Sensor Data Fusion using Unscented Kalman Filter for VOR-based Vision Tracking System for Mobile Robots

PRE-PRINT VERSION

Sensor Data Fusion using Unscented Kalman Filter for VOR-based Vision Tracking System for Mobile Robots

Muhammad Latif Anjum[1], Omar Ahmad[1], Basilio Bona[2], Dong-il "Dan" Cho[3]

[1]Department of Mechanical and Aerospace Engineering (DIMEAS), [2]Department of Control and Computer Engineering, (DAUIN), Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129, Torino, Italy
[1,2]{muhammad.anjum, omar.ahmad, basilio.bona}@polito.it
[3]Department of Electrical Engineering and Computer Science/ASRI/ISRC, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 151-742, Korea
[3]dicho@snu.ac.kr

Abstract. This paper presents sensor data fusion using Unscented Kalman Filter (UKF) to implement high performance vestibulo-ocular reflex (VOR) based vision tracking system for mobile robots. Information from various sensors is required to be integrated using an efficient sensor fusion algorithm to achieve a continuous and robust vision tracking system. We use data from low cost accelerometer, gyroscope, and encoders to calculate robot motion information. The Unscented Kalman Filter is used as an efficient sensor fusion algorithm. The UKF is an advanced filtering technique which outperforms widely used Extended Kalman Filter (EKF) in many applications. The system is able to compensate for the slip errors by switching between two different UKF models built for slip and no-slip cases. Since the accelerometer error accumulates with time because of the double integration, the system uses accelerometer data only for the slip case UKF model. Using sensor fusion by UKF, the position and orientation of the robot is estimated and is used to rotate the camera mounted on top of the robot towards a fixed target. This concept is derived from the vestibulo-ocular reflex (VOR) of the human eye. The experimental results show that the system is able to track the fixed target in various robot motion scenarios including the scenario when an intentional slip is generated during robot navigation.

Keywords: Sensor Fusion, Unscented Kalman Filter, VOR, Vision Tracking Systems

1 Introduction

Applications of robotics in various fields of science and engineering are increasing. Robots need a set of accurate and reliable sensors and control techniques to perform various tasks assigned to them. Accurate vision systems are especially crucial for robots involved in target tracking, visualization and vision based decision making. Accurate localization of mobile robots is an essential requirement for mobile robots navigation and target tracking. Various sensors have been utilized to achieve an accurate positioning system for mobile robots. The system proposed in this paper estimates the robot motion information and further uses this information to implement vision tracking system by rotating the camera mounted on the mobile robot towards the target. This concept is based on the Vestibulo-Ocular Reflex (VOR) of the human eye. The system can be divided into two blocks. Block 1 estimates robot motion information using sensor data fusion and block 2 uses this information to rotate the camera mounted on the robot to track the fixed target.

Shim, E.S. *et al* presented a stable vision system for mobile robots using encoder data [2]. It detects a target within a few meters and maintains it fixed in the center of the image frame during locomotion. The proposed system uses encoder data to calculate robot motion and periodically uses vision sensor signals to compensate for the errors. Their system is highly dependent on vision sensor information in the slip case and therefore performance deteriorates.

Jaehong Park *et al* developed a high performance vision tracking system for mobile robots using sensor data fusion via Kalman filter [1]. The robot motion information is computed by low cost accelerometer data, gyroscope data, and encoder data. The vision information is obtained by camera images of the object on locomotion during vision tracking. Researchers have also used fuzzy controller [10] and double Kalman filters [11] to obtain vision tracking systems for mobile robots.

We are using UKF which is an advanced filtering technique as compared with EKF. Furthermore, we are no longer using vision sensor (camera) image information which makes the system computationally heavy causing problems for tracking the target in continuous real time environment. UKF is much more efficient to deal with systems with severe nonlinearities because it is not dependent on first order linearization. The concept of unscented Kalman filter, also known as Sigma Point Kalman Filter (SPKF), has been studied by many researchers. Julier and Uhlmann compared the performance of this new filter (UKF) with the previously used EKF [6]. They argued that because EKF is based on the first order linearization of the nonlinear systems, its performance deteriorates especially when the systems are highly nonlinear. They introduced this new filter which is based on selecting a deterministic set of points, called the sigma points. These sigma points capture the true mean and covariance of the system when propagated through the nonlinear systems.

2 The Vision Tracking System

The vision tracking system implemented in this paper is inspired by Vestibulo-Ocular Reflex (VOR) of the human eye. The vestibulo-ocular reflex is the eye reflex movement that stabilizes images on the retina during head movement, by producing an eye movement in the direction opposite to head movement, thus preserving the image on the center of the visual field. For example, when the head moves to the right, the eyes move to the left, and vice versa. Since slight head movement is present all the time, the VOR is very important for stabilizing vision. Patients whose VOR is impaired find it difficult to read, because they cannot stabilize the eyes during small head tremors. The VOR does not depend on visual input and works even in total darkness or when the eyes are completely closed [4].

The VOR has both rotational and translational aspects. When the head rotates about any axis (horizontal, vertical, or torsional), distant visual images are stabilized by rotating the eyes about the same axis, but in the opposite direction. When the head translates, for example during walking, the visual fixation point is maintained by rotating the gaze in the opposite direction, based on the translational distance covered by the head.

The vestibulo-ocular reflex needs to be fast. For clear vision, head movement must be compensated almost immediately; otherwise, vision corresponds to a photograph taken with a shaky hand. To achieve clear vision, signals from the semicircular canals are sent as directly as possible to the eye muscles. The connection involves only three neurons, and is correspondingly called the three neuron arc [4]. Using these direct connections, eye movement lags the head movement by less than 10 ms, and thus the vestibulo-ocular reflex is one of the fastest reflexes in the human body.

The system proposed in this paper uses data from various sensors onboard the robot to detect the robot motion, and then rotates the camera based on the robot motion information. The whole concept can be partitioned into two blocks:

- Robot motion information using sensor fusion by UKF
- Camera rotation in opposite direction of the robot motion (VOR based concept)

Low cost inertial sensors (MEMS based gyroscope and accelerometer) are used along with robot wheel encoders to implement the robot motion information algorithm. The motion information obtained from the motion information block is used to rotate the camera towards the target based on the VOR concept of the human eye.

3 The Mathematical Modeling

The robot position and orientation can be estimated based on the encoder measurements using the following state model. The linear and angular velocities v, w can be obtained from differential drive kinematics.

$$p_k = \begin{bmatrix} x_k \\ y_k \\ \theta_k \end{bmatrix} = \begin{bmatrix} x_{k-1} + v_{k-1}\cos(\theta_{k-1})\Delta t \\ y_{k-1} + v_{k-1}\sin(\theta_{k-1})\Delta t \\ \theta_{k-1} + w_{k-1}\Delta t \end{bmatrix} \tag{1}$$

Odometry is based on the assumption that wheel revolutions can be translated into linear displacement relative to the floor. This assumption is only of limited validity. One extreme example is wheel slippage: if one wheel was to slip on, say, an oil spill, then the associated encoder would register wheel revolutions even though these revolutions would not correspond to a linear displacement of the wheel. To avoid this error, our proposed system does not use encoder data in the case of slippage.

Gyroscopes measure the angular velocities which can be integrated to give the orientation. Gyroscopes have bias and drift errors which should be properly tackled to avoid the unbounded accumulation of errors in position and orientation [3]. The data from gyroscope can be used to calculate robot orientation as follows.

$$\theta_k = \theta_{k-1} + w_k \Delta t \tag{2}$$

The accelerometer measures the linear acceleration of the robot which can be integrated to give velocity and position for mobile robots. Accelerometer data is very noisy because it naturally incorporates the gravity vector. Linear position estimation with information from accelerometers is more susceptible to errors due to the double integration process. The following equations are used to obtain linear position from accelerometer data.

$$\begin{aligned} v_{k+1} &= v_k + a_k \Delta t \\ x_{k+1} &= x_k + v_k \Delta t \end{aligned} \tag{3}$$

We have used a two-wheeled robot with 3 degrees of freedom containing motion in *x/y* axis and rotation in the *z*-axis. Two coordinate frames are used to model the robot navigation; the earth coordinate frame (X,Y,Z), and the robot coordinate frame (x,y,z) as shown in fig. 1.

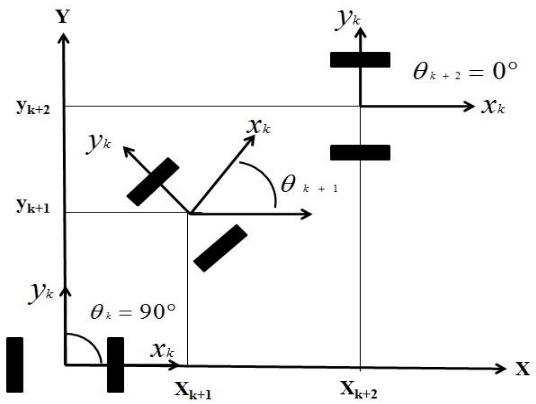

Fig. 1. The coordinate system used for robot navigation

4 The Unscented Kalman Filter

The unscented Kalman filter (UKF) is a recursive minimum mean square estimator (MMSE) based on the optimal Gaussian approximate Kalman filter framework that addresses some of the approximation issues of the EKF. Unlike the EKF, the UKF does not explicitly approximate the nonlinear process and observation models; the state distribution is still represented by a Gaussian random variable (GRV), but it is specified using a minimal set of deterministically chosen sample points called the sigma points. These sample points completely capture the true mean and covariance of the GRV, and when propagated through the true nonlinear system, captures the posterior mean and covariance accurately to the 2nd order for any nonlinearity, with errors only introduced in the 3rd and higher orders [6].

An unscented transformation is based on two fundamental principles. First, it is easy to perform a nonlinear transformation on a single point (rather than an entire *pdf*). Second, it is not too hard to find a set of individual points (sigma points) in state space whose sample *pdf* approximates the true *pdf* of a state vector. These sigma points $x^{(i)}$ can be calculated from the mean $\bar{x}$ and a deviation from the mean $\chi^{(i)}$ obtained from the square root decomposition of covariance matrix P. For an *n*-element state vector x, the sigma points can be calculated as follows:

$$
\begin{aligned}
x^{(i)} &= \bar{x} + \chi^{(i)} && i = 1, \ldots, 2n \\
\chi^{(i)} &= \left(\sqrt{nP}\right)_i^T && i = 1, \ldots, n \\
\chi^{(n+i)} &= -\left(\sqrt{nP}\right)_i^T && i = 1, \ldots, n
\end{aligned}
\tag{4}
$$

where $\sqrt{nP}$ is the matrix square root of nP such that $(\sqrt{nP})^T(\sqrt{nP}) = nP$, $(\sqrt{nP})_i$ is the ith row of $\sqrt{nP}$, $\bar{x}$ is mean and P is the error covariance of the state vector x. These sigma points when propagated through the nonlinear equation capture the true mean and covariance of the random variable. The mathematical details of UKF algorithm have been intentionally skipped and readers are referred to [5-8] for more details.

4.1 The UKF State Model

The state model for the UKF is formulated based on the error states, i.e., the difference between the encoder and gyroscope measurement is used as state variable for the UKF. The 3-element state vector consists of position error states e_x , e_y , e_θ . This can be obtained by calculating the position of the robot based on gyroscope and encoders separately and then taking the difference of the two values. The results obtained from the UKF (error states) are then added to the encoder based position values to exactly find the position of the robot. The state vector x_k therefore serves as an error compensator for encoder data. The state model of UKF can be mathematically put as:

$$
x_k = \begin{bmatrix} e_{x,k} \\ e_{y,k} \\ e_{\theta,k} \end{bmatrix} = \begin{bmatrix} e_{x,k-1} + v_{k-1}(\cos\theta_{g,k-1} - \cos\theta_{e,k-1})\Delta t \\ e_{y,k-1} + v_{k-1}(\sin\theta_{g,k-1} - \sin\theta_{e.k-1})\Delta t \\ e_{\theta,k-1} + (w_{g,k-1} - w_{e,k-1})\Delta t \end{bmatrix} + w_{k-1} \tag{5}
$$

The subscript ‘g’ and ‘e’ refer to the data from gyro and encoder, and w is the zero mean process noise with covariance Q_k. The error state ‘$e_{x/y/\theta}$’ is defined as the difference in gyro and encoder measurement. This can be mathematically put as:

$$x_k = \begin{pmatrix} e_{x,k} \\ e_{y,k} \\ e_{\theta,k} \end{pmatrix} = \begin{pmatrix} p_{x,gyro} - p_{x,enc} \\ p_{y,gyro} - p_{y,enc} \\ \theta_{gyro} - \theta_{enc} \end{pmatrix} \tag{6}$$

In the case of slip detection, the encoder data is no more reliable, so a different formulation of the above state vector is used. When slip is detected, we replace the encoder data with the accelerometer data in the above equation.

4.2 The UKF Measurement Model

The measurement model is based on the difference of the orientation of the robot measured by the gyroscope and encoders. This model can be mathematically written as follows:

$$z_k = e_{\theta,k} + v_k = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} e_{x,k} \\ e_{y,k} \\ e_{\theta,k} \end{bmatrix} + v_k \tag{7}$$

$$e_{\theta,k} = \left[\theta_{gyro} - \theta_{enc}\right]$$

where v_k is the zero mean measurement noise with covariance R_k. As in the case of state model, the measurement model is also changed when slip occurs. Here again, the encoder data is replaced with the accelerometer data in case of slip.

Finding the UKF noise parameters (Q, R) is a tedious job and requires repeated experiments with the model and sensors. The diagonal elements of the measurement noise covariance matrix represent the square of the standard deviation of the error in corresponding parameters. So the measurement noise covariance matrix can be found by experimenting several times with the robot encoders, gyroscope, and accelerometer to calculate the square of the standard deviation of the error. Finding the process noise covariance is more subtle. Often times the best method of estimating Q is by tuning of filter i.e. adjusting the value of Q to obtain the optimal state [12].

5 The VOR Modeling

After the execution of the first block, i.e. the robot motion information block, we now have the robot position information with respect to the previous position. The figure 2 shows the coordinate system used to calculate the camera rotation angle.

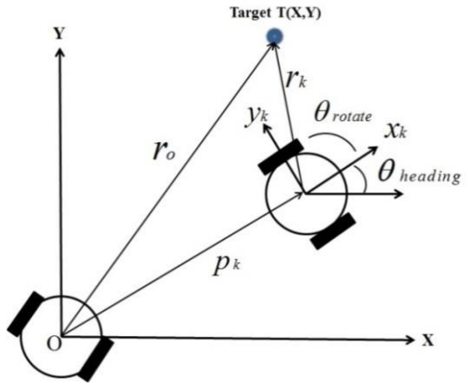

Fig. 2. Coordinate frame for calculating camera rotation angle

The camera rotation angle can be calculated from the following equation

$$\theta_{rotate} = \tan^{-1}\left(\frac{r_{y,k}}{r_{x,k}}\right) = \tan^{-1}\left(\frac{-(X - p_{x,k})\sin\theta_{heading} + (Y - p_{y,k})\cos\theta_{heading}}{(X - p_{x,k})\cos\theta_{heading} + (Y - p_{y,k})\sin\theta_{heading}}\right) \tag{8}$$

The DC motor is used to rotate the camera towards the target based on the voltage generated corresponding to the θ_{rotate} .

6 The Slip Detector

Our system has two different models for slip and no-slip cases. The slip therefore must be detected beforehand. The data from accelerometer and encoders is first compared to ascertain the occurrence of slip. The occurrence of slip is confirmed if the difference of encoder and accelerometer position is greater than a threshold value k . The threshold value can be adjusted as desired and should also accommodate the accelerometer errors. Based on the output of this slip detector, either of the two UKF models is used.

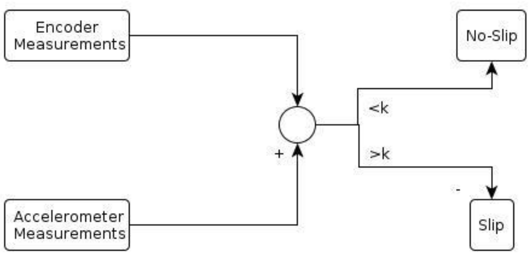

Fig. 3. The slip detector

7 The Complete System

The complete system consists of two UKF models incorporating data from three sensors i.e. gyroscope, accelerometer and encoders. The two UKF models are labeled as the slip UKF model and no-slip UKF model. Based on the output of this slip detector, either of the two UKF models is used. The slip-UKF model integrates data from inertial sensors only while the no-slip UKF model integrates data from encoders and gyroscope. The transition between the UKF models will be automatic based on whether slip occurs or not. The overall system is shown in the block diagram below.

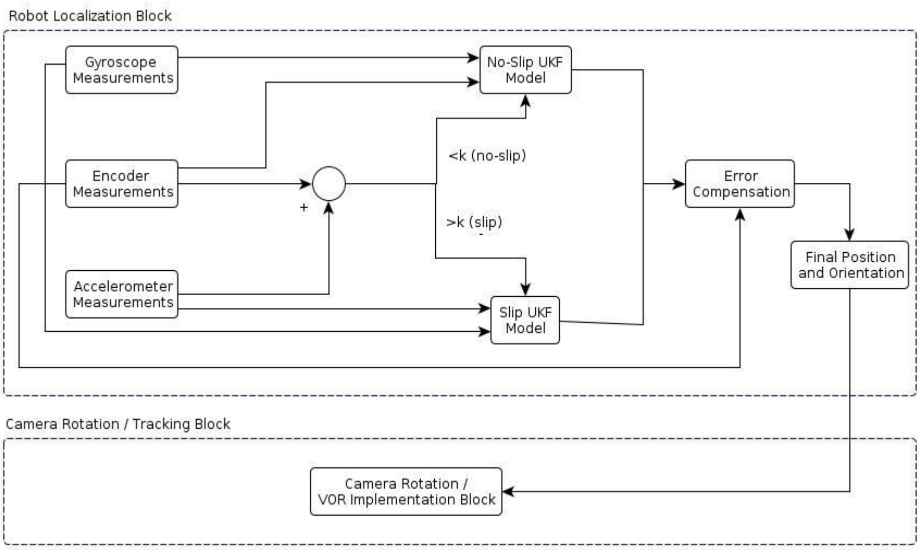

Fig. 4. The complete system including the slip detector. The constant value, k, is chosen to accommodate the error in accelerometer data.

8 The Experimental Setup and Results Analysis

A mobile robot (Mobile Robot, Customer & Robot Co., Ltd) with build-in wheel encoders was used for experiments. MEMS based IMU unit containing an accelerometer (KXPS5-3157, Kionix Inc.), and a gyroscope (ADIS1 6255, Analog Devices Inc.), a rotating camera (SPC 520NC, Philips) with inputs from DC motor (Series 2619, MicroMo Electronics Inc.) and a robot control PC were onboard the robot platform. The signals from the inertial sensors are sent to the host control PC via RS232 communication. The Bluetooth is used as a communication source between the robot control and the host PC. Signals are sent to the camera control motor using RS485 communication protocol. The program to get the sensor data from encoders and inertial sensors is written in Microsoft Visual C++ 6.0.

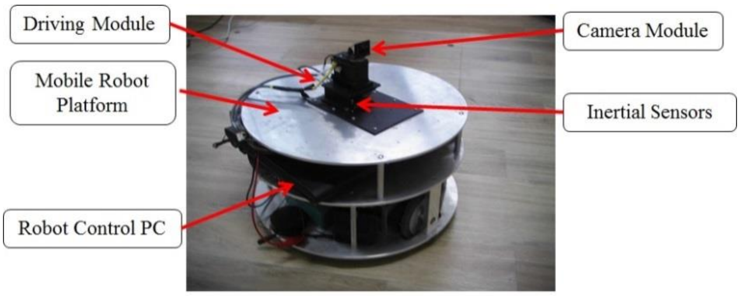

Fig. 5. System Configuration

8.1 The Localization Experiments

Experiments are first conducted to test the robot motion information block (block1). The results have been tested by comparing position output with an accurate distance laser sensor (DLS-BH 30, Dimitix). Intentional slip is generated by rolling a paper sheet under the robot wheels. It would be interesting to compare block1 results in slip case. During the slip experiment, using only encoder data gives 238mm position error (for a 1m linear motion) while our UKF system gives only 96mm error. The comparison of our system block1 performance with encoders and laser sensor is given in figure 6.

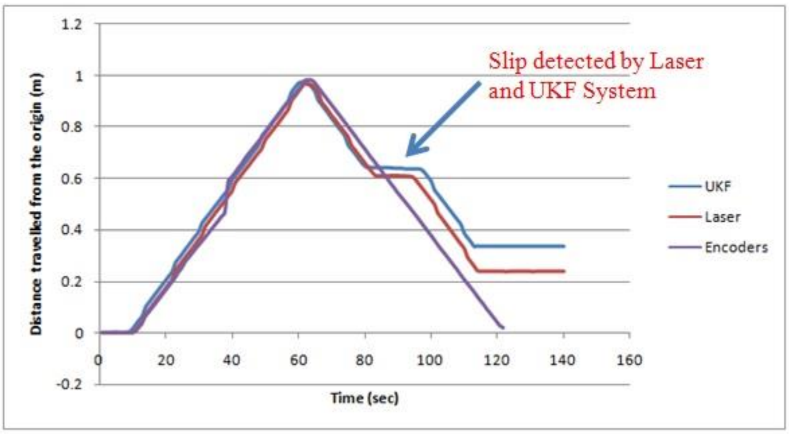

Fig. 6. Distance from the origin as measured by our algorithm, Encoders and accurate Laser Sensor during simple translational motion (slip case).

8.2 The Tracking System Experiments

Experiments were also conducted to test the accuracy of the vision tracking system. The robot camera is initially fixed at the target. The robot is then moved in various trajectories to test the efficiency of vision tracking system. The SURF (Speeded

Up Robust Features) algorithm is used to detect the target inside the camera image. The following three types of experiments were conducted:

1. Simple Translational Motion: Robot moved 1m forward, and then moved 1m backward. The initial distance between the robot camera and the target was kept at 1.5 m and the velocity of the robot is 0.2 m/s.
2. Translational Rotational Combined Motion: Robot moved 1m forward, rotated 90° counterclockwise and then moved 1m forward. The initial distance between the robot camera and the target is once again kept at 1.5m and the translational and rotational velocities of robot are 0.2 m/s and 30 °/s respectively.
3. The Square Motion: The robot followed the square path described in figure 6. The initial distance, translational and rotational velocities of robot remain same.

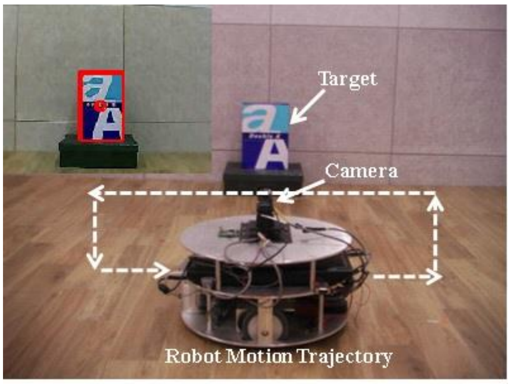

Fig. 7. Experiment 3: Square Motion

The results obtained are summarized in the table below.

Experiment	Tracking Success Rate	Pixel Error (RMS)	Angle Error (RMS)	Recognition Success Rate
Experiment 1 Linear Motion	100 %	21.62 pixel	1.96°	100 %
Experiment 2 Rotation + Translation	99 %	32 pixel	2.83°	92 %
Experiment 3 Square Motion	96 %	33 pixel	2.59°	92 %

Table 1. Summary of the experimental results

Successful tracking is established when the target lies within the camera image. Successful recognition is established when target is recognized by SURF, forming a red boundary around the target as shown in figure 7. Tracking success rate is calculated by dividing the number of successful tracking by number of total image frame. Similary the recognition success rate is calculated by the number of successful recognition divided by the number of total image frames. Angle error gives the degree by which target image is displaced from the centre of camera image. SURF is used for all those calculations.

9 Conclusion

We have presented the implementation of a vestibulo-ocular-reflex (VOR) based vision tracking system for mobile robots. Since the cost of inertial sensors is very low and most of the mobile robots have built-in wheel encoders, this system can be implemented with a very low cost. The system used an advanced filtering technique called the unscented Kalman filter (UKF) for sensor data fusion.

The system is also designed to work in the case of slip in the robot wheels. The system detects the slip by comparing data from encoders and accelerometer and accordingly switches to one of the sensor fusion model built for slip and no-slip cases. The no-slip sensor fusion model integrates data from encoders and gyroscope while the slip case sensor fusion model integrates data from gyroscope and accelerometer. The system implemented in this work has been experimentally tested for accuracy and gives satisfactory results both in slip and no-slip cases. However, the performance of the system is not up to the expected level in the case of robot wheel slip. This is due to the fact that the slip sensor fusion model uses data from accelerometer, which is very noisy. A more careful and robust modeling of the accelerometer noises is required to achieve better performance in the case of slip.

10 References

1. Jaehong P., Wonsang H., Hyun-il K., Jong-hyeon K., Chang-hun L., Anjum M. L., Kwang-soo K. and D. Cho.: High Performance Vision Tracking System for Mobile Robot Using Sensor Data Fusion via Kalman Filters. In: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems, (IROS 2010), Taipei, Taiwan.
2. Shim E. S., Hwang W., Anjum M. L., Kim H. S., Park K. S., Kim K., D. Cho: Stable Vision System for Indoor Moving Robot Using Encoder Information. In: Proceedings of 9th IFAC Symposium on Robot Control, (SYROCO), 2009.
3. Barshan, B., and Durrant-Whyte, H.F.: Inertial Navigation Systems for Mobile Robots. IEEE Transaction on Robotics and Automation, Vol. 11, No. 3, June 1995, pp. 328-342.
4. Dieterich M., Brandt T.: Vestibulo-ocular Reflex. Current Opinion in Neurology, Volume 8, Issue 1, February 1995.
5. Merwe, R.V., Wan, E.A., and Julier, S.I.: Sigma-point Kalman filter for nonlinear estimation and sensor-fusion, application to integrated navigation. American Institute of Aeronautics and Astronautics.
6. Julier, S., Ulhmann, J., and Durrant-Whyte, H.F., "A New Method for the Nonlinear Transformation of Means and Covariances in Filters and Estimators", IEEE Transaction on Automatiac Control, Vol. 45, No. 3, March 2000, pp. 477-482.
7. Van der Merwe, R.: Sigma-Point Kalman Filter for Probabilistic Inference in Dynamic State Space Models. PhD Thesis, OGI School of Science and Engineering at Oregon Health & Science University, Portland, OR, April 2004.

8. D. Simon.: Optimal State Estimation: Kalman, H-Infinity, and Nonlinear Approaches. 1st ed. Wiley & Sons, 2006.
9. W Hwang, J Park, H Kwon, ML Anjum, J Kim, C Lee, K Kim, DD Cho.: Vision Tracking System for Mobile Robots using Two Kalman Filters and a Slip Detector. In: Proceedings of International Conference of Control, Automation and Systems (ICCAS), October, 2010.
10. H Kwon, J Park, W Hwang, J Kim, C Lee, ML Anjum, K Kim, D Cho.: Sensor Data Fusion using Fuzzy Control for VOR-based Vision Tracking System. In: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems, (IROS 2010), Taipei, Taiwan.
11. Zhen Jia, A. Balasuriyaa and S. Challa.: Sensor Fusion-based Visual Target Tracking for Autonomous Vehicles with the Out-of-Sequence Measurements Solution. Robotics and Autonomous Systems. Volume 56, Issue 2, 29 February 2008, pp 157-176.
12. Anjum M. L., Jaehong P., Wonsang H., Hyun-il K., Jong-hyeon K., Changhun L., Kwangsoo K., and D. Cho.: Sensor Data Fusion using Unscented Kalman Filter for Accurate Localization of Mobile Robots. In: Proceedings of International Conference of Control, Automation and Systems (ICCAS), October, 2010.
13. P. Corke, J. Lobo, and J. Dias.: An Introduction to Inertial and Visual Sensing. International Journal of Robotic Research, Vol. 26, pp. 519-535, 2007.
14. S. Hutchinson, G.D. Hager, and P. Corke.: A Tutorial Introduction to Visual Servo Control. IEEE Transaction on Robotics and Automation, vol. 12, no. 5, 1996.
15. http://en.wikipedia.org/wiki/Vestibulo-ocular_reflex