Proceedings of the 17th World Congress
The International Federation of Automatic Control
Seoul, Korea, July 6-11, 2008

A Virtual Range Finder based on Monocular Vision System in Simultaneous Localization and Mapping

X.Z. Zhang*. A.B. Rad. and Y.K. Wong***

**Department of Electrical Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong (e-mail: {eexz.zhang, eeykwong}@polyu.edu.hk)*
*** School of Engineering Science, Simon Fraser University, 250-13450, 102 Avenue, Surrey, B.C., V3T 0A3,, Canada (e-mail:arad@sfu.ca)*

Abstract: This paper presents a virtual range finder model with the monocular vision system for simultaneous localization and mapping (SLAM). It relaxes the constraint often cited in the literature that the motion of the optical axis has to be parallel, and reduces the errors for range extraction by a single camera. This model could also provide a supplementary range measurement for landmark initialization in bearing-only SLAM. As the sensor data transformation from pixel to metric value is a nonlinear process, the uncertainty for observation model adopted in Extended Kalman Filter (EKF) SLAM framework can not be in the Gaussian form, which probably makes difficult for data association and SLAM. Concerning this problem, we present a new data association technique based on the homography transformation by a sequence of images and integrate it into the update process of the EKF to assist the innovation computation. The experimental results on real data validated the performance of the virtual range finder model and the new data association approach.

1. INTRODUCTION

Simultaneous localization and mapping (SLAM) is considered as a fundamental ability for autonomous mobile robots. The robot is required to form a consistent representation from various sensor measurements and concurrently determine its pose by using this map. The perceptive devices, including intrinsic and extrinsic ones, take a central role in SLAM and ensure the reliability of the generated maps. The useful information is automatically extracted from these sensors accompanied by appropriate algorithms to estimate the states of the environment where the robots navigate.

Range finders, such as laser and sonar, are typical sensors that offer the range and orientation data of an object and are widely applied in feature-based SLAM. Ip et al. (Ip, Rad, Chow, & Wong, 2002) suggested an enhanced adaptive fuzzy clustering (EAFC) algorithm from sonar array data to extract the segments features that were adopted to map building process in static and dynamic environments. Garulli et al. (Garulli, Giannitrapani, Rossi, & Vicino, 2005) also developed a line-based environment representation with the laser range finder. Similarly, employing the laser, Pfister (Pfister, Roumeliotis, & Burdick, 2003) suggested a weighted line algorithm for describing the environment. Besides the line features can be elicited, another simple representation, points/landmarks (Andrade-Cetto & Sanfeliu, 2002; Di Marco, Garulli, Giannitrapani, & Vicino, 2003, 2004; Madhavan & Durrant-Whyte, 2005) can be also obtained from these sensors. Furthermore, Diosi and Kleeman (Diosi & Kleeman, 2004) incorporated the sonar and laser information to generate the metric point feature map.

In recent years, vision system has attracted many researchers as it provides abundant information than the range finders do. With the characteristics of numerous vision systems, related algorithms on feature extraction and SLAM have been developed. By stereoscopic camera (Miro, Dissanayake, & Weizhen, 2005), there is sufficient information for depth extraction as well as the bearings. In contrast, it is a challenge job to retrieve the depth information from monocular images. Fortunately, the depth can be approximately recovered by a sequence of images. Saxena et al. (Saxena, Chung, & Ng, 2005) applied the supervised learning to predict the depth-map with a function of the image by integrating multi-scale local and global image features. Additionally, depth can be estimated from the change of blur depending on the zoom shift, i.e. from defocus (Baba, Oda, Asada, & Yamashita, 2006). Considering the real-time implementation, Davison et al. (Davison, Reid, Molton, & Stasse, 2007) presented a unified inverse depth parametrization and conversion between inverse depth and depth for point features with monocular SLAM in room space with few moving object. Murphey et al. (Murphey et al., 2000) designed a DepthFinder to detect the distances of the objects through at least two images, which acts like a stereo vision system. The errors of this model, however, are serious when the camera motion is unparallel, that is, the ambiguity of image points affects the accuracy. To eliminate variety of ambiguity, Martin (Martin, 2006) introduced several powerful domain specific constraints and presented an evolving visual sonar, but some of these constraints may generate the improper bound and confuse the robots to perceive the pseudo obstacles. Another alternative method only focuses on the bearing measurement reflected in one image when the monocular cameras are adopted. By using this technique, the bearing-only SLAM strategy is derived (Costa, Kantor, &

Choset, 2004; Fitzgibbons & Nebot, 2002; Klippenstein, 2007; Xiang & Hong, 2006).

A single camera has few limitations one of which is depth extraction. Without using the depth data, the bearing-only SLAM framework probably causes landmark initialization problems due to non-availability of range measurements. In this work, we suggested a virtual range finder with a single calibrated camera and extracted the range and bearing measurements of the interested points which are to be applied in the EFK-SLAM framework. Either in static or dynamic environments, this model relaxes the constraint that the motion of the optical axis has to be parallel and provides a complementary range data for landmark initialization in bearing-only SLAM. Additionally, if the real range finder, such as laser, has a serious fault, this model can be used as an alternate strategy to ensure supplying sufficient perceptive information. Since the transformation from the pixel values in images to metric ranges and bearings in the real world is a substantially nonlinear process, the normal distribution for the error model of observations is not maintained. If the uncertainty in the observation model is still assumed as the normal distribution, this can cause some difficulties for data association as well inducing the divergence of the update procedure in the EKF SLAM. To address this problem we present a new data association policy employing the homography transformation based on Scale-invariant Feature Transform (SIFT) (Lowe, 2004) and incorporate the result of this new data association algorithm into the update process to avoid the divergence of the SLAM. The experiments and simulations validate our proposed methods.

The paper is structured as follow: After discussing the geometric formulations of virtual range finder and diverse coordinate references in Section 2, Section 3 depicts the landmark detection by the monocular camera. In Section 4, the novel data association methodology is introduced as well as SLAM framework. Experiments and simulations test the proposed methods in Section 5. Finally Section 6 presents the conclusion and future path of this research.

2. GEOMETRIC FORMULATION

There are four main coordinate references in the autonomous mobile robots system and the overview of the geometric relationship among them is displayed in Fig.1a. The subscript **R** refers to robot or local frame and the related plane is $\mathbf{\Pi}_R$. Similarly, **C** denotes camera reference, **I** refers to the image frame and **W** expresses world or global coordinate system.

For convenience, as shown in Fig. 1b we assume that the origins of robot and camera coordinate frame (O_R and C) are identical. Given a world point **P** on the ground plane, **P**' is the projection of **P** on line **GE** that is the projection of optical axis on the ground. Here **E** is the intersection of the optical axis with the ground plane. The images of **P** and **P**' in $\mathbf{\Pi}_I$ are **p** and **p**'. The height **h** of the optical center from the ground plane and the tilt angle α of the optical axis from the $\mathbf{X_R}$ axis are hand-measured (cf. Fig. 1c). We can recover the range r which is the distance from the robot to the interested point **P** following the equations (1)~(4)

$$\beta = \arctan\frac{v_p - v_o}{f_v}, \gamma = \arctan\frac{u_p - u_o}{(v_{p'} - v_o)/\cos\beta} \tag{1}$$

$$P'C = h/\sin(\alpha+\beta), \quad \text{(cf. Fig. 1c)} \tag{2}$$

$$PC = P'C/\cos\gamma, \quad \text{(cf. Fig. 1b)} \tag{3}$$

$$r = \sqrt{PC^2 - h^2} = \frac{h}{sin(\alpha+\beta)}\sqrt{1 - sin^2(\alpha+\beta)cos^2\gamma} \tag{4}$$

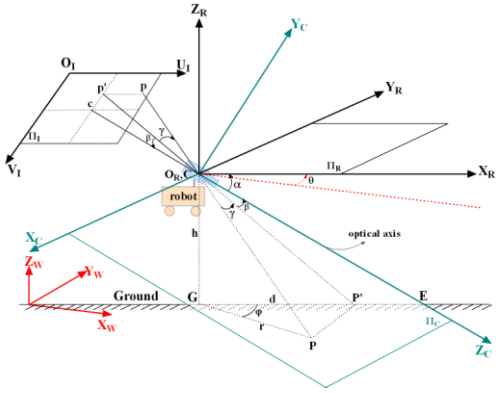
where f_u and f_v are the focal length at the direction of **U** and **V** axis respectively; u_p and v_p are the pixel coordinates of **p**; u_o and v_o are those of principal point. With the same technique, in Fig. 1c we can also retrieve the distance d from the robot to the projection point **P**' according to

$$d = \frac{h}{\tan(\alpha+\beta)} \tag{5}$$

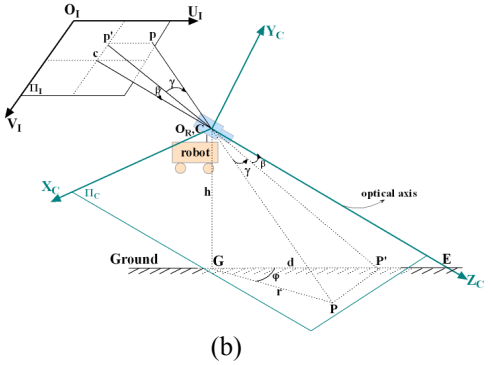
Therefore, the bearing of **P** from the $\mathbf{X_R}$ axis is

$$\begin{aligned}\varphi &= -\mathrm{sgn}(u_p - u_o)\arccos\frac{d}{r} \\ &= -\mathrm{sgn}(u_p - u_o)\arccos\frac{cos(\alpha+\beta)cos\gamma}{\sqrt{1 - sin^2(\alpha+\beta)cos^2\gamma}}\end{aligned} \tag{6}$$

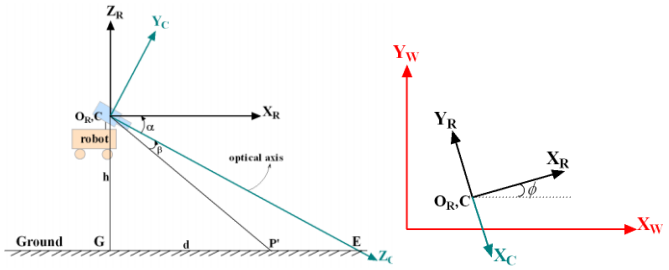
Here the sign function w.r.t. u_p and u_o manifests the bearing is converted into the robot reference system. Equation (4) and (6) are the virtual range and bearing measurements extracted from one image. Note that the monocular camera intrinsic parameters, such as focal length, principal points, etc., are acquired in advance by calibration process (Zhang., 1999).



(a)



(b)



(c) (d)

Fig.1. Geometric relationship in several coordinate references

3. LANDMARK EXTRACTION

In previous section, we suggested that the virtual range and bearing information of the points can be extracted from the monocular images. Usually points may be treated as the landmarks for the mapping or SLAM task, thus it is necessary to define a special type of interested points as the features. In this study we selected the points located on the ground plane in that the corresponding height **h** in computing the range is measured from the floor. With this constraint, we assigned the intersections of vertical and horizontal edges detected from one image as the expected landmarks. The point extraction process is interpreted in detail as follow:

Step1: Pre-processing an acquired image to filter out different noise signals;
Step2: Detecting the vertical edges by the Sobel operator;
Step3: Implementing the morphological operation to lessen improper edges and strengthen appropriate ones;
Step4: Performing Hough transform (Wong, Shi, & Chan, 1997) to obatain the line parameters ρ and θ of the edges;
Step5: Selecting the region of interest (ROI). In this study, we defined a rectangel window which satisfies $48 \le v \le 240$ and $0 < u \le 320$ as the ROI, as is shown in Fig.2;
Step6: In the ROI, detecting the horizontal edges which reveal the bound of the ground by the Sobel operator with a reasonable threshold for the operator and a pixel limitation for the line length;
Step6: Re-running *Step3* and *Step4*;
Step7: Calculating the intersections (i.e. landmarks) of these two distinct edges by cross product as expressing in equation (7).

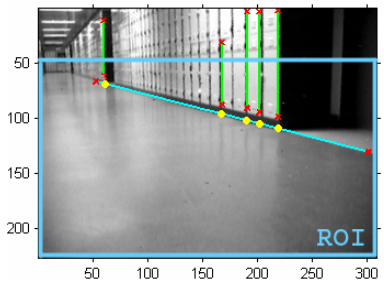


Fig.2. Two different edges and the landmarks (yellow points) calculated by these edges.

$$x_{lm} = l_{gnd} \times l_{vert} = \begin{pmatrix} i & j & k \\ \cos\theta_{gnd} & \sin\theta_{gnd} & -\rho_{gnd} \\ \cos\theta_{vert} & \sin\theta_{vert} & -\rho_{vert} \end{pmatrix} \quad (7)$$

where θ_{gnd} and ρ_{gnd} are the parameters for the horizontal edges, θ_{vert} and ρ_{vert} for the vertical ones. Fig.2 illustrates the two different edges and the determined landmarks exhibited in yellow dots.

4. SLAM FRAMEWORK

The SLAM framework applied in this paper is the widely used algorithm: extended Kalman filter (EKF) same as our previous work (Ip et al., 2002). The motion and observation models are

$$X_k = f(X_{k-1}, u_k) + \varepsilon_k = \begin{bmatrix} x_{k-1} + v_k \Delta t \cos\phi_k \\ y_{k-1} + v_k \Delta t \sin\phi_k \\ \phi_k + \omega_k \Delta t \end{bmatrix} + \varepsilon_k \quad (8)$$

$$z_k = h(X_k, z_{k-1}) + \zeta_k = \begin{bmatrix} \sqrt{(x_{lm} - x_k)^2 + (y_{lm} - y_k)^2} \\ \arctan\dfrac{y_{lm} - y_k}{x_{lm} - x_k} - \phi_k \end{bmatrix} + \zeta_k \quad (9)$$

where $X=[x,\ y,\ \phi]^T$ is the state variable of robot pose; $u_k=[v_k,\omega_k]^T$ is the control variable; $x_{lm}=r\cos\varphi$, $y_{lm}=r\sin\varphi$ are the Cartesian coordinates of the landmark. ε_k is the Gaussian noise of the motion model with zero mean and covariance $\mathbf{Q}_k$; ζ_k is the random noise for observation model, however, it could not be treated as Gaussian form because computing the metric ranges and bearings from the pixel points is an essentially nonlinear process. If ζ_k is simply taken as Gaussian, it probably produces a slight inaccuracy for data association and divergence for update process of EKF. As regarding to these problems, we presented a new data association mechanism based on the homography transformation within several images and then integrate it into the EKF to assist the update procedure.

4.1 Data Association based on Homography Transformation

Although the noise of the observation model is probably not Gaussian, the pixel noise can be considered as the normal distribution and the statistical parameters are accessed by camera calibration. Consequently, it is possible to handle the data association by the pixel coordinates of the landmark instead of the range and bearing measurements. The idea is applying the homography transformation (Hartley, 2003) computed from the Scale-invariant Features between any two images to calculate the Mahalanobis distance. Gil.et al. (Gil, Reinoso, Martinez Mozos, Stachniss, & Burgard, 2006) managed the data association with the SIFT features from the pattern classification viewpoint, and the Mahalanobis distance was established by the average SIFT descriptors and a high-dimensional covariance matrix. By contrast, following Lowe's algorithm (Lowe, 2004), we used SIFT mechanism to determine the stable matched points between any two images and then estimated the homography transformation matrix **M** and its covariance $\Sigma_{\mathbf{M}}$ using MLE technique (Hartley, 2003) by these points. The Mahalanobis distance was calculated based on **M**, $\Sigma_{\mathbf{M}}$ and the pixel coordinates of landmarks. Another reason why we considered SIFT as the matched points selection mechanism for estimating **M** and $\Sigma_{\mathbf{M}}$ is that

Scale-invariant Features maintain stable in dynamic environments. As is shown in Fig.3, it is obvious that those matched points determined by SIFT are unsusceptible to the moving object (the person).

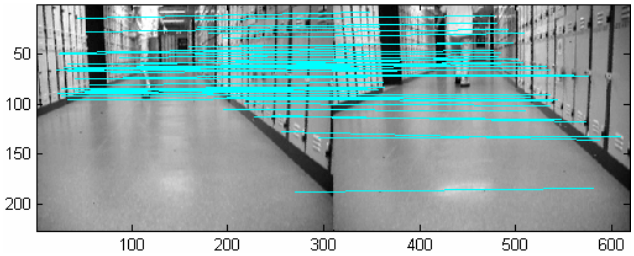


Fig.3. Matched points determined by SIFT in dynamic environments, which can be used to estimate **M** and $\Sigma_{\mathbf{M}}$.

With the homography transformation matrix **M**, the estimated pixel coordinates of one landmark is expressed as

$$\hat{p}_f = Mp_m \tag{10}$$

where p_m is the pixel coordinates of the landmarks existed in the map and $\hat{p}_f$ is those of predicted landmarks. The current captured feature is marked as p_f, and the Mahalanobis distance for data association is

$$d_m = (p_f - \hat{p}_f)^T \Sigma_M^{-1} (p_f - \hat{p}_f) \tag{11}$$

Compared with the formula in Gil's work (Gil et al., 2006), the main differences are the Mahalnobis distance in (11) is constructed by the pixel coordinates of landmarks directly without any SIFT descriptor and the covariance of **M**, and $\Sigma_{\mathbf{M}}$ has low dimension. If d_m is less than a threshold, then a current feature is associated with an existing one in the map, otherwise the feature is a new one. An example of predicting the pixel coordinates according to formula (10) and the data association procedure is demonstrated in Fig. 4. The yellow points in Image 1 are landmarks stored in map, and those in Image 2 are captured landmarks. To make the data association understandable, we stressed 3rd feature (labeled 3) of Image 1 in red cross, and the predicted landmark in Image 2 is estimated according to (10) and highlighted in red cross as well. It is clear that the prediction sufficiently approximates to the captured feature labeled in 3'. With the criterion (11), we may claim that the captured landmark 3' is same as previous feature 3. Additionally, it seems that landmark 6' in Image 2 is a new one which can also be determined by (11).

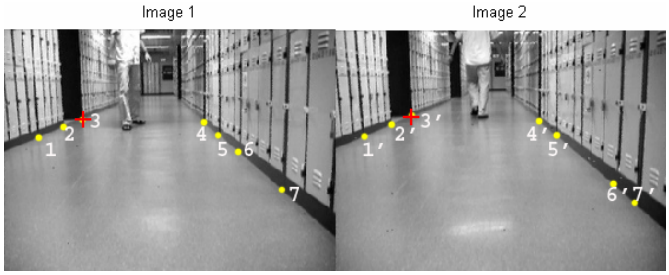


Fig.4. The example of the pixel coordinates prediction and data association based on homography transformation.

4.2 EKF SLAM Framework

In this subsection, we describe the EKF SLAM framework, and incorporate the results of the homography transformation based data association into the EKF to serve for update process and enhance the performance of SLAM. Equations (8) and (9) indicate the motion and observation model. The prediction phase of EKF is

$$\begin{aligned} \hat{X}_k &= f(X_{k-1}, u_k) \\ \hat{P}_k &= F_X P_{k-1} F_X^T + F_u Q_k F_u^T \end{aligned} \tag{12}$$

and the update phase is

$$\begin{aligned} X_k &= \hat{X}_k + K_k(z_k - \hat{z}_k) \\ P_k &= (I - K_k H_X)\hat{P}_k \\ K_k &= \hat{P}_k H_X^T \left(H_X \hat{P}_k H_X^T + R_k\right) \end{aligned} \tag{13}$$

Where F_X and F_u are the Jacobian matrices w.r.t state variable X and control variable u respectively; H_X is the Jacobian matrices w.r.t. X; R_k is the covariance of the metric measurements and is obtained approximately by linearized method with Taylor series expansion. Note that the range and bearing measurements contribute to the innovation in (13), which implies the results of the data association based on the pixel coordinates. This is because the range and bearing data is calculated from pixel information of the landmarks. If the data association has been established in pixel measurements, the relationship of the landmarks still retains identical even though in metric representation. In this viewpoint, our proposed data association methodology can help to ascertain the existed and new features, and release the complexity of the innovation computation by ranges and bearings directly.

5. EXPERIMENTAL RESULTS

To validate the virtual range and bearing extraction method and its application in the EKF SLAM, we divided the experiments into two scenarios. We first tested the virtual range finder model and then examined the performance of this virtual sensor and the data association based on homography transformation in dynamic SLAM. The mobile robot platform used for experimental studies was the Pioneer 2DX mounted with a Canon VCC4 monocular camera, a SICK LMS200 laser range finder and a sonar array.

5.1 Testing the Virtual Range Finder Model

To calculate the range and bearing from an image, the intrinsic parameters of the camera are the important components. We calibrated the camera by using the Camera Calibration Toolbox (Bouguet, Last updated April 13th, 2007) and the intrinsic parameters are listed in Table 1.

With these calibration parameters and equations (4), (6), we extracted the ranges and bearings of some random selected points in our control lab. Fig.5 shows the positions of the numbered points. Table 2 enlists the results compared with the hand-measured data. It can be seen from Table 2 that the

errors of the range are within 5% wherever these points located. However, the deviation on bearings is slightly higher, around 10%. This is because the bearing is not directly detected but computed by measuring the two sides of the angle. Another experiment in the next subsection will further validate the accuracy of the bearing data by comparing with the related laser data.

Table 1. Intrinsic Parameters

Focal length	fc=[365.12674 365.02905]
Principal point	cc=[145.79917 114.50956]
Skew factor	alpha_c=[0.00000]
Distortion factor	kc=[-0.22776, 0.36413, -0.00545, -0.00192, 0.00000]
Pixel std	err=[0.10083 0.10936]

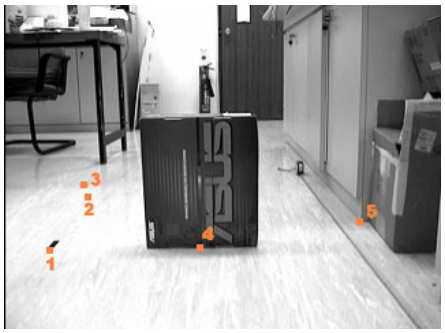


Fig.5. Positions of measured points in control lab.

Table 2. Range and Bearing Comparison

Point	Measured		Calculated		Deviation	
	r(m)	φ(rad)	r(m)	φ(rad)	r(%)	φ(%)
1	1.29	0.330	1.28	0.314	0.7	4.8
2	1.91	0.261	1.82	0.236	4.7	9.6
3	2.13	0.266	2.03	0.240	4.7	9.7
4	1.20	0	1.17	-0.02	2.5	≈0
5	1.52	-0.305	1.54	-0.330	1.3	8

5.2 Testing Application of Virtual Range Finder and the Data Association Algorithm in Dynamic SLAM

To test the performance of the virtual range finder model and our proposed data association algorithm, by using ARIA and OpenCV class library a sequence of images were collected when the mobile robot was moving with an average speed of 200mm/s in the corridor outside the control laboratory of the electrical engineering department. There were several people walking through the corridor in a typical speed around the robot accompanying with slowing down and stopping at some place. Meanwhile, the related laser information was also recorded as ground truth for model comparison and algorithm validation. After obtaining the sensor information, without considering real-time property temporarily we implemented the SLAM offline in MATLAB environment. For promoting efficiency, the landmark extraction procedure was carried out previously in an individual routine.

Fig.6 illustrates the EKF SLAM results with the virtual range finder and our proposed data association algorithm. The red asterisks are the landmarks represented by their Cartesian coordinates from the extracted range and bearing measurements, and as the reference the raw map built by laser sensor data is also overlaid in black. As is shown in Fig.6, it appears that the landmarks are almost identical with the laser data, further validating that the error of extracting bearing measurements can be reduced. The estimation errors of the robot pose are displayed in Fig.7. Most of the pose errors are well within the 3-sigma region, elucidating the performance of our proposed data association algorithm and probably indicating that the proposed methodology has robustness in dynamic environments.

In this work, we did not examine the efficiency of this model in bearing-only SLAM. However, regarding the range measurements as the auxiliary policy for landmark initialization is promising.

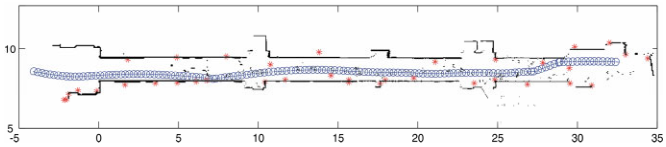


Fig.6. The EKF-SLAM results with the virtual range finder model and homography transformation based data association.

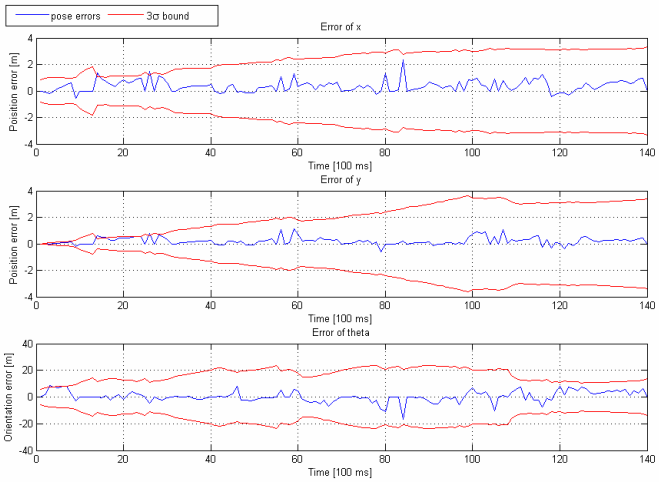


Fig.7. Estimation errors of the robot pose.

6. CONCLUSIONS

A virtual range finder model based on the monocular camera is proposed in this work. This model alleviates the deviation problem caused by unparallel motion of the optical axis, and it can offer a supplementary range measurement for bearing-only SLAM. As the uncertainty of the virtual observation is non-Gaussian, it could weaken the performance of the data association implied in update procedure of EKF. Therefore we proposed a new data association technique based on homography transformation within a sequence of images and incorporated it into the update process. Experiments validated the performance of the virtual range finder model and the homography transformation based data association in dynamic SLAM. However, the limitation of the proposed model is that the landmarks should be on the flat-floor and the external knowledge on the camera installation is required.

In the future research, we will relax these constraints by fusing other sensor information and active vision mechanism in computer vision community. The research work on sensor fusion is ongoing, the preliminary results of which is encouraging. Another path of the future study is on an appropriate error archetype for the virtual range finder model.

REFERENCES

Andrade-Cetto, J., & Sanfeliu, A. (2002). Concurrent Map Building and Localization on Indoor Dynamic Environments. *International Journal of Pattern Recognition & Artificial Intelligence, 16*(3), 361.

Baba, M., Oda, A., Asada, N., & Yamashita, H. (2006). Depth from defocus by zooming using thin lens-based zoom model. *Electronics and Communications in Japan (Part II: Electronics), 89*(9), 53-62.

Bouguet, J.-Y. (Last updated April 13th, 2007). Camera Calibration Toolbox for Matlab. 2007, from http://www.vision.caltech.edu/bouguetj/calib_doc/index.html

Costa, A., Kantor, G., & Choset, H. (2004). *Bearing-only landmark initialization with unknown data association.* Paper presented at the Robotics and Automation, 2004. Proceedings. ICRA '04. 2004 IEEE International Conference on.

Davison, A. J., Reid, I. D., Molton, N. D., & Stasse, O. (2007). MonoSLAM: Real-Time Single Camera SLAM. *Transactions on Pattern Analysis and Machine Intelligence, 29*(6), 1052-1067.

Di Marco, M., Garulli, A., Giannitrapani, A., & Vicino, A. (2003). Simultaneous localization and map building for a team of cooperating robots: a set membership approach. *Robotics and Automation, IEEE Transactions on, 19*(2), 238-249.

Di Marco, M., Garulli, A., Giannitrapani, A., & Vicino, A. (2004). A Set Theoretic Approach to Dynamic Robot Localization and Mapping. *Autonomous Robots, 16*(1), 23-47.

Diosi, A., & Kleeman, L. (2004). *Advanced sonar and laser range finder fusion for simultaneous localization and mapping.* Paper presented at the 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Sendai, Japan.

Fitzgibbons, T., & Nebot, E. (2002, 27-29 November, 2002). *Bearing-Only SLAM using Colour-based Feature Tracking.* Paper presented at the Proc. 2002 Australasian Conference on Robotics and Automation, Auckland.

Garulli, A., Giannitrapani, A., Rossi, A., & Vicino, A. (2005). *Mobile robot SLAM for line-based environment representation.* Paper presented at the 44th IEEE Conference on Decision and Control, and the European Control Conference, CDC-ECC '05, Seville, Spain.

Gil, A., Reinoso, O., Martinez Mozos, O., Stachniss, C., & Burgard, W. (2006). *Improving Data Association in Vision-based SLAM.* Paper presented at the Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on.

Hartley, R. (2003). *Multiple view geometry in computer vision / Richard Hartley, Andrew Zisserman* (2nd ed. ed.). Cambridge [England] :: Cambridge University Press.

Ip, Y. L., Rad, A. B., Chow, K. M., & Wong, Y. K. (2002). Segment-Based Map Building Using Enhanced Adaptive Fuzzy Clustering Algorithm for Mobile Robot Applications. *Journal of Intelligent and Robotic Systems, 35*(3), 221-245.

Klippenstein, J., Zhang, H. and Wang, X. . (2007, August 5-8). *Feature Initialization for Bearing Only Visual SLAM Using Triangulation and the Unscented Transform.* Paper presented at the Proc 2007 IEEE International Conference on Mechatronics and Automation. , Harbin, China.

Lowe, D. G. (2004). Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision, 60*(2), 91-110.

Madhavan, R., & Durrant-Whyte, H. F. (2005). 2D map-building and localization in outdoor environments. *Journal of Robotic Systems, 22*(1), 45-63.

Martin, M. C. (2006). Evolving visual sonar: Depth from monocular images. *Pattern Recognition Letters, 27*(11), 1174-1180.

Miro, J. V., Dissanayake, G., & Weizhen, Z. (2005). *Vision-based SLAM using natural features in indoor environments.* Paper presented at the Intelligent Sensors, Sensor Networks and Information Processing Conference, 2005. Proceedings of the 2005 International Conference on.

Murphey, Y. L., Chen, J., Crossman, J., Zhang, J., Richardson, P., & Sieh, L. (2000). *DepthFinder, a real-time depth detection system for aided driving.* Paper presented at the Intelligent Vehicles Symposium, 2000. IV 2000. Proceedings of the IEEE.

Pfister, S. T., Roumeliotis, S. I., & Burdick, J. W. (2003). *Weighted line fitting algorithms for mobile robot map building and efficient data representation.* Paper presented at the 2003 IEEE Int. Conf. on Robotics and Automation.

Saxena, A., Chung, S. H., & Ng, A. Y. (2005). *Learning Depth from Single Monocular Images*. Paper presented at the 19th Annual Conference on Neural Information Processing Systems.

Wong, Y. K., Shi, K. L., & Chan, T. F. (1997). Object Detection by Hough Transform. *Advances in Modelling and Analysis B: Signal Processing and Pattern Recognition, 38*(1), 1-14.

Xiang, W., & Hong, Z. (2006). *Good Image Features for Bearing-only SLAM.* Paper presented at the Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on.

Zhang., Z. Y. (1999, Sept.). *Flexible Camera Calibration By Viewing a Plane From Unknown Orientations.* Paper presented at the International Conference on Computer Vision (ICCV'99), Corfu, Greece.