Using Range and Bearing Observation in Stereo-Based EKF SLAM

Yao-Chang Chen[1(✉)], Tsung-Han Lin[2], and Ta-Ming Shih[3]

[1] School of Defense Science, Chung-Cheng Institute of Technology,
National Defense University, 33551 Tauyuan County, Taiwan
joseph.chen4210@gmail.com

[2] Department of Computer Science and Information Engineering,
National Taiwan University, 10617 Taipei, Taiwan

[3] Department of Avionics, China University Science and Technology,
31241 Hsinchu County, Taiwan

Abstract. In this work, we have developed a new observation model for a stereo-based simultaneous localization and mapping (SLAM) system within the standard Extended-Kalman filter (EKF) framework. The observation model was derived by using the inverse depth parameterization as the landmark model, and contributes to both bearing and range information into the EKF estimation. In this way the inherently non-linear problem cause by the camera projection equations is resolved and real depth uncertainty distribution of landmarks features can be accurately estimated. The system was tested by real-world large-scale outdoor data. Analysis results show that the landmark feature depth estimation is more stable and the uncertainty noise converges faster than the binocular stereo-based approach. We also found minor drift in the vehicle pose estimation even after extended periods demonstrating the effectiveness of the new model.

Keywords: Simultaneous localization and mapping (SLAM) · Stereo · Inverse depth parameterization · Stereo observation model · Extended Kalman filter (EKF)

1 Introduction

Over the past years, Simultaneous Localization and Mapping (SLAM) has received extensive research interest because it serves as a basic methodology for robots moving autonomously in an unknown environment. Current methods of SLAM have found to achieve accurate mapping for extended periods of time especially by using laser sensor with well bounded result for both indoor and outdoor environments [1].

Since vision systems have properties of being low cost, lightweight and contain rich information when compared to traditional robotic sensors, like laser scanners or sonars, vision-based systems are employed in a wide range of robotic applications. These applications include object recognition, obstacle avoidance, navigation, topological global localization and, more recently, in simultaneous localization and mapping, which in this case is the so-called visual SLAM.

A. Natraj et al. (Eds.): TAROS 2013, LNAI 8069, pp. 340–352, 2014.
DOI: 10.1007/978-3-662-43645-5_37, © Springer-Verlag Berlin Heidelberg 2014

One can distinguish visual SLAM as either monocular or binocular approaches. The first remarkable work in monocular visual SLAM was done by Davison et al. [2], in which a single camera is used under the Extended Kalman filtering (EKF) framework. Since camera is a bearing-only sensor, crucial limitation of monocular SLAM is the unobservability of the scale, and this cause the scale of the map to slowly drift in large environment. On the other hand, stereo vision systems are often applied in which the absolute measurement of 3D space, especially the feature depths, can be directly estimated to avoid the scale ambiguity.

In stereo systems, the observation pair (u_1, v_1, d) contains both bearing and range information, where (u_1, v_1) is the left image coordinate and d_i is the disparity. By projecting this observation pair through the pinhole model one can get the 3D Euclidean XYZ position of landmark relative to the camera [3]. Many stereo SLAM systems, including indoor and outdoor, build the map using the 3D Euclidean representation for landmark model [3–6].

Standard pinhole model projection equations used in the vision systems with EKF framework suffers from nonlinearity [7, 8]. Due to this nonlinearity, the true uncertainty of 3D Euclidean XYZ landmark can be modeled by Gaussian only for nearby features. However, true distributions of faraway features are non-Gaussian, which makes the EKF filter estimation inconsistent [9]. Other work also tried using UKF (Unscented Kalman Filter) for a stereo system on an unmanned aerial vehicle [10] to solve the nonlinearity issue, but a better way is to find a landmark model that has a high degree of linearity. Therefore, Montiel et al. [11] proposed an inverse depth parameterization to represent the landmark model. The key concept is to parameterize the inverse depth of features relative to the camera locations from which they were first viewed directly. This way of parameterization would achieve a high degree of linearity, and furthermore, the features are initialized with no delay and can successfully estimate for both near and distant features.

The drawback of inverse depth parameterization is that the 6-D state vector representation is computational intensive. Therefore, Civera et al. [12] proposed a linearity index, that inverse depth representation can be safely converted to Euclidean XYZ form; once the depth estimate of a feature has converged. The speed of convergence therefore is important.

The most closely related work is by Paz et al. [13], in which they proposed a binocular stereo-based EKF SLAM. They combine both the inverse depth parameterization and Euclidean XYZ parameterization in the map, which alleviate the nonlinearity issue effectively. The system can map both near and far features with proper uncertainty distribution, and it can be used in large-scale outdoor environment.

However, the observation model in previous studies is bearing-only. When a 3D feature is acquired simultaneously by left and right camera images, the stereo system are treated as two bearing-only observers. In each instance, EKF update is done once for left and right camera without using any range information. Thus, range information such as disparity does not directly contribution to camera poses and map spatial location estimation. In contrast we wish to incorporate not just bearing but also range information.

In this research, we want to focus on using EKF to solve the stereo-based SLAM problem. It is essential to find an appropriate probabilistic models for observations of a stereo camera that is still consistent to the linear property. Thus, our contribution is to derive a new stereo observation model that incorporates the inverse depth parameterization with observation pair (u_1, v_1, d). In this way, both range and bearing information can be directly injected into the EKF estimation process to handling the nonlinear projection issue. In order to develop the new observation model, two Jacobians needs to be derived for the EKF framework. The first instance is in the feature initialization step and the second instance is in the feature prediction step. Based on our knowledge this is the first time this observation model is proposed in the stereo-based EKF SLAM.

2 Stereo-Based EKF SLAM System Models

2.1 State Vector Definition

Following the standard EKF-based approach of SLAM, the system state vector x consists the current estimated pose of camera and physical location of features. x will change in size dynamically as features are added to or deleted from the map.

$$x = (x_C, y_1, \ldots y_i, \ldots y_n)^T \tag{1}$$

The camera state X_C is composed by the position r^{WC} with respect to a world reference frame W, and q^{WC} quaternion for orientation, and linear and angular velocity v^W and ω^C relative to world frame W and camera frame C, respectively.

$$x_C = (\, r^{WC} \quad q^{WC} \quad v^W \quad \omega^C \,)^T \tag{2}$$

The feature y_i is defined by the inverse depth parameterization [11] using a 6-D state vector:

$$y_i = (x_i y_i z_i \theta_i \phi_i \rho_i)^T \tag{3}$$

The y_i vector encodes the ray from the first camera position from feature observed by x_i, y_i, z_i, the camera optical center, and θ_i, ϕ_i azimuth and elevation (coded in the world frame) defining unit directional vector $m(\theta_i, \phi_i)$. The feature point's depth along the ray d_i is encoded by its inverse $\rho_i = 1/d_i$.

y_i models a three-dimensional point located at

$$\begin{pmatrix} X_i \\ Y_i \\ Z_i \end{pmatrix} = \begin{pmatrix} x_i \\ y_i \\ z_i \end{pmatrix} + \frac{1}{\rho_i} m(\theta_i, \phi_i) \tag{4}$$

$$m = (\cos \emptyset_i \sin \emptyset_i, \ -\sin \emptyset_i, \ \cos \emptyset_i \cos \theta_i)^T \tag{5}$$

2.2 Motion Model

The motion model in this work describes an ego motion with 6 DOF. The camera orientation is represented in terms of quaternions, which can deal with the issue of gimbal lock in Euler angles. It is assumed to be both in constant velocity and angular velocity with a zero-mean Gaussian acceleration noise $n = \begin{pmatrix} a^W & \alpha^C \end{pmatrix}^T$ uncertainty. At each step, there is an impulse of linear velocity $V^W = a^W \Delta t$ and angular velocity $\Omega^C = \alpha^C \Delta t$, with zero mean and known Gaussian distribution.

$$x_{C_{k+1}} = \begin{pmatrix} r^{WC}_{k+1} \\ q^{WC}_{k+1} \\ v^{W}_{k+1} \\ \omega^{C}_{k+1} \end{pmatrix} = f_v(x_{C_k}, n) = \begin{pmatrix} r^{WC}_k + \left(v^W_k + V^W_k\right)\varDelta t \\ q^{WC}_k \times q\left(\left(\omega^C_k + \Omega^C\right)\varDelta t\right) \\ v^W_k + V^W \\ \omega^C_k + \Omega^C \end{pmatrix} \tag{6}$$

where $q\left(\left(\omega^C_k + \Omega^C\right)\Delta t\right)$ is the quaternion defined by the rotation vector $\left(\omega^C_k + \Omega^C\right)\Delta t$.

2.3 Stereo Observation Model

In this work, we define a new nonlinear function $h(X_C, y_i)$, which allows the prediction of the value of observations measurement $\hat{z}_i$ given the current estimatiom camera pose x_C and the ith feature y_i in the map. The observation model can be written in a general form as:

$$\hat{z}_i = \begin{bmatrix} u_{li} \\ v_{li} \\ d_i \end{bmatrix} = h(x_C, y_i) + w_{u_i v_i d_i} \tag{7}$$

The vector $\hat{z}_i$ is a observation of the feature y_i relative to the camera pose x_C, where $w_{u_i v_i d_i}$ is a vector of uncorrelated observation errors with zero mean Gaussian noise and covariance matrix $R_{u_i v_i d_i}$.

The observation model can be divided in three steps. In step 1, through inverse depth parameterization, feature y_i is transformed to Euclidean XYZ landmark representation with respect to the world reference frame W:

$$\begin{pmatrix} X_i \\ Y_i \\ Z_i \end{pmatrix} = \begin{pmatrix} x_i \\ y_i \\ z_i \end{pmatrix} + \frac{1}{\rho_i} m(\theta_i, \phi_i) \tag{8}$$

In step 2, Euclidean XYZ is transformed into camera reference frame C, while plugging into x_C camera pose:

$$h^C = \begin{bmatrix} h^C_x \\ h^C_y \\ h^C_z \end{bmatrix} = \left(R^{WC}\right)^T \left(\begin{pmatrix} X_i \\ Y_i \\ Z_i \end{pmatrix} - r^{WC} \right) \tag{9}$$

Next, using the typical pinhole camera model [3], we will project this 3D position h^C to its expected image coordinates and compute its expected disparity in the new view:

$$\hat{z}_i = \begin{bmatrix} \hat{u}_{li} \\ \hat{v}_{li} \\ \hat{d}_i \end{bmatrix} = \begin{bmatrix} \frac{h_x^C}{h_z^C} f + C_x \\ \frac{h_y^C}{h_z^C} f + C_y \\ \frac{fb}{h_z^C} \end{bmatrix} \tag{10}$$

where $(C_x C_Y)$ are the camera center in pixels, f is the focal length, b the baseline of stereo.

3 The Estimation Process of Stereo-Based EKF SLAM

3.1 The Prediction Step

In the prediction stage, the camera motion model Eq. (6) is used to produce a state prediction from the previous state. Since the camera motion model only propagates the pose of previous state, we leave the map states unchanged:

$$x(k+1|k) = f_v(x_C(k|k), n) \tag{11}$$

In the prediction of covariance, state-augmentation methods [14] is used, which results in an optimal SLAM estimate with reduced computation from cubic complexity to linear complexity, and has the form:

$$P(k+1|k) = FP_{xx}(k|k)F^T + GQ_kG^T \tag{12}$$

where $F = \frac{\partial f_v}{\partial x_C}$ is the Jacobian of $f_v()$ evaluated at the estimate $x_C(k|k)$,$Q_k$the Gaussian noise covariance and $G = \frac{\partial f_v}{\partial n}$ is the Jacobian of $f_v()$ evaluated with the noise n.

3.2 The Update Step

In the update step, an observation $z_i(k+1) = (u_{li}, v_{li}, d_i)$ of the i^th feature will be available. Stereo observation model Eq. (7) is used to form an observation prediction $\hat{z}_i(k+1|k)$ and innovation $v_i(k+1)$

$$\hat{z}_i(k+1|k) = h(x_C(k+1|k), y_i(k|k)) \tag{13}$$

$$v_i(k+1) = z_i(k+1) - \hat{z}_i(k+1|k) \tag{14}$$

and then, one can calculate the innovation covariance matrix:

$$S_i(k+1) = HP(k+1|k)H^T + R_{u_i v_i d_i}(k+1) \tag{15}$$

where H is the Jacobian of h(.) evaluated at $x_C(k+1|k)$ and $y_i(k|k)$, and the Kalman gain $K_i(k+1)$ can be obtained as

$$K_i(k+1) = P(k+1|k)H^T + S_i(k+1)^{-1} \tag{16}$$

The observation matrix $z_i(k+1)$ is used to update the predictions and form a new estimation of the state by using the standard EKF update equations:

$$x(k+1|k+1) = x(k+1|k) + K_i(k+1)v_i(k+1) \tag{17}$$

$$P(k+1|k+1) = P(k+1|k) - K_i(k+1)HP(k+1|k) \tag{18}$$

The main focus here is the innovation $v_i(k+1)$ (14), which represents the difference between the actual sensor measurement $z_i(k+1)$ and the predicted measurement $z_i(k+1|k)$, both containing range and bearing information. It means that when multiplying innovation with Kalman gain $K_i(k+1)$, both bearing and range information optimize the state estimation directly

3.3 Landmark Initialization

The initialization process includes both the feature state initial values and the covariance assignment. Therefore we use inverse depth parameterization to represent features initial values, and derived the feature initialization model, $g\left(r^{WC}_{(k+1|k+1)}, q^{WC}_{(k+1|k+1)}, z_{i(k+1|k+1)}\right)$ to describe the initial values in terms of current camera pose $r^{WC}_{(k+1|k+1)}$, $q^{WC}_{(k+1|k+1)}$ and a new sensor observation pair $z_{i(k+1|k+1)} = (u_{li}, v_{li}, d_i)$

$$y_i = g\left(r^{WC}_{(k+1|k+1)}, q^{WC}_{(k+1|k+1)}, z_{i(k+1|k+1)}\right) = (x_i y_i z_i \theta_i\ \phi_i\ \rho_i)^T \tag{19}$$

The end-point of the projection ray is taken from the camera location estimate:

$$(x_i y_i z_i)^T = r^{WC}_{(k+1|k+1)} \tag{20}$$

The feature spatial location vector (in the camera reference frame) is computed from the observation pair $z_i = (u_{li} v_{li} d_i)^T$, by rearranging stereo observation model Eq. (10), we have

$$h^C = \begin{pmatrix} (u_{li} - C_x)\frac{b}{d_i} \\ (v_{li} - C_Y)\frac{b}{d_i} \\ \frac{fb}{d_i} \end{pmatrix} \tag{21}$$

where $(u_{li} v_{li})$ are the pixels on the left image, and d_i is the horizontal disparity.

The inverse depth prior ρ_i can be computed from h^C

$$\rho_i = \frac{1}{\|h^C\|} \tag{22}$$

Using the current camera orientation estimation from the state vector, h^C can be transformed to the world reference frame and the azimuth and elevation angles are extracted as:

$$h^W = R^{WC}(q^{WC}_{(k+1|k+1)})h^C \tag{23}$$

$$\begin{pmatrix} \theta_i \\ \emptyset_i \end{pmatrix} = \begin{pmatrix} \arctan(h^W_x, h^W_z) \\ \arctan\left(-h^W_y, \sqrt{h^W_x + h^{W^2}_z}\right) \end{pmatrix} \tag{24}$$

The newly initialized feature $y_i = (x_i y_i z_i \theta_i \phi_i \rho_i)^T$ is added to the state vector $x(k+1|k+1)$.

In order to model the uncertainty of the newly initialized feature, we derived the Jacobian matrix of the functions in (19), using a first-order error propagation to approximate the distribution of the variables in (19) as multivariate Gaussians. The covariance matrix of newly feature is:

$$P_{y_i y_i}(k + 1|k + 1) = J_{xc}P_{xx}J^T_{xc} + J_R R J^T_R \tag{25}$$

R includes σ_u, σ_v, σ_d, which represents the pixel uncertainties in image $(u_{li} v_{li})$ location and disparity d_i . In our experiments, we use $\sigma_u = 1\,\text{pixel}$, $\sigma_v = 1\,\text{pixel}$, $\sigma_d = 1\,\text{pixel}$. Since R is the error covariance describing the noisy measurements of the stereo system, the uncertainty through J_R is propagated to the newly feature yi of landmark model space. J_R is the Jacobian of g(.) which is derived by the observation pair z_i. P_{xx} is the camera pose covariance matrix, representing current pose estimation uncertainty. This uncertainty through J_{xc} is propagated to the newly feature y_i of landmark model space. J_{xc} is the Jacobian of g(.) which is derivative by $r^{WC}_{(k+1|k+1)}, q^{WC}_{(k+1|k+1)}$.

4 Experimental Results

4.1 Analysis of Landmark Uncertainty

In order to validate that our proposed observation model describes the uncertainty of 3D points accurately, we have simulated an experiment where the true uncertainty of the landmark location (derived from a Monte Carlo simulation) is compared to the estimated uncertainty from Eq. (25) and the traditional Euclidean XYZ landmark model.

The actual intrinsic parameters of the stereo camera, such as the baseline, are accounted in the simulation. The origin of the left camera is set as the reference frame, with the principal axis pointing to Z and X axis pointing to the right.

Consider a landmark point in front of the left camera that is at 70 m distance along the Z axis, a Monte Carlo simulation has been performed by drawing a set of 10,000 samples from the Gaussians distributions of u_{li}, v_{li}, and d_i (assuming a standard deviation of $\sigma_u = \sigma_v = 1$ and $\sigma_d = 2$ pixels, respectively), and by projecting them through Eq. (21), yielding a set of 10,000 samples of the landmark 3D position (X Y Z). In Fig. 1 (Left), the black sample points show the true measurement uncertainty from stereo systems, green point shows the real position of the landmark point. Next, the estimated uncertainty is calculated using first-order error propagation based on our observation model Eq. (25), shown by the enclosing red lines. The traditional Euclidean XYZ model is shown by the enclosing blue lines. One can see that the red lines enclose the true uncertainty noise, while blue lines do not. In Fig. 1 (Right), histogram is used to show the uncertainty distribution. Gray rectangles show the true uncertainty of the landmark location, red rectangles show our proposed observation model uncertainty, and blue indicates the traditional Euclidean XYZ model. The red rectangles have more closely covered the true gray rectangles distribution.

In Fig. 2(a), (b), (c), black sample points indicates the real distributions with various distances, using 15 m, 30 m, 45 m respectively, and the red points shows the real position of the landmark point. The estimated uncertainty is calculated using first-order error propagation using our observation model, shown by blue enclosing lines. One can see that for any distance close or far, the uncertainty region estimated by our model accurately bounds the true uncertainty.

From the simulation result, we shown that the measure error can be accurately estimated basing on our proposed observation model, which will help the EKF filter estimation to be consistent and avoid filter divergence.

4.2 Real World Experiments

4.2.1 Dataset and Feature Points Matching

All experiments are verified by using the Karlsruhe dataset [15], a real-world, large-scale, grayscale stereo sequences. Odometry data is available from OXTS RT 3000 GPS/IMU system. An experimental vehicle is equipped with a stereo camera rig

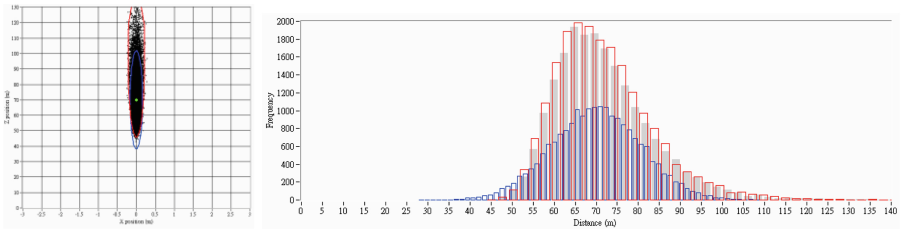
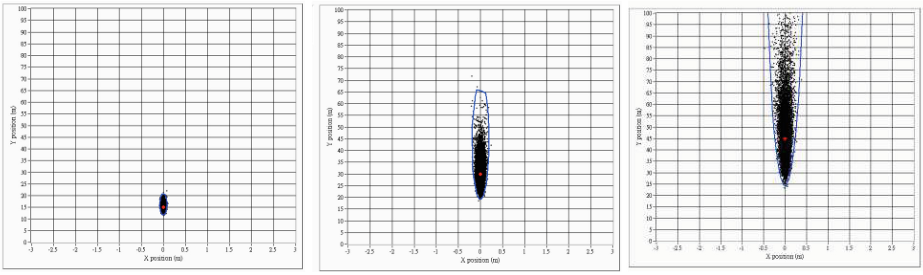


Fig. 1. (Left) The black point clouds represent the true uncertainty which are samples of distribution of a real landmark point position, given that the pixel noise in the images is Gaussian. Red line enclosed regions represent the estimated uncertainty using our proposed observation model. Blue line enclosed regions represent the estimated uncertainty using the traditional Euclidean XYZ model. (Right) The histogram is used to show the uncertainty distribution, gray rectangles show the true uncertainty of the landmark location, red rectangles show the estimated uncertainty using our proposed observation model, and blue indicates the traditional Euclidean XYZ model. Our model describes true noise distribution more accurately.



(a) (b) (c)

Fig. 2. Simulated experiment of a point reconstruction from a stereo pair observation for a point at (a) 15 m, (b) 30 m, (c) 45 m distance. The point clouds are samples of distribution of a real landmark point position, given that the pixel noise in the images is Gaussian. The black sample points show the distribution. Blue line enclosed regions represent the estimated uncertainty using our proposed observation model (color figure online).

Pointgrey Flea2 firewire cameras covering inner city traffic. All the frames are rectified with a resolution of 1344 × 391 pixels running at 10 fps. In our experiment we set the left image frame as the reference coordinate system. Our algorithm is implemented by LabVIEW on an Intel Core i5 with 1.7 GHz and 4 GB RAM computer.

In order to obtain the matching stereo observation pairs, stereo points matching based on LIBVISO2 [16], an open-source library is used to demonstrate real-time computation of point feature match of left and right images. The matching stereo pairs are done only in the first frame for initialization.

We use a different method to track the initialized points in the subsequent frames. When each stereo observation pair is initialized and saved in the map, it also save corresponding 11 × 11 surrounding patch, which serves as a photometric identifier. When the next image comes, the saved pairs in 3D space are projected back to image plane. The pairs are deleted if it is outside the visible field of view of left and right images. If they are within the field of view, we use active search concept to decide the searching region, in which the region size is determined by the EKF innovation covariance. The corresponding patch of the features first will be warped according to the predicted camera motion, and then normalized cross-correlation is performed in the searching region to find the matching point, for details see [17].

4.3 Analysis of Features Location Estimation and Stability

In this part of the experiment, we want to compare the two monocular observers based visual SLAM systems and our proposed stereo observer based visual SLAM systems with regard to the accuracy of landmark features spatial location estimation.

We used a subset of the Karlsruhe dataset, the 2009_09_08_drive_0010 from frame 61 to frame 72. The scenario is that the vehicle is moving forward on the road, and then turns right. The features are initialized first in frame 61 (Fig. 3 Left), and then were continuously tracked until frame 72 (Fig. 3 Right). Therefore the camera

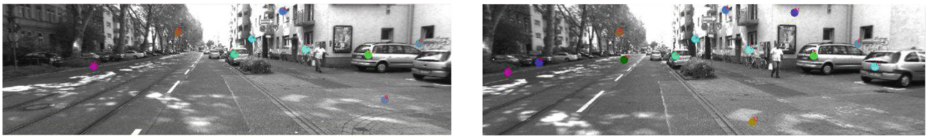


Fig. 3. (Left) At frame 60 features are initialized. (Right) At frame 70, features tracking result.

poses and features locations were updated 10 times. We recorded every state vector and covariance matrix along the way, and select a near (number 7) and far (number 0) feature point for analysis.

In Fig. 4 (Left) and (Right) each time the feature updates its depth uncertainty, no matter the feature is either far or near, it can be seen that our proposed method depth and uncertainty converge faster.

Figure 5 (Left) and (Right) shows the raw measurement of each feature's depth (green points), and the depth estimation. The blue points shows the two monocular observers based visual SLAM systems estimation result, and the red points shows the stereo observer based visual SLAM systems estimation result.

From Fig. 5 it shows that our proposed method estimation is more stable, therefore the curve is smoother. Also after 10 updates, the estimation result is also closer to the mean values of raw measurement. After 10 updates, only the estimation of near feature is closer to the raw measurement mean values, Fig. 5 (Left), while the features far away cannot get close to the raw measurement mean values Fig. 5 (Right).

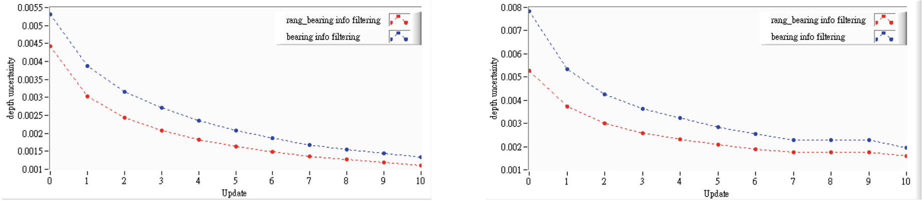


Fig. 4. (Left) Depth uncertainty of feature number 7, over 10 times EKF updates, (Right) depth uncertainty of feature number 0, over 10 times EKF updates.

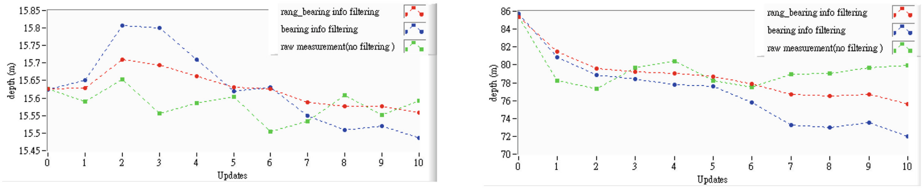


Fig. 5. (Left) The depth estimated value of feature number 7 (close by), over 10 times EKF updates, (Right) the depth estimated value of feature number 0 (far away), over 10 times EKF updates. The proposed model (red line) is more stable and closer to raw measurement mean (color figure online).

4.4 Motion Estimate Results

To evaluate how the capability of the proposed algorithm can correctly estimate the camera pose and velocity in a large-scale environment after extended periods of time, we use one of the Karlsruhe dataset, the 2009_09_08_drive_0015 images. Within the dataset we pick out 800 stereo frames, with a total length of the route approximately 350 m. These images are sequential scenes of inner city urban driving, including two wide angle turns.

Figure 6 depicts the trajectory estimated by our visual SLAM algorithm (in red) and 'groundtruth' output of a OXTS RT 3003 GPS/IMU system (in green). Note that the GPS/IMU system can only be considered as 'weak' groundtruth [16], because localization errors of up to two meters may occur in inner-city scenarios due to limited satellite availability. Thus instead, LIBVISO2 [16] visual odometry (in black) is argued to be a better groundthruth. As we can see, our proposed model estimated trajectory is quite close to the 'groundtruth'.

From Fig. 6, because the car at frame 250 made a negative x-directional turn, yaw angle started to increase drastically. This cause the forward velocity Vz to decrease drastically, while the negative x-direction Vx started to increase. Afterward at frame 600, both Vz and Vx started to decrease, and then at frame 650 the car turns toward the z-axis direction, making the yaw decreases. Finally the car continues toward the z-axis direction while Vz velocity started to increase.

From the experimental result, we can see that the proposed algorithm can be used in large-scale outdoor scenario, that the car pose and velocity can be decently estimated quite well, close to the groundtruth.

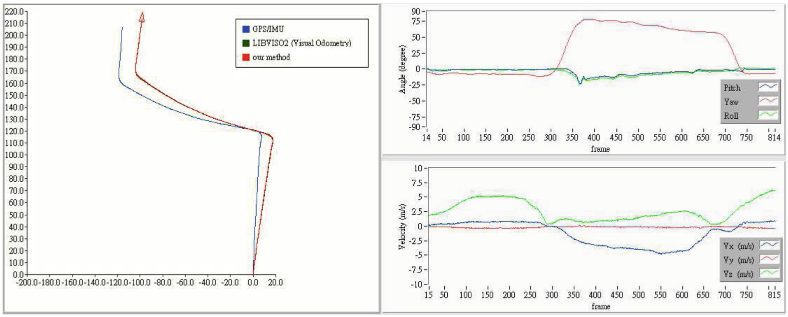


Fig. 6. (Left) Depicts the trajectory estimated by our visual SLAM algorithm (red) and the trajectory using the OXTS RT 3003 GPS/IMU system provided by the Karlsruhe dataset. (Right) The progression of vehicle orientation and velocity, estimated by our visual SLAM algorithm. Our proposed method has estimated well in large-scale scenario (color figure online).

5 Conclusions and Future Works

The focus of this work was to develop a new probabilistic observation model in EKF-based stereo SLAM. The statistical behavior in stereo vision is known to have inherently non-linear problem. Therefore, we use inverse depth parameterization as the landmark model to deal with the non-linear problem, and in contrast to the binocular stereo-based approach, we used stereo observation pair (u, v, d) to derive a new observation model, allowing both bearing and range information to be incorporated into the EKF estimation process. Furthermore, our new observation model is also computationally faster than the previously mentioned binocular stereo-based approach. In our approach, we only have to do projection and update in the EKF framework once, in contrast to the binocular approach which requires projection and update two times each. Moreover, because update step requires inverse of large innovation matrix, doing update step twice would be computationally intensive in large-scale SLAM. Based on our knowledge this is the first time this observation model with inverse depth parameterization is proposed in the stereo-based EKF SLAM.

From our experiments, it shows that even in large-scale outdoor environment, there is only a little 'scale drift', which means that our observation model together with inverse depth parameterization has kept true to the real noise distribution of the landmark feature. This also means the proposed observation models that is designed with the additional range information has helped and worked consistently under the strict requirement of EKF framework. We also demonstrated that the proposed system has converged faster and has more stable depth estimation from experiment data. These convergence and stableness characteristics will be critical to our future work. For future work we want to develop moving objects detection along with static map into a nice single EKF framework.

References

1. Cole, D.M., Newman, P.M.: Using laser range data for 3D SLAM in outdoor environments. In: Proceedings of IEEE International Conference on Robotics and Automation (ICRA), pp. 1556–1563 (2006)
2. Davison, A.J.: Real-time simultaneous localisation and mapping with a single camera. In: Proceedings of Ninth IEEE International Conference on Computer Vision, vol. 2, pp. 1403–1410. IEEE (2003)
3. Se, S., Lowe, D., Little, J.: Mobile robot localization and mapping with uncertainty using scale-invariant visual landmarks. Int. J. Robot. Res. **21**(8), 735–758 (2002)
4. Herath, D., Kodagoda, S., Dissanayake, G.: Simultaneous localisation and mapping: a stereo vision based approach. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 922–927 (2006)
5. Lemaire, T., Berger, C., Jung, I.-K., Lacroix, S.: Vision-based SLAM: stereo and monocular approaches. Int. J. Comput. Vision **74**(3), 343–364 (2007)
6. Berger, C., Lacroix, S.: Using planar facets for stereovision SLAM, intelligent robots and systems. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), vol. 22, no. 26, pp. 1606–1611 (2008)
7. Sibley, G., Matthies, L., Sukhatme, G.: Bias reduction and filter convergence for long range stereo. In: 12th International Symposium of Robotics Research (ISRR), San Francisco, CA, USA (2005)
8. Sibley, G., Sukhatme, G., Matthies, L.: The iterated sigma point filter with applications to long range stereo. In: Robotics: Science and Systems II, Cambridge, USA (2006)

9. Tim, B., Nieto, J., Guivant, J., Stevens, M., Nebot, E.: Consistency of the EKF-SLAM algorithm. In: International Conference on Intelligent Robots and Systems (IROS), Beijing, China (2006)
10. Li, X., Aouf, N., Nemra, A.: 3D mapping based VSLAM for UAVs. In: 2012 20th MediterraneanConference on Control & Automation (MED), pp. 348–352 (2012)
11. Montiel, J.M.M., Civera, J., Davison, A.J.: Unified inverse depth parametrization for monocular slam. In: Robotics: Science and Systems, Philadelphia, USA, August 2006
12. Civera, J., Davison, A.J., Montiel, J.M.: Inverse depth to depth conversion for monocular SLAM. In: Proceedings of IEEE International Conference on Robotics and Automation, pp. 2778–2783. IEEE (2007)
13. Paz, L., Piniés, P.: Large-scale 6-DOF SLAM with stereo-in-hand. IEEE Transactions on Robot. **24**(5), 946–957 (2008)
14. Durrant-Whyte, H., Bailey, T.: Simultaneous localization and mapping: part I. IEEE Robot. Autom. Mag. **13**(2), 99–110 (2006)
15. Karlsruhe Dataset. http://www.cvlibs.net/datasets/karlsruhe_sequences.html
16. Geiger, A., Ziegler, J., Stiller, C.: StereoScan: dense 3d reconstruction in real-time. In: IEEE Intelligent Vehicles Symposium (IV), (Iv), pp. 963–968 (2011)
17. Civera, J., Davison, A.J., Martínez, M.: Structure from motion using the extended Kalman filter. In: Civera, J., Davison, A.J., Montiel, J.M.M. (eds.) Springer Tracts in Advanced Robotics, vol. 75. Springer, Heidelberg (2012)