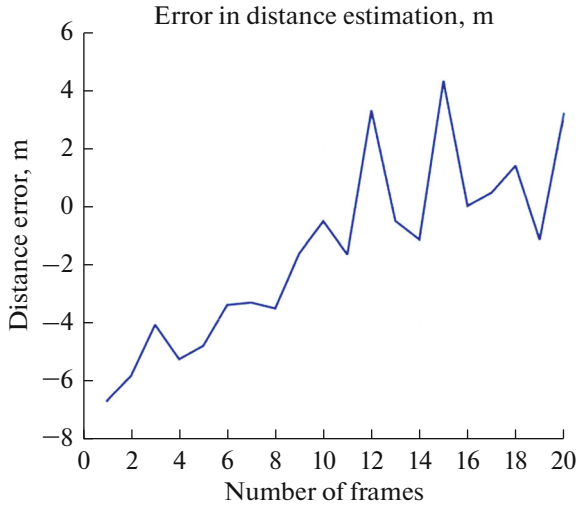

Fig. 7. Estimation error in distance to the feature point for a good observability without linear and angular velocity measurement errors. Image noise is 100 pixels.

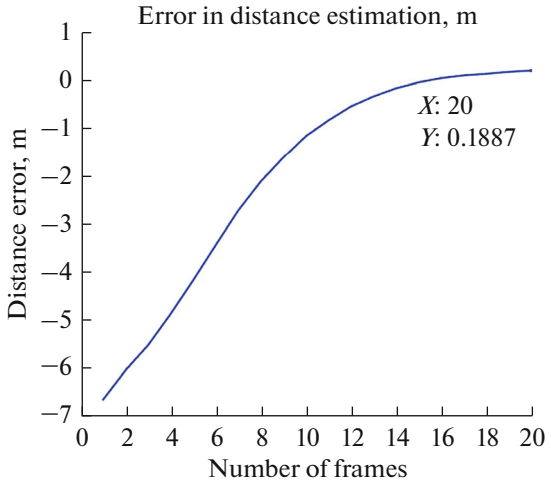

Fig. 8. Estimation error in distance to the feature point for a good observability without linear velocity measurement error. Gyro bias is 10 deg/hr. Image noise is 1 pixel.

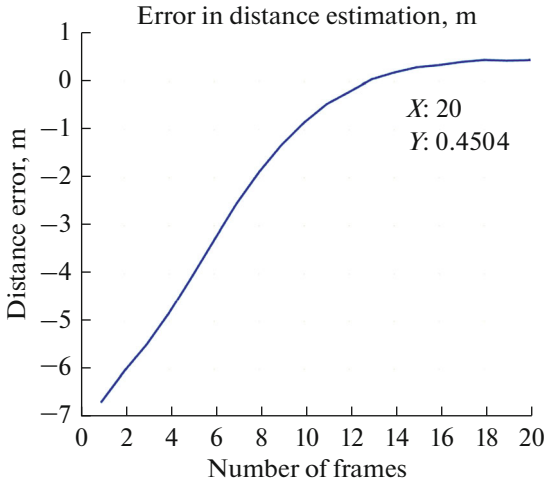

Fig. 9. Estimation error in distance to the feature point for a good observability without linear velocity measurement error. Gyro bias is 100 deg/hr. Image noise is 1 pixel.

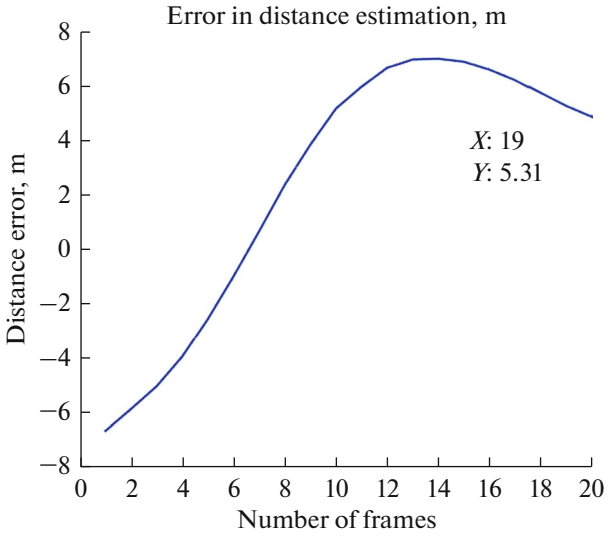

Fig. 10. Estimation error in distance to the feature point for a good observability without linear velocity measurement error. Gyro bias is 1000 deg/hr. Image noise is 1 pixel.

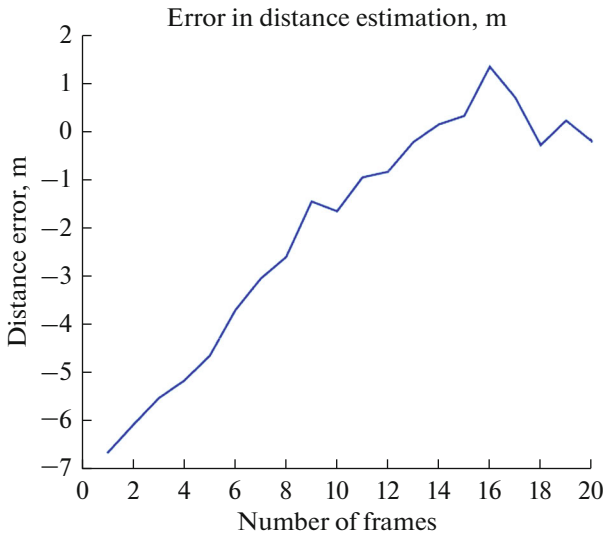

Fig. 11. Estimation error in distance to the feature point for a good observability without linear velocity measurement error. Angular velocity measurement error is random noise with standard deviation 1000 deg/hr. Image noise is 1 pixel.

To investigate the effect of angular velocity measurement errors on estimation accuracy we can assume good observability without linear velocity measurement error. The following cases will be considered: constant bias in angular velocity of 10 deg/hr, 100 deg/hr and 1000 deg/hr, and random error additive error in angular velocity measurement with the standard deviation of 1000 deg/hr. The depth estimation accuracy for 10 deg/hr bias is shown in Fig. 8. The error in distance estimation after 20 images is about 0.18 m, therefore the accuracy degradation because of small angular velocity is small. The estimation accuracy for the 100 deg/hr gyro bias is shown in Fig. 9. The error in distance estimation after 20 images is about 0.45 m. The estimation accuracy for the 1000 deg/hr gyro bias is shown in Fig. 10. In this case the filter did not converge to the true distance. The error in distance estimation after 20 images is about 5 m. This measurement error is well above the limit that can be tolerated by the filter. The estimation accuracy for the 1000 deg/hr random error in angular velocity is shown in Fig. 11. The distance estimation is noisy, but the filter converges to the true distance. The

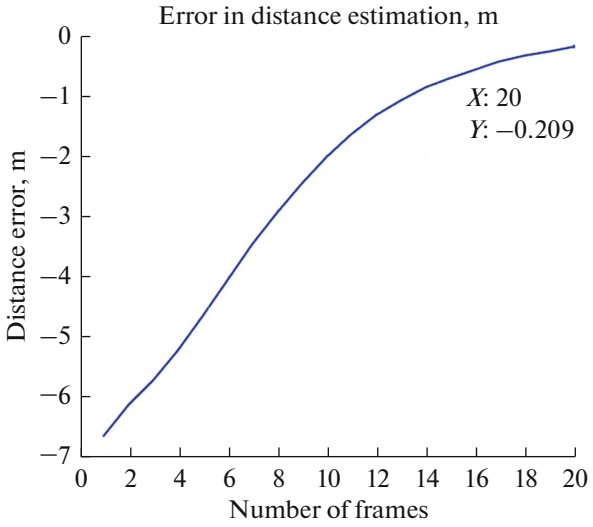

Fig. 12. Estimation error in distance to the feature point for a good observability without angular velocity measurement error. Velocity bias is 0:02 m/s Image noise is 1 pixel.

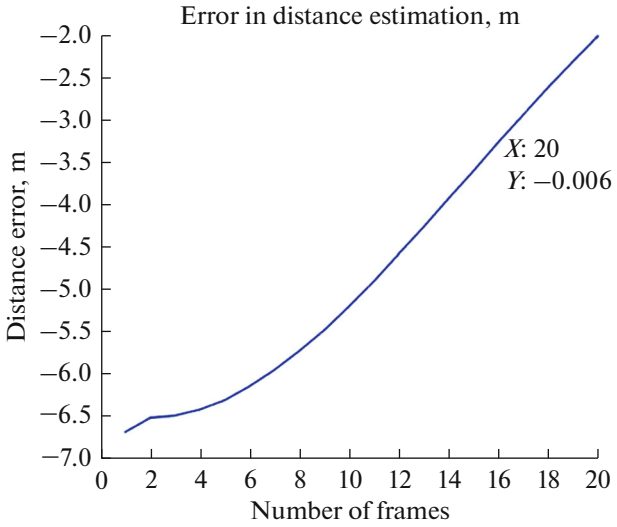

Fig. 13. Estimation error in distance to the feature point for a good observability without angular velocity measurement error. Velocity bias is 0.1 m/s. Image noise is 1 pixel.

error in distance estimation after 20 images is about less than 1 m.

The Effect of Linear Velocity Measurement Errors

To investigate the effect of linear velocity measurement errors on estimation accuracy we can assume good observability without angular velocity measurement error. The following cases will be considered: constant bias in linear velocity of 0.02 m/s and 0.1 m/s. The depth estimation accuracy for 0.02 m/s velocity bias is shown in Fig. 12. The error in distance estimation after 20 images is about 0.2 m. So, no significant accuracy degradation was observed. The estimation accuracy for the 0.1 m/s velocity bias is shown in Fig. 13. The error in distance estimation after 20 images is about 2 m. This velocity measurement error is too large for obtaining good distance estimation accuracy.

FUTURE WORK

We are planning to implement the presented algorithms in the two different mobile robots shown in Fig. 14. The first is a robotic manipulator arm mounted on a mobile base and the second is a commercially available Robotnik mobile platform. Both of these robots are equipped with the Linux Ubuntu computer with robot operating system (ROS), Microstrain 3DM-GX3-25 inertial measurement unit and odometer. These mobile platforms are also equipped with machine vision cameras that are connected to the computer with USB cables. The instrumentation includes a triaxial accelerometer, triaxial gyro, triaxial magnetometer, temperature sensors, and an on-board processor that fuses the measurements to provide static and dynamic orientation and translation measurements. The range of the attitude and heading in IMU unit is 360 deg in all three axes and the sampling rate is 100 Hz. The angular attitude accuracy of the IMU is 0.5 deg in static conditions and 2.0 deg in dynamic conditions. The Renishaw RM22I odometer has the resolution of 0.7 deg and the sampling rate of 40 Hz. The machine vision cameras are manufactured by Point Grey and have the resolution 4 Mpixels and the imaging rate 10 Hz. The camera is mounted on holders with servomotors and can be tilted up to 90 deg and panned up to ±90 deg. The field of view of cameras is 48 deg.

Feature Points Detection

Real environments present the typical challenges of identifying good visual features, establishing feature correspondences and robust tracking. The feature-tracking algorithm tracks a set of points using the Kanade–Lucas–Tomasi algorithm. The image captured with camera describes the intensity of light, which is reflected from the surface of objects. The intensity can rapidly change in the image for example because of the color changes or the orientation changes of the object. If the intensity changes at the same time in two directions such spot is called a corner in image. There exist a large variety of corner detection algorithms [18, 19] for instance. In this work the algorithm described in [20] was applied. In such method the corner detection is based on the gradient covariance matrix, which is computed from the area of 25 pixels (5 × 5) pixels. Such covariance matrix is computed for each pixel location in the image and the minimal eigenvalue of matrix is computed which is the quality measure of the corner. Next the local maxi-

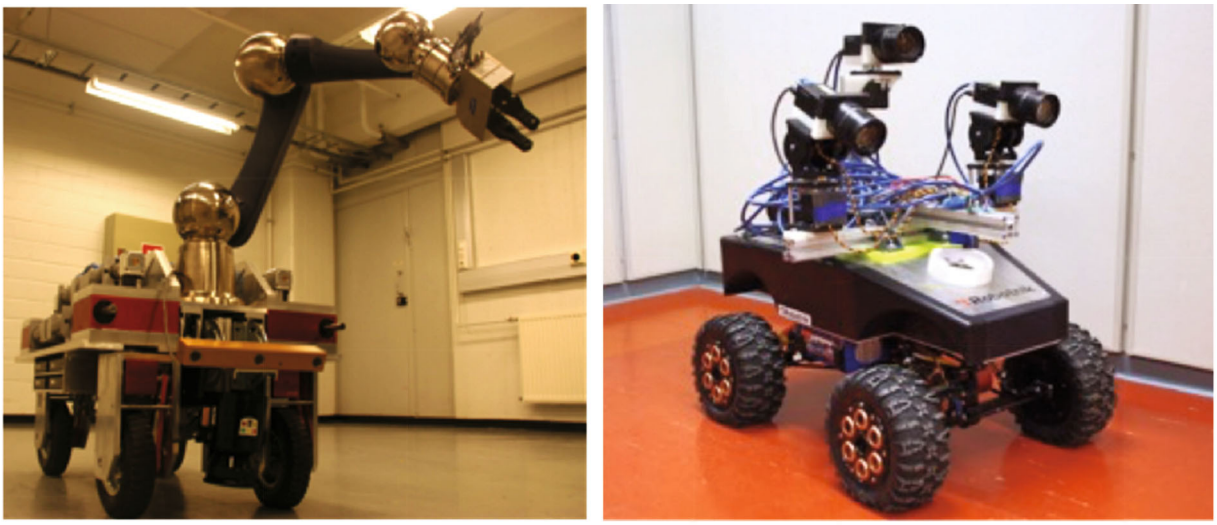

Fig. 14. Robotic manipulator arm mounted on a mobile base (left) and commercially available Robotnik mobile platform (right).

mums from the 3 × 3 neighborhoods are obtained. Finally the corners whose minimal eigenvalue is less than the product between the quality measure of best corner and the value 0.01 are rejected. Furthermore, the algorithm rejects corners for which there is a stronger corner at a distance less than 2 pixels. These two threshold limits in this algorithm are adjusted such that at least 500 feature points are found from each image.

In this work the detected feature points are not associated with any specific object and the current goal is to obtain an image where each chosen feature point contains information about the depth or distance to the camera. In our future work we are planning to develop approaches for vanishing points identification and data association. The latter can help to identify specific objects including moving ones and build a map of the environment. The vanishing points correspond to distant landmarks and therefore their projections to the image plane are not moving unless the camera is rotating.

Tracking Feature Points

Feature points can be found from each image frames separately. However, the location of feature points differs from one frame to another and therefore the features between the consecutive frames must be matched by their tracking over the image sequence. Tracking method is based on the optical flow algorithm, which uses the iterative Lucas–Kanade method with image pyramids [21, 22]. The Lucas–Kanade method assumes that the displacement of features (optical flow vector) between the two consecutive images is small. This condition was fulfilled in our tests because of low speed of the ground robot and high imaging rate. The optical flow is defined as a vector $\mathbf{d} = [d_x, d_y]$ that minimizes the residual function defined as:

$$e(d_x, d_y) = \sum_{x=u_x-\omega_x}^{u_x+\omega_x} \sum_{y=u_y-\omega_y}^{u_y+\omega_y} (I(x,y) - J(x+d_x, y+d_y))^2, \quad (4)$$

where x and y are the pixel locations in image plane, I and J are the intensity functions of consecutive images, u_x and u_y the coordinates of the image point, and ω_x and ω_y determine the size of the search area. This minimization is done for each level in image pyramid. An image pyramid is a collection of images that all are successively down-sampled from the original image until some desired stopping point is reached. In this work we used three pyramids levels. The minimization starts from the smallest image and proceeds to middle image and finally ends to the original image. Such pyramid procedure speeds up the minimization and helps to match features from different distances.

CONCLUSIONS

This paper describes the algorithm for fusion of monocular vision measurements, inertial rate sensor measurements, and camera motion. The outcome of this approach is a relative distance between a camera and objects in its environment and it can be implemented in many important practical applications such as obstacle avoidance, mapping, localization of the robot position relative to mapped objects, and relative position control. The simulation study shows that this approach can be very accurate in depth (distance to objects) estimation with estimation error as low as 2% of the distance to the object. However, the distance estimation accuracy strongly depends on the accuracy of gyroscopes and odometer as well as observation sce-

narios. Based on the simulation study the accuracy requirements to the instrumentation and observation scenarios have been established.

It was shown that this approach could tolerate quite large (up to 100 deg/hr) angular velocity measurement errors. To obtain good results medium accuracy MEMS IMUs such as Microstrain 3DM-GX3-25 and Xsens MTi-10 are required. Accurate velocity measurement is also important since it is used to compute the baseline between successive frames. Since velocity is scaled to the distance to landmark the velocity measurement error also should be scaled accordingly and given in terms of relative error. Based on the simulation study it was concluded that relative speed measurement errors as large as 5% of camera speed can be tolerated. This is not a very restrictive requirement, because typical wheel encoders have better accuracy. However, wheel slippage also has to be taken into account.

Mutual camera and feature point geometry is also very important. For obvious reasons this approach cannot compute distance to a feature point located in the focus of expansion (FOE). For practical reasons distance estimation to feature points near FOE is not accurate. The simulation study showed that there is no significant degradation in accuracy of the distance estimation when the angle between feature points and the FOE is larger than 10 deg. However, accuracy starts to deteriorate if a feature point is getting closer to the FOE, and for feature points located at the angle of less than 5 deg to the FOE the method does not produce reliable depth estimation. If the application allows some freedom in observer path then the accuracy can be improved by moving to create a geometry between a feature point and an observer that maximizes system observability.

ACKNOWLEDGMENTS

The research leading to these results has been funded by the Finnish Academy of Science.

REFERENCES

1. Lazaros, N., Sirakoulis, G.C., and Gasteratos, A., Review of stereo vision algorithms: from software to hardware, *International Journal of Optomechatronics*, 2008, no. 2(4), pp. 435–462.
2. Hartley, R. and Zisserman, A., *Multiple View Geometry in Computer Vision*, Cambridge university press, 2003.
3. Vishnyakov, B.V., Vizilter, Y.V., Knyaz, V.A., Malin, I.K., Vygolov, O.V. and Zheltov, S.Y., Stereo sequences analysis for dynamic scene understanding in a driver assistance system, in *SPIE Optical Metrology* 2015, International Society for Optics and Photonics.
4. Kytö, M., Nuutinen, M. and Oittinen, P., Method for measuring stereo camera depth accuracy based on stereoscopic vision, in *IS&T/SPIE Electronic Imaging 2011*, International Society for Optics and Photonics.
5. Beloglazov, I.N., Accuracy of stereoscopic navigation system in flights above natural landscapes and urban land, *Journal of Computer and Systems Sciences International,* 2010, no. 49.5, pp. 802–810.
6. Gordon, D.A., Static and dynamic visual fields in human space perception, *JOSA*, 1965, vol. 55, no. 10, pp. 1296–1302.
7. Gibson, J.J., *The Ecological Approach to Visual Perception*, Psychology Press, 2013.
8. Regan, D., Beverley, K., and Cynader, M., *The Visual Perception of Motion in Depth, Scientific American*, 1979.
9. Gibson, J.J., *The Perception of the Visual World*, 1950.
10. Wexler, M., Panerai, F., Lamouret, I., and Droulez, J., Self-motion and the perception of stationary objects, *Nature*, 2001, no. 409(6816), pp. 85–88.
11. Longuet-Higgins, H.C. and Prazdny, K., The interpretation of a moving retinal image, *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 1980, vol. 208, no. 1173, pp. 385–397.
12. Landy, M.S. et al., Measurement and modeling of depth cue combination: in defense of weak fusion, *Vision research,* 1995, no. 35.3, pp. 389–412.
13. Tkocz, M. and Janschek, K., Metric velocity and landmark distance estimation utilizing monocular camera images and IMU data, in *Proc. 11th Workshop on Positioning, Navigation and Communication,* IEEE, 2014.
14. Huster, A., Relative position sensing by fusing monocular vision and inertial rate sensors, *Ph.D. dissertation*, Citeseer, 2003.
15. Huster, A. and Rock, S.M., Relative position estimation for manipulation tasks by fusing vision and inertial measurements, in *Proc. 11th International Conference on Advanced Robotics*, Coimbra, Portugal, June 30–July 3 2003, vol. 2. ICAR, 2003, pp. 1562–1567.
16. Hammel, S., Liu, P., Hilliard, E., and Gong, K., Optimal observer motion for localization with bearing measurements, *Computers & Mathematics with Applications*, 1989, vol. 18, no. 1, pp. 171-180.
17. Oshman, Y. and Davidson, P., Optimization of observer trajectories for bearings-only target localization, *Aerospace and Electronic Systems, IEEE Transactions on*, 1999, vol. 35, no. 3, pp. 892–902.
18. Harris, C. and Stephens, M., A combined corner and edge detector, in *Proc. of the 4th Alvey Vision Conference,* 1988, pp. 147–151.
19. Moravec, H., *Obstacle Avoidance and Navigation in the Real World by a Seeing Robot Rover*, Tech Report CMU-RI-TR-3 Carnegie-Mellon University, Robotics Institute, 1980.
20. Shi, J. and Tomasi, C., Good features to track, in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 1994, pp. 593–600.
21. Lucas, B.D. and Kanade, T., An iterative image registration technique with an application to stereo vision, in *Proc. of Imaging Understanding Workshop,* 1981, pp. 121–130.
22. Bouguet, J.-Y., *Pyramidal implantation of the Lucas-Kanade Feature Tracker Description of the algorithm*, Intel Corporation, Microprocessor Research Labs, 1999.