The 2010 IEEE/RSJ International Conference on
Intelligent Robots and Systems
October 18-22, 2010, Taipei, Taiwan

Noise Model Creation for Visual Odometry with Neural-Fuzzy Model

Atsushi Sakai, Masahito Mitsuhashi and Yoji Kuroda
Meiji University, Department of Mechanical Engineering,
1-1-1 Higashimita, Tama-ku, Kawasaki, Kanagawa, Japan
Email: {asakai.amsl, mitsuhashi.amsl}@gmail.com, ykuroda@isc.meiji.ac.jp

Abstract— **In this paper, we propose a technique of learning a noise pattern of visual odometry for accurate and consistent 6DOF localization. The noise model is represented by three parameters of feature points as input: I) The number of inliers among feature points, II) Average of distances between feature points, III) Variance of interior angles. The correlation between these parameters and estimation error is also described. To approximate the complicate noise model accurately, our technique adopts Hybrid neural Fuzzy Inference System (HyFIS) for a learning engine. The noise model is created with HyFIS beforehand, and then the error of visual odometry is estimated by the noise model and compensated on the fly. Learning results of the noise model and results of 6DOF localization in untextured and dynamic environments are presented, effectiveness of our technique is shown.**

Index Terms— **6DOF Localization, Neuro-Fuzzy Learning, Noise Model, Visual Odometry.**

I. Introduction

THE ability of a mobile robot to know its position and attitude, which is known as localization, is critical to its autonomous operation and navigation. In recent years, navigation methods mainly using vision sensors have gotten a lot of attentions [1]-[3]. Because image sensors are relatively inexpensive, compact, light-weight, and low-power. Above all, the image sensors can represent a potential answer to the need of new and improved perception capabilities for mobile robots.

Visual odometry (VO) is a localization technique only using vision sensors. VO estimates a 6 Degrees-Of-Freedom (6DOF) pose of a robot with obtained images for each time. VO can localize the robot in various environments where are difficult to localize by general localization methods. For example, a field where GPS data can not be obtained, rough terrain, and sandy fields. A number of VO algorithms have been proposed (e.g. using a single camera [4]-[6], using stereo cameras[7]-[9]). These techniques enable to localize accurately in various environments.

However, the VO might lead to fatal error from various causes. For instance, being in untextured environments, dynamic objects, inaccurate feature positions, outliers, and rapid image alterations. The error would induce a crucial localization error and spoil the accuracy and consistency of VO. To solve these problems, several techniques have been proposed: outlier rejection is achieved using preemptive RANSAC [10] in [6], bundle adjustment is used to reduce the accumulation error [8], and key frame adjustment is used instead of the bundle adjustment [11]. These approaches enable to localize accurately and stably only using image sensors.

In this paper, we focus on pulsive noise of VO. For VO to be suitable for practical use, it is important for the estimation of VO to be stable and accurate. However, the pulsive noise of VO often occurs when a robot is in untextured and dynamic environments. The pulsive noise can result in fatal error of localization. Even if the techniques described above are used, it is difficult to cope with the pulsive noise. Because enough feature points cannot be detected or observed feature positions are inaccurate.

To solve the pulsive noise problem, in this paper, we propose a technique to reduce the error of VO which is different from the techniques described above. Especially, the technique can reduce the effect of pulsive noises of visual odometry. The technique detects a occurrence of the pulsive noise using a noise model. The noise model of VO which is represented by some parameters of feature points is created, then is used for compensating the pulsive error. The noise model has three parameters as input: I) The number of inliers among feature points, II) Average of distances between feature points, III) Variance of interior angles. These parameters represent various situations of feature points and express accuracy of VO. The noise model is obtained by learning with Hybrid neural Fuzzy Inference System (HyFIS) [12]. HyFIS can learn a nonlinear dynamical system and obtain not only its optimal parameters but also its optimal structure by a learning process. Finally, an error ratio of VO is presented from the noise model on the fly, the estimation error of VO is compensated using the error ratio.

Our technique is robust to variations in illumination. Because the created noise model is only represented by geometric parameters of feature points. Moreover, the technique is low computational cost because the noise model is created by the advance learning.

II. 6DOF Localization with Visual Odometry

In this section, our visual odometry method is described. The visual odometry is a process of determining equivalent odometry information using only camera images. Our VO system has two calibrated cameras and they are used for feature tracking to estimate a relative incremental motion between two frames. The VO estimates a travel distance as following steps. At first, feature points are detected in a image by Shi-Tomasi feature detector [13] and Normalized Cross-Correlation (NCC) is used for stereo matching. Then,

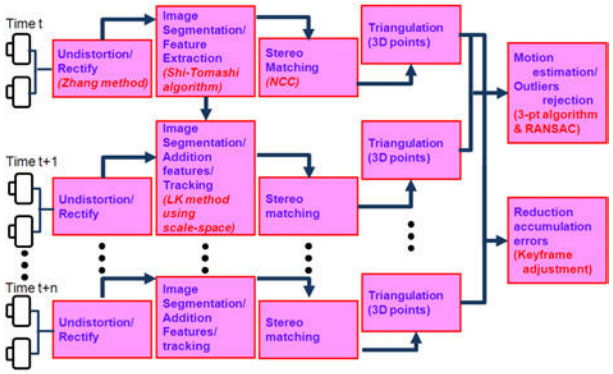

Fig. 1. The flow of our visual odometry system.

the tracking of visual landmarks between consecutive frames is performed by LK method using Gaussian pyramid. Next, these feature points are triangulated at each frame based on stereo correspondences. Finally, a motion parameter is estimated using the framework of RANdom SAmple Consensus (RANSAC) and 3-point algorithm [14][15][16]. They help to reduce the number of outliers and make our system more robust and accurate. Moreover, to keep accuracy and consistency of VO in untextured and dynamic environments, every image is segmented into squired grids. The most characteristic feature point is chosen form the extracted feature points in each grid. In order to reduce the accumulation error, we use key frame adjustment [11]. Fig. 1 shows the flow of our visual odometry. The details of our VO method is described in [16].

III. Hybrid neural Fuzzy Inference System

Hybrid neural Fuzzy Inference System (HyFIS) is an adaptive neural-fuzzy system for building and optimizing fuzzy model [12]. Neural-Fuzzy system like HyFIS has two advantages: the learning power of neural networks and the defined structure of Fuzzy connection. Moreover, HyFIS has a great learning engine which includes not only optimal parameter learning but also optimal structure learning. The details of HyFIS architecture and the learning engine are described in following sections.

A. The architecture of HyFIS

Fig. 2 shows the architecture of HyFIS. HyFIS has five layers for creating a fuzzy system. The structure eliminates the disadvantage of a normal neural network which it's structure is indefinite. We use indices, i, j, k, and l for nodes in layers 2, 3, 4, and, 5 respectively. The output from the nth node of mth layer is denoted by $y_n^{[m]}$.

1) Layer 1: Nodes in layer 1 normalize each input to the standard interval $[-1\ 1]$ and transmit each input to the next layer. In each parameter, a maximum absolute value of all teaching data is used for the normalizing.

2) Layer 2: Nodes in layer 2 act as membership functions to represent linguistic variables. In this paper, we use five fuzzy sets for the Neural-Fuzzy model: Large Negative (LN), Small Negative (SN), Zero (ZE), Small Positive (SP), and

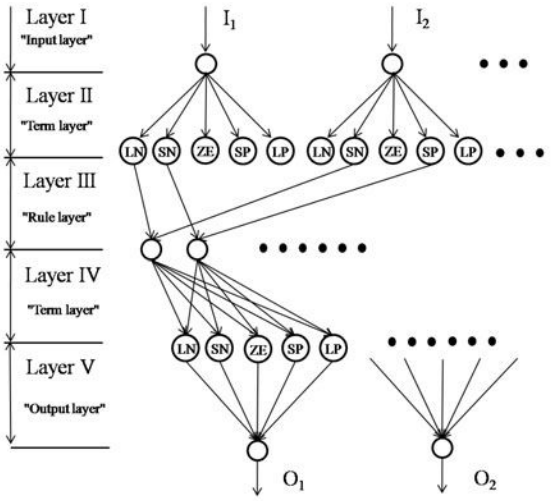

Fig. 2. Architecture of HyFIS.

Large Positive (LP). We use following Gaussian functions as the membership:

$$y_i^{[2]} = \exp(-\frac{(x-c_i)^2}{\sigma_i^2}) \tag{1}$$

Where i is a index of each membership function, c and σ are the mean and the covariance of Gaussian distribution respectively, and x is an input. Initially, the parameters of membership function are spaced equally over the weight space.

3) Layer 3: Each node in layer 3 represents a possible IF-part of a fuzzy rule. The nodes perform the AND operation as follows:

$$y_j^{[3]} = \min_{i \in I_j}(y_i^{[2]}) \tag{2}$$

Where I_j is the set of indices of the nodes which are connected to the node j in layer 3.

4) Layer 4: Each node in layer 4 represents a possible THEN-part of a fuzzy rule. The nodes perform the OR operation to integrate the information leading to the same output linguistic variables. Each node has a connection weight $w_{k,j}$ of the node j in layer 3 to the node k in layer 4. The functions of this layer are expressed as follows:

$$y_k^{[4]} = \max_{j \in I_k}(w_{kj} y_j^{[3]}) \tag{3}$$

Where I_k is the set of indices of the nodes which are connected to the node k in layer 4.

5) Layer 5: Nodes in layer 5 act as a defuzzifier. A node in this layer outputs a crisp output value. In this paper, Center of Gravity method is used, the output is computed as follows:

$$y_l^{[5]} = \frac{\sum_{l \in I_l} y_k^{[4]} \sigma_{lk} c_{lk}}{\sum_{k \in I_k} y_k^{[4]} \sigma_k} \tag{4}$$

Where I_l is the set of indices of the nodes which are connected to the node l in layer 5.

B. Hybrid learning algorithm

HyFIS has two kinds of learning algorithm; Structure generation and Parameter learning. This section describes both learning algorithms.

1) Structure generation: HyFIS generates a set of fuzzy rules from teaching data to solve the conflict problem (e.g. rules have the same IF-part but a different THEN-part) and to delete redundant rules. This process consists of the two steps. At first, maximum membership degrees are computed to each input and output from teaching data. And then, the linguistic pairs are selected from the result and generates the fuzzy rules. If two or more generated fuzzy rules have the same preconditions and consequents, the rule which has maximum degree is used:

$$S = \underset{t \in I_{TD}}{max} (\prod_{n=1}^{M} \mu_{n,t}(x_{n,t})) \tag{5}$$

Where S is a set of pairs of IF and THEN part, I_{TD} is a set of indices of the data which has same rules, $\mu_t(x_t)$ is each membership degree, and M is the number of membership functions of input and output.

2) Parameter learning: After the whole network structure is established, HyFIS adjusts optimally the parameters of membership functions (e.g. mean c and variance σ of Gaussian distribution). In HyFIS, the gradient descent learning is used to minimize an error function. The error function is as follow:

$$E = \frac{1}{2} \sum_{n=1}^{L} (d_n - y_n^{[5]})^2 \tag{6}$$

Where E is an error degree, L is the number of teaching data, d_n and $y^{[5]}$ are a target and an actual output respectively. The update rule is as follow:

$$P_{t+1} = P_t - \eta \frac{\partial E}{\partial P} \tag{7}$$

Where P_t is a parameter at the time t and η is the learning rate. The details of the update process are described in [12].

IV. Parameters of Noise Model

This section describes the parameters of noise model and correlations between the noise parameters and estimation error. Our noise model estimates the accuracy of visual odometry by three parameters. Other kinds of parameters (e.g., The maximum difference in height, average of Normalized Cross Correlation for the stereo matching, and the number of feature pairs when the number of inliers is maximum, etc) have also been evaluated. From these results, we concluded that the three parameters are important especially for the VO accuracy. Fig. 3 is the simplified schematic of the noise parameters.

A. Noise parameters

1) The number of inliers among feature points: The first parameter is the number of inliers among feature points. The number of inliers can be obtained using the framework of RANSAC and 3-point algorithm in motion estimation process [14][15][16]. Firstly, a hypothesis of motion parameter is generated from a set of 3 sample feature points selected randomly from all points. And then, all feature points observed before movement are reprojected into the current image frame using the hypothesis. Next, we calculate the reprojection error to each point. If the reprojection error is smaller than a threshold (in this paper, 1.5 pix), the data is regarded as inliers. Finally, we count the number of inliers.

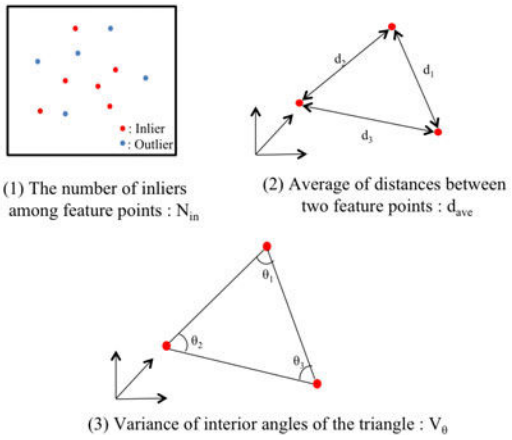

Fig. 3. Simplified schematic of noise parameters.

2) Average of distances between two feature points: The second parameter is the average of distances between two feature points. Each distance is calculated with feature's locations in 3D space, and the average of distances is computed as follows:

$$d_{ave} = \frac{1}{2N} (\sum_{i=1}^{N} d_{B,i} + \sum_{j=1}^{N} d_{A,j}) \tag{8}$$

where d_{ave} is the average of distances, $d_{A,j}$ and $d_{B,i}$ are the distance of each feature pair in a current stereo image (*After moving*) and previous one (*Before moving*) respectively, and N is the number of feature pairs. In this paper, N = 3 because we adopt the 3-point algorithm for motion estimation.

3) Variance of interior angles of triangle: The third parameter is variance of interior angles of the triangle which is made of selected 3 feature points. The variance is computed as follows:

$$V_\theta = \frac{1}{2N} (\sum_{i=1}^{N} (\bar{\theta_B} - \theta_{B,i})^2 + \sum_{j=1}^{N} (\bar{\theta}_A - \theta_{A,j})^2) \tag{9}$$

where V_θ is the variance of interior angles, $\theta_{A,j}, \theta_{B,i}$ and $\bar{\theta_A}, \bar{\theta_B}$ are interior angles and their mean values respectively. The variance being close to 0 means that the triangle is close to the regular triangle.

B. Correlations between noise parameters and estimation error.

We conducted a experiment for showing the correlations between each parameter and estimation error. The stereo camera translated by 10 cm and images were obtained. Then,

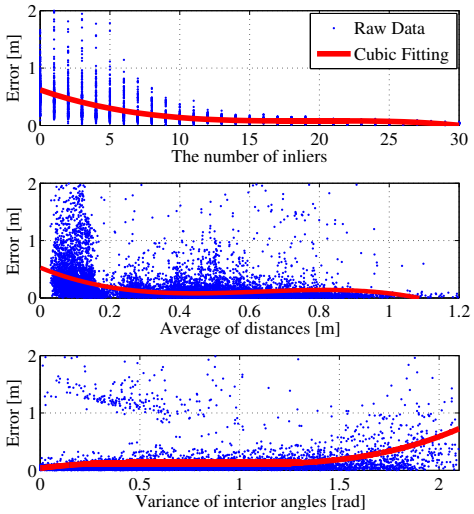

Fig. 4. Correlations between noise parameters and estimation error.

each noise parameter and estimation error were calculated. The estimation error is a norm between an estimation distance and the true distance in 3D space.

Fig. 4 shows the results of the experiment. Each graph has raw data plot and a cubic fitted curve of the raw data. The fitted curves represent that there is a correlation between each noise parameter and estimation error. The result shows that the noise parameters can represent the estimation accuracy of VO.

V. Error Compensation with the Noise Model

This section describes an error compensation method with the noise model. In this research, the noise model has 3 inputs and 1 output. The 3 inputs are the noise parameters described above, the output is a noise ratio of VO in the interval $[0\ 1]$. We use the noise ratio for the error compensation. The error compensation uses the weighting addition of two incremental motion hypotheses. The first one is the hypothesis of VO and the second one is the hypothesis of motion history and motion assumption. The conclusive motion distance is the expected value calculated by the two hypotheses and the noise ratio as follows:

$$d_t = p_t d_{vo,t} + (1 - p_t) d_{m,t}\ , \quad p_t \in [0\ 1] \tag{10}$$

where d_t is the conclusive moving distance, p_t is the noise ratio, $d_{vo,t}$ is the hypothesis of VO, and $d_{m,t}$ is the hypothesis of motion history and motion assumption. In this paper, the $d_{m,t}$ is obtained from a previous motion distance and assumption of plane motion. On the moving direction along the x axis (Fig. 5) and rotation directions of each angle, the previous motion distance is used as the second hypothesis. On the horizontal direction along the y axis and the vertical direction along the z axis, the second hypothesis is 0 (no moving) with the plane motion assumption. Because, considering our control system, structured pavement, and the robot's mechanical structure, we can say that the forward and heading velocity don't change rapidly and the robot hardly

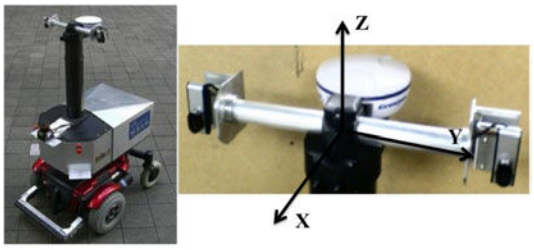

Fig. 5. The mobile robot, named *Infant* and our stereo camera.

move vertically and horizontally. Therefore, each incremental motion distance and rotation angle are computed from Eq. 10 by following equations:

$$dx_t = p_t dx_{vo,t} + (1 - p_t) dx_{t-1} \tag{11}$$

$$dy_t = p_t dy_{vo,t} \tag{12}$$

$$dz_t = p_t dz_{vo,t} \tag{13}$$

$$d\theta_t = p_t d\theta_{vo,t} + (1 - p_t) d\theta_{t-1} \tag{14}$$

where dx, dy, dz are incremental distances on each direction, and $d\theta$ is a incremental rotation angle on each direction.@

VI. Experimental Results

This section describes results of created noise model and localization. We used a mobile robot platform named *Infant* and a stereo camera for the experiments (Fig. 5). The cameras were Qcam for Notebooks Pro (QVX-13NS) made by Logicool. Our stereo camera was calibrated by Zhang calibration method [17]. When observing a place about 5 m away, distance accuracy was 10 cm (the error ratio is 2%). The interface between the cameras and a computer was USB2.0, and the resolution of images was VGA (640 480 pix). The stereo image sequences were acquired by about 4 Hz, the baseline of the stereo camera was 37 cm, and the depression angle of cameras was 23 degrees. The experiments were done by the system that the CPU is Intel Core2 Duo 2.33 GHz, the RAM is 3.25 GB. OpenCV was used as an image-processing library.

A. Creating noise model

Experiments were conducted for obtaining teaching data. We translated the stereo camera by $10cm$ on the directions along x or y axis, and obtained stereo images. Then, we estimated moving distances with VO using the images and calculated estimation error and the noise parameters for teaching data. The estimation error is a norm between an estimation distance and the true distance in 3D space. The total number of the teaching data was 190.

Table I shows the results of structure learning. 16 rules were selected from 625 of a total fuzzy rules. The maximum degrees in the table were used for the weights of HyFIS network w_{kj}. Fig. 6 shows the learning results of parameters of each membership function. The upper left figure shows the initial distribution of all membership function, and the others shows the learning results. The learning results of HyFIS was used for the error compensation.

TABLE I

RESULTS OF STRUCTURE LEARNING

	IF			THEN	
Rules	A	B	C	y	Degree
1	ZE	SP	ZE	SP	0.450141
2	ZE	SP	SP	SP	0.383080
3	ZE	SP	LP	SP	0.212231
4	ZE	LP	ZE	ZE	0.422216
5	SP	ZE	ZE	SP	0.549189
6	SP	ZE	SP	ZE	0.447151
7	SP	SP	ZE	ZE	0.960294
8	SP	SP	SP	ZE	0.878823
9	SP	SP	LP	LP	0.792584
10	SP	LP	ZE	ZE	0.806533
11	SP	LP	SP	ZE	0.462327
12	LP	ZE	ZE	ZE	0.578093
13	LP	ZE	SP	SP	0.391987
14	LP	SP	ZE	ZE	0.496780
15	LP	SP	SP	ZE	0.636455
16	LP	SP	LP	SP	0.397486

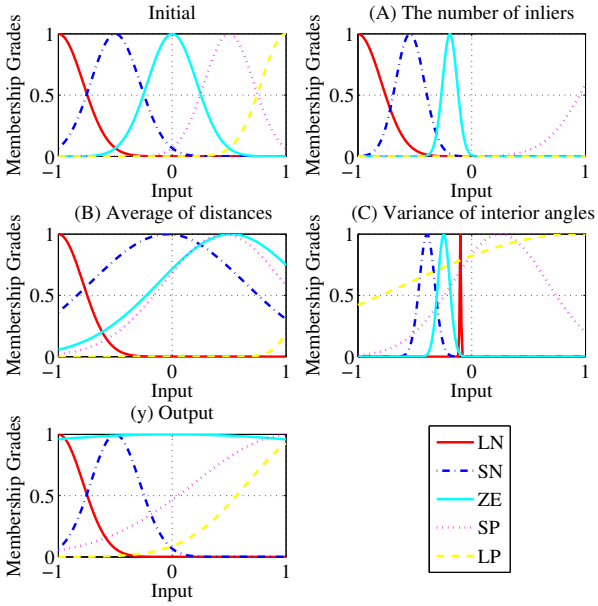

Fig. 6. Learning results of member functions.

B. Localization result

In this section, we show experimental results of 6DOF localization: $x = [x, y, z, \psi, \phi, \theta]$. Two test runs were conducted in untextured and dynamic environments. Fig. 7 shows examples of tracking images in the experimental fields. The most of the fields were made of untextured pavement and had many moving objects (e.g. pedestrians). In the experiments, the robot moved about $0.3\ m/s$, traveled about $45\ m$ in the first run, about $70\ m$ in the second run. Ground truth was obtained by a sensor fusion by Unscented Kalman filter using DGPS, IMU, and wheel odometry [18].

Fig. 8 shows the 6DOF localization result at the first run and Fig. 9 shows the result on the x-y surface. Fig. 8 and Fig. 9 have three trajectories; Ground truth, VO, and VO with

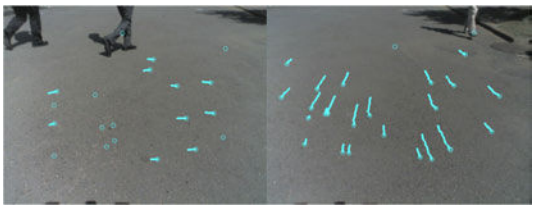

Fig. 7. Feature tracking in an untextured and dynamic environment.

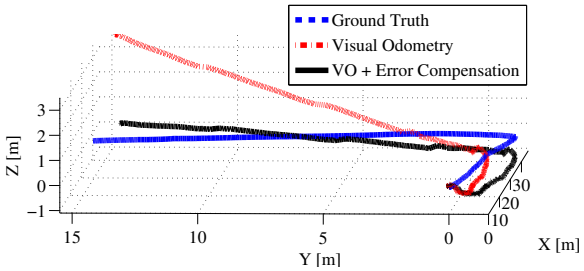

Fig. 8. 6DOF localization result (Run1).

error compensation. The localization result of simple VO had a big error. On the other hand, the accuracy of VO with error compensation was improved as compared with the simple VO. Especially, the error of the roll angle was reduced. Fig. 10 shows variations of roll angle as a function of time in the first run. This figure shows that pulsive noises sometimes occur when the robot starts to move and curve. This pulsive noises resulted in a big error of the roll angle. However, the figure also shows that the learned noise model can detect the pulsive noises and the error compensation reduced the effect of the pulsive noises. The result shows that the localization accuracy is improved.

Fig. 11 shows the 6DOF localization result in 3D space at the second run and Fig. 12 shows the result on the x-y place. The result shows that the accuracy of VO with error compensation is improved as well as the result at the first run.

Fig. 13 is a comparison of resultant error of simple VO and VO with error compensation as a function of time in both experiments. In both results, we obtained precise results as compared with the simple VO all of the time. Especially, the

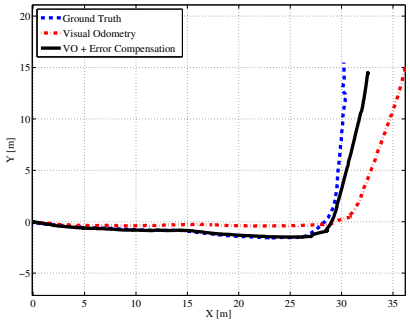

Fig. 9. Localization result on the x-y surface (Run1).

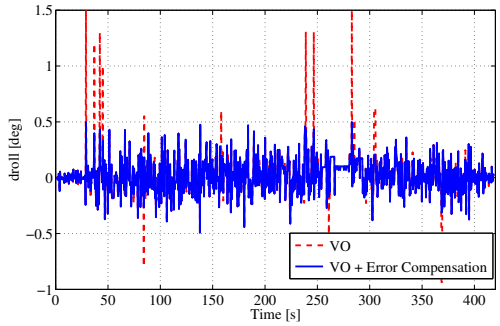

Fig. 10. Variations of the roll angle (Run1).

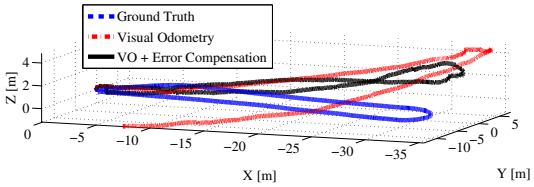

Fig. 11. 6DOF Localization result (Run2).

final position error was reduced by half or more.

VII. Conclusion

In this paper, we proposed a technique for learning a noise model of visual odometry for accurate and consistent 6DOF localization. The technique learns the noise model represented by three parameters of feature points. To approximate accurately the complicate noise model, our technique adopts Hybrid neural Fuzzy Inference System (HyFIS) for a learning engine. The noise model is learned on HyFIS beforehand, and then the error of visual odometry is estimated with the noise model and compensated on the fly. Learning results and 6DOF localization results in untextured and dynamic environments were presented, the effectiveness of our technique was shown.

References

[1] P. Sermanet, R. Hadsell, M. Scofier, M. Grimes, J. Ben, and A. Erkan. "A Multirange Architecture for Collision-Free Off-Road Robot Navigation", Journal of Field Robotics 26(1), 52-87, 2009.

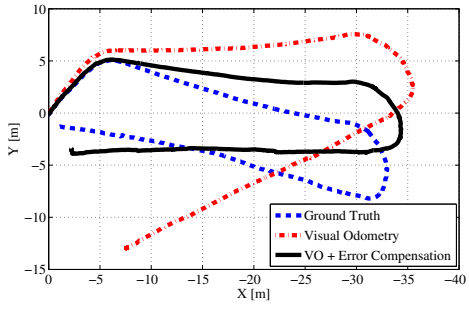

Fig. 12. Localization result on the x-y surface (Run2).

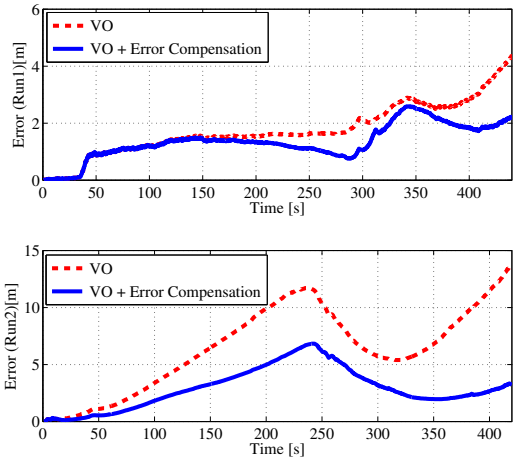

Fig. 13. Comparing errors of simple VO and VO with error compensation.

[2] M. Bajracharya, A. Howard, L. H. Matthies, B. Tang, and M. Turmon. "Autonomous Off-Road Navigation with End-to-End Learning for the LAGR Program", Journal of Field Robotics 26(1), 3-25, 2009.
[3] K. Konolige , M. Agrawal, M. R. Blas, R. C. Bolles, B. Gerkey, J. Sola, and A. Sundaresan. "Mapping, Navigation, and Learning for Off-Road Traversal", Journal of Field Robotics 26(1), 88-113, 2009.
[4] S. I. Roumeliotis, A. E. Johnson, and J.F. Montgomery, "Augmenting Inertial Navigation with Image-Based Motion Estimation", Proceedings of the 2002 IEEE International Conference on Robotics Automation, Washington, pp. 4326-4333, 2002.
[5] A.J. Davison, "Real-Time Simultaneous Localization and Mapping with a Single Camera ", IEEE Int. Conf. on Computer Vision, Nice, pp. 1403-1410, 2003.
[6] D. Nister, O. Naroditsky, and J. Bergen, "Visual Odometry", Proc. IEEE Computer Society Conference on Computer Vision and Patterrn Recognition, Volume 1, pp.652-659, 2004.
[7] A. Milella and R. Siegwart, "Stereo-Based Ego-Motion Estimation Using Pixel Tracking and Iterative Closest Point", Proceedings of the Fourth IEEE International Conference on Computer Vision System (ICVS 2006), 2006.
[8] K. Konolige, M. Agrawal, and J. Sola. "Large-scale visual odometry for rough terrain.", In Proceedings of the International Symposium on Research in Robotics, Hiroshima, Japan, November 26-29 2007.
[9] C.F. Olson, L.H. Matthies, M. Schoppers, and M.W. Maimone, "Rover Navigation Using Stereo Ego-Motion", Robotics and Autonomous Systems, 43, pp. 215-229, 2003.
[10] D. Nister, "Preemptive RANSAC for Live Structure and Motion Estimation", IEEE International Conference on Computer Vision, Nice, pp. 199-206, 2003.
[11] M. Tomono, "Robust Stereo SLAM Based on Edge-Point ICP and Error Recovery", 17-th Robotics Symposia, pp.217-222, 2009
[12] Kim, and N. Kasabov, "HyFIS: Adaptive neuro-fuzzy systems and their application to non-linear dynamical systems", Neural Networks, v.12 (9), pp.1301-1321, 1999.
[13] J. Shi and C. Tomasi, "Good Feature to Track", IEEE Conference of Computer Vision and Pattern Recognition, CA, pp. 593-600, 1994.
[14] M.W. Maimone, Y. Cheng, L. Matthies, "Two Years of Visual Odometry on the Mars Exploration Rovers", Journal of Field Robotics, Volume 24 number 3, special issue on Space Robotics, Mar. 2007.
[15] K. Ni, and F. Dellaert, "Stereo tracking and three-point/one point algorithms - a robust approach in visual odometry", In International Conference on Image Processing, 2006.
[16] Y. Tamura, M, Suzuki, A.Ishii, Y. Kuroda, "Visual Odometry with Effective Feature Sampling for Untextured Outdoor Environment", 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS09), St. Louis, USA, Oct, 10-15, 2009.
[17] Z. Zhang, "A Flexible New Technique for Camera Calibration", Pattern Analysis and Machine Intelligence, vol. 22, 2000.
[18] Julier and J. K. Uhlmann, "General method for approximating nonlinear transformations of probability distributions", Robotics Research Group Department of Engineering Science, University of Oxford 1996.