Proceedings of the 35th Chinese Control Conference
July 27-29, 2016, Chengdu, China

Visual Odometry Based on Improved Feature Matching and Unscented Kalman Filter

YU Huan, XIE Ling, Chen Jiabin, SONG Chunlei, Fei Guo

School of Automation, Beijing Institute of Technology, Beijing 100081, China
E-mail: 846363719@qq.com

Abstract: In this paper, we present an improved vision-based navigation method and proposed an improved feature matching method for improving the matching accuracy. In the matching process, we divide it into two steps, coarse and fine matching. During the coarse matching step, we adopt SURF feature detector for feature detection and Fast Library for Approximate Nearest Neighbors for feature matching, and then use the constraints of epipolar geometry, major orientation of feature points, and the uniqueness of feature matching to roughly eliminate error matching. In the fine matching process, Random Sample Consensus method with outlier rejection is employed, which will reduce the effects on motion estimation by moving objects in the scenes. The visual odometry algorithm is based on trifocal geometry, which is no need for the reconstruction of the 3d object points. Finally, we employ Unscented Kalman Filter for ego-motion estimation, which is better than Extended Kalman Filter and the experimental result shown that it can fully adapt to environment with high uncertainty. The experimental results prove that the method proposed in this paper is superior to other algorithm in terms of positioning precision.

Key Words: Vision-based navigation, SURF feature detector, Unscented Kalman Filter

1 Introduction

Navigation and positioning technology is the key technology for intelligent vehicle to perform specific tasks. In recent years, vision-based navigation has made rapid development. Vision navigation is entirely autonomous and no need for auxiliary information. When GPS is compromised and IMU drift is serious, and especially in slippery terrain where speedometer yields wrong data, vision navigation will play a key role [1][2][3]. In this paper, the vision positioning is based on binocular stereo camera and we calculate the six degrees of freedom of a car by the stereo camera. However, at present, with regard to this kind of Visual Odometry(VO) there are still exists many issues: 1. The existence of dismatch, which will result low precise motion estimation. 2. Nonuniform distribution of feature points in the space, so features in a small area may influence the overall system performance. 3. Many visual odometry algorithm is based on feature points, which are extracted from static objects. But our environment always exists many mobile objects [4]. To realize high accuracy, strong robustness and real-time visual odometry, domestic and international scholars have done a lot of researches mainly from two aspects, which are hardware and software. DSP, FPGA and GPU accelerator are employed to improve the real-time calculating capacity in terms of hardware [5]. In the studies of algorithm of visual odometry, high accuracy feature points and lines matching methods, 3D-Construction, and filtering techniques are used for high accuracy motion estimation[4][6]. To further improve location precision, GPS, IMU and Lidar information can be fused with vision data, which will reduce position error in some special scenarios [7][8].

In this paper, to solve the issues mentioned above we improved the currently common used method by using a. two-steps matching method and nonlinear filtering technology, which improved the accuracy of the feature points matching and motion estimation a lot. In the experimental part, we test the algorithm proposed in this paper on the KITTI benchmark [9]. The epipolar of images in the dataset have been rectified, and SURF detector is employed for the feature extraction [10]. We first use Fast Library for Approximate Nearest Neighbors (FLANN) for feature matching and then remove obvious mismatch according to constrains of parallel epipolar, orientation of feature points, and the uniqueness of feature points matching[11]. For the feature points of independently moving objects, Outliers are eliminated by means of a Random Sample Consensus (RANSAC) based rejection plan[12]. We employ trifocal geometry as the mathematical framework and then use Unscented Kalman Filter for robust nonlinear motion estimation [13][14]. At the end of paper, we adopt two image sequences for algorithm verification and the results shown that the method proposed above can improve the current visual odometry algorithm a lot.

The structure of the paper is the following. In section 2, we first briefly introduce vision positioning theory and review the trifocal tensor. Then, the two steps feature matching method and motion estimation method based on UKF are introduced. In section 3, the ego-motion estimation results are shown in figures. Finally, section 4 gives out the conclusion of the paper.

2 Basic Theory of the Algorithm

2.1 Geometrical Model of the Visual System

In the experiments of this paper, we firstly rectify the images collected from two calibrated cameras before feature detection and motion estimation. The images corrected

*This work is supported by National Natural Science Foundation of China under Grant 91120010.

according to the epipolar line can be seen as images created by two parallel optical axis cameras [15]. The imaging model is as the Fig. 1 [10].

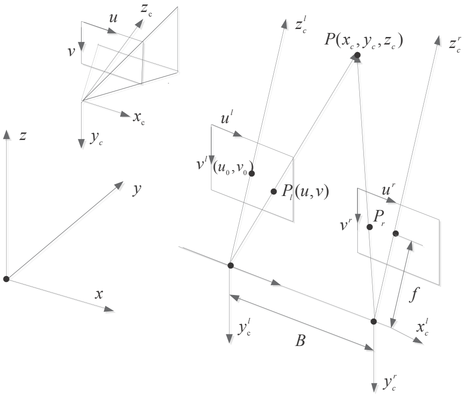

Fig. 1: Imaging model and Different Coordinates system

Supposed the M is the 3x3 calibration matrix which contains intrinsic parameters of camera, then the transformation relationship of a feature point in the camera coordinate to the image coordinate is:

$$\bar{x} = \left(u, v, w\right)^T = M \cdot X_c \tag{1}$$

Transformation from Camera coordinate to the world coordinate can be realized by rotation matrix R and translation vector t. Let the coordinate of a feature points in the world coordinate is $X_W = \left(X_W, Y_W, Z_W\right)^T$ and in the camera coordinate is $X_C = \left(X_C, Y_C, Z_C\right)^T$, then

$$X_C = R \cdot X_W + t \tag{2}$$

According to (1) and (2):

$$\bar{x} = P \cdot X_W \tag{3}$$

Here, $P = K \cdot [Rt]$ is the 3x4 projection matrix [13].

2.2 Trifocal Geometry

A trifocal tensor employ a 3x3x3 matrix describe the geometrical relationship of three pictures. The projection matrices are $P_1 = K_1 \cdot [R_1 | t_1]$, $P_2 = K_2 \cdot [R_2 | t_2]$, and $P_3 = K_3 \cdot [R_3 | t_3]$, and then the trifocal tensor is

$$T_i^{qr} = (-1)^{i+1} \cdot \det \begin{pmatrix} a^i \\ b^j \\ c^k \end{pmatrix} \tag{4}$$

Where a^i represents that the i-th row is removed from P_1. b^q and c^r denotes P_2, P_3 without the j-th and k-th row respectively[13].

According to the calculated trifocal tensor, the relationship of the matched points of m_1, m_2, m_3 can be described as follows:

$$m_3^k = m_1^k \cdot l_{2,j} \cdot T_i^{jk} \tag{5}$$

We use the rotation matrix $R(\Theta, \Phi, \Psi)$ and the translation vector $t = (t_X, t_Y, t_Z)^T$ to describe the relationship between the camera coordinate and the world reference coordinate. If ΔT the time difference of two consecutive frames is known, the ego-motion is given by:

$$t = (V_X \cdot \Delta T, V_Y \cdot \Delta T, V_Z \cdot \Delta T)^T \tag{6}$$

$$R(\omega_Z \cdot \Delta T, \omega_X \cdot \Delta T, \omega_Y \cdot \Delta T) \tag{7}$$

Where V_i and ω_i represent translational and rotational velocities, respectively.

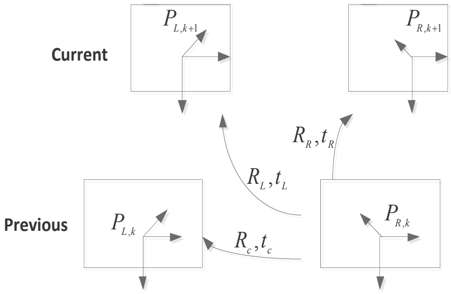

Fig. 2: Motion estimation sequence

According to the intrinsic and extrinsic parameters of the camera, and let the previous right camera as the world reference frame, then Fig. 2 shows the transformation and the projection matrices are given by:

$$P_{R,k} = K_R \cdot [I | 0] \tag{8}$$

$$P_{L,k} = K_L \cdot [R_C | t_C] \tag{9}$$

$$P_{R,k+1} = K_R \cdot [R_R | t_R] \tag{10}$$

$$P_{L,k+1} = K_L \cdot [R_L | t_L] \tag{11}$$

By equation (4), we get:

$$T_R = T(K_R, K_L, R_C, t_C, R_R, t_R, \Delta T) \tag{12}$$

$$T_L = T(K_R, K_L, R_C, t_C, R_L, t_L, \Delta T) \tag{13}$$

T_R and T_L denote the trifocal tensor between the two images of previous frame to the image of current right and left camera image. According to definitions above, we have the mapping relationship of feature points. Feature points of previous frame $m_{R,k} \leftrightarrow m_{L.k}$ are mapped to the current frame via non-linear mapping:

$$m_{R,k+1} = h_R(T_R, m_{R,k}, m_{L,k}) \tag{14}$$

$$m_{L,k+1} = h_R(T_L, m_{R,k}, m_{L,k}) \tag{15}$$

2.3 Two-steps Feature Points Matching

In the feature points detection, we choose SURF detector for its good robustness, movement and rotation invariant. After the feature detection, a two-steps feature matching method is employed, which includes coarse and fine matching. We can refer some papers about SURF detector [10].

1. Coarse matching

We get feature points descriptor based on SURF feature detection, and then Fast Library for Approximate Nearest Neighbors (FLANN) is employed for the first step feature points matching. Muja and Lowe first proposed FLANN algorithm in 2009 [11], which is based on K-means tree or KD-tree and is an excellent method for matching feature points described with a 64-dimensional matrix.

For the images have been corrected by epipolar, we can remove some wrong matching points according parallel epipolar constraint. Under this restriction, if the vertical coordinates of two images differ by two pixel value, we define this match is false. Then under the uniqueness constraint of matching, if one point corresponds to multiple points, they are wrong matching. Finally, in SURF feature descriptor, the major orientation of a point is defined as Fig. 3. If the major orientation of two feature points has 5 degrees difference, this matching will be abandoned.

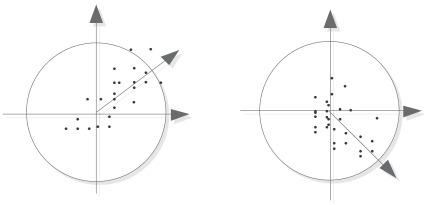

Fig. 3: Major orientation of a feature point

2. Fine matching

Fine matching process is based on Random Sample Consensus (RANSA), which can reduce the effects on motion estimation by moving objects in the scenes. The motion estimation method based on RANSAC firstly choose a subset from the matched points, then estimate the ego-motion based on this data, and the number of points is given by

$$n = \frac{\log(1 - p)}{\log(1 - (1 - \varepsilon)^s)} \qquad (16)$$

Here, s is the minimum quantity needed for calculation. P represent the probability that one or more sample contains inliers and ε is the assumed percentage of outliers in the data. We get all inliers using Euclidean re-projection error. When the Euclidean re-projection error is lower than a certain threshold, a feature is an inlier. To get the final ego-motion estimation, we use all inliers of the best sample after the iterative process.

The two-step feature matching process yields a robust estimation of vehicle even in the presence of moving objects.

2.4 Unscented Kalman Filter and Motion Estimation

The Kalman filtering is a linear minimum variance estimation, including a prediction step and an update step. Kalman filter is the best filter for systems with zero-mean white noise. To non-Gaussian state space, we can get linearized model by implementing Tayor expansion, which is called Extended Kalman Filter. The linearization method of EKF cannot estimate the mean and variance of the Guassian random vector precisely after nonlinear transmission. So the filtering precision is low, especially in the systems with high nonlinear properties. Additionally, we must solve Jacobian matrix in EKF, which restricts the application of EKF [15]. To solve those problems, Julier proposed Unscented Kalman filtering, which use Unscented Transform in the filter. For the unscented transform integrates information of higher order moments in the estimation process [17].

If the discrete time nonlinear system model is given by

$$X_{k+1} = f(X_k, u_k, W_k) \qquad (17)$$

$$Z_k = h(X_k, V_k) \qquad (18)$$

Where $X_k \in R^n$ is the system state; $f(\cdot)$ represents the N-dimensional vector function; $h(\cdot)$ is the M-dimensional vector function; W_k denotes the processing noise; V_k is the measurement noise. The calculation process of UKF is as follows:

Step1: parameters initialization

$$\hat{X} = E[X_0] \qquad (19)$$

$$P_0 = E[(X_0 - \hat{X}_0)(X_0 - \hat{X}_0)^T] \qquad (20)$$

$$\hat{X}_0^a = E[X_0^a] = \begin{bmatrix} \hat{X}_0^T & \overline{W}_0^T & \overline{V}_0^T \end{bmatrix}^T \qquad (21)$$

$$P_0^a = \begin{bmatrix} P_{X_0} & 0 & 0 \\ 0 & R_W & 0 \\ 0 & 0 & R_V \end{bmatrix} \qquad (22)$$

Step2: Sigma points ξ_{k-1}^i (i=1,2,…,2n) calculation

$$\xi_{k-1}^{(0)} = \hat{X}_{k-1}^a \qquad (23)$$

$$\xi_{k-1}^{(i)} = \hat{X}_{k-1}^a + (\sqrt{(n+\lambda)P_{k-1}^a}),\ i = 1, 2, 3...n \qquad (24)$$

$$\xi_{k-1}^{(i)} = \hat{X}_{k-1}^a - (\sqrt{(n+\lambda)P_{k-1}^a}),\ i = n+1, n+2..., 2n \qquad (25)$$

Step3: Time updating

$$\xi_{k-1}^{(i)} = f(\xi_{k-1}^{(i)}),\ i = 0, 1, 2...2n \qquad (26)$$

$$\hat{X}_{k,k-1}^a = \sum_{i=0}^{2n} \omega_i^{(m)} \xi_k^{(i)} \qquad (27)$$

$$P_{k,k-1}^a = \sum_{i=0}^{2n} \omega_i^{(c)} (\xi_k^{(i)} - \hat{X}_{k,k-1}^a)(\xi_k^{(i)} - \hat{X}_{k,k-1}^a)^T + P_{k-1}^a \qquad (28)$$

Step4: Measurement updating

$$\chi_{k.k-1} = h(\xi_{k,k-1}^a) \qquad (29)$$

$$\hat{Z}_{k,k-1} = \sum_{i=0}^{2n} \omega_i^{(m)} \chi_{i,(k,k-1)} \qquad (30)$$

$$P_{\tilde{Z}_K} = \sum_{i=0}^{2n} \omega_i^{(m)} (\chi_{i,(k,k-1)} - \hat{Z}_{k,k-1})(\chi_{i,(k,k-1)} - \hat{Z}_{k,k-1})^T \qquad (31)$$

$$P_{X_k Z_K} = \sum_{i=0}^{2n} \omega_i^{(m)} (\xi_{i,(k,k-1)} - \hat{X}_{k,k-1}^a)(\chi_{i,(k,k-1)} - \hat{Z}_{k,k-1})^T \qquad (32)$$

$$K_k = P_{X_k Z_K} P_{\tilde{Z}_K}^{-1} \qquad (33)$$

$$\hat{X}_k^a = \hat{X}_{k,k-1}^a + K_k (Z_k - \hat{Z}_{k,k-1}) \qquad (34)$$

$$P_{ak} = P_{k,k-1}^a - K_k P_{\tilde{Z}_K} K_k^T \qquad (35)$$

In this paper, the state is defined as:

$$X = (V_X, V_y, V_z, \omega_X, \omega_y, \omega_Z)^T \qquad (36)$$

Where V_i and ω_i represent translational and rotational velocities, respectively. The relations of feature coordinates in the current frames and the motion is given by:

$$m_{R,k+1} = h_R(T_R, m_{R,k}, m_{L,k}) \qquad (37)$$

$$m_{L,k+1} = h_L(T_L, m_{R,k}, m_{L,k}) \qquad (38)$$

In our case, we choose the discrete-time system model as follows:

$$X_{k+1} = f(X_k) + \omega_k \qquad (39)$$

$$Z_{k+1} = h(X_{k+1}) + \upsilon_{k+1} \qquad (40)$$

Where $Z_{k+1} = [u_{r,k+1,1}, \ldots, v_{L,k+1,N}]^T$ represents the 4N-dimensional measurement vector. The state error covariance of ω_k is a 6×6 matrix Q_k and the measurement error covariance is a 4Nx4N matrix R_{k+1}. N is the amount of matched feature pairs.

2.5 Experimental Results

In this section, we evaluate the proposed approaches by an off-the-shelf dataset from KITTI Vision Benchmark Suit, which provides dataset with ground truth and collects data from GPS, Gyroscope and Stereo camera with a resolution of 1344x391 and 10 frames/s [16].

Firstly, the comparison of common feature matching method and the proposed two-steps method is given in the following Figures. The result is based on the OpenCV library, which is a cross platform computer vision/image processing algorithms library. Fig. 4 is the matching result of SURF and FLANN; Fig. 5 represents the false matching according to the constraints we have mentioned; Fig. 6 is the final result we get, According to the parallel epipolar constrains, which obviously has higher matching precision than the first matching.

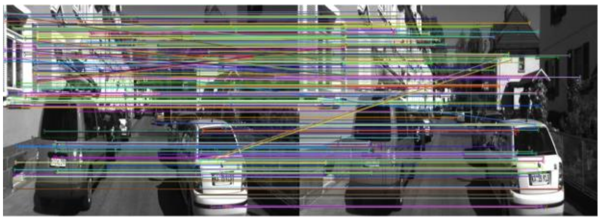

Fig. 4: Surf feature detection and FLANN feature matching

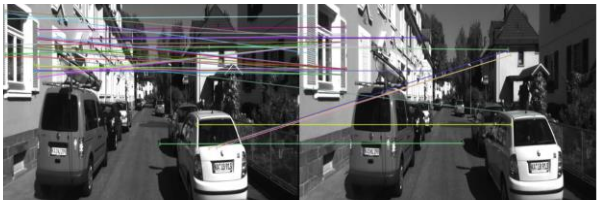

Fig. 5: False matches of FLANN matching

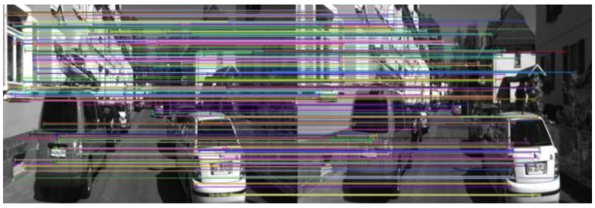

Fig. 6: Coarse matching based on matching constraints

Secondly, we choose two sequences of the dataset, which are sequence 04 2011_09_30_drive_0016 and sequence 07 2011_09_30_drive_0027.The first is a straight road about 402.52 meters and the trajectory of the latter is a curve. In Fig. 7 and Fig. 8, the ego-motion trajectory of the vehicle are estimated by three algorithms. In Fig. 7, the red line is the ground truth, which is measured by GPS and IMU; the estimation based on two steps feature matching and UKF is shown as the blue line, which is defined as UKF-Vision is close to the reference line; the green line is estimated by EKF-Vision.

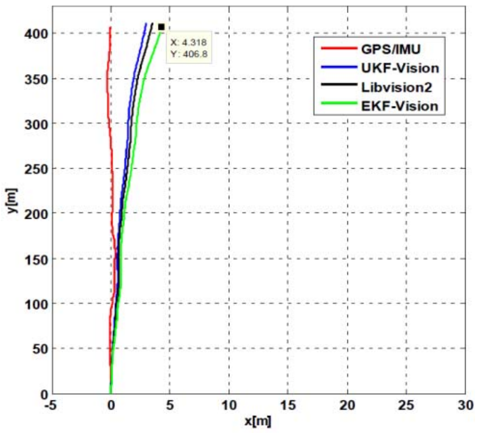

Fig. 7: Motion estimation of sequence 04

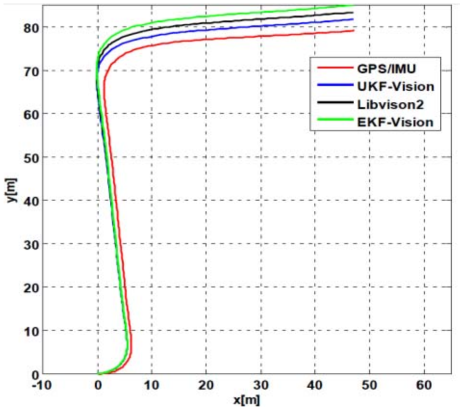

Fig. 8: Motion estimation of Sequence 07

The position error is shown in Table 1. We can obviously see that the method proposed by the paper get the best accuracy, both on straight road and on curve road.

Table 1: The localization error of different methods

Paper Size	Straight Road	Curve Road
Length(m)	402.52	132.76
UKF-Vision	0.76%	1.55%
Libvision2	0.98%	3.03%
EKF-Vision	1.08%	3.93%

3 Conclusion

This paper presents a stereo visual odometry algorithm based on two steps feature matching and Unscented Kalman Filter. The odometry algorithm is based on trifocal geometry. The experiment in the paper denotes that the coarse and accurate matching process can improve the matching accuracy and robustness a lot, which is the precondition of precisely estimating the motion of the car. The Extended Kalman Filter cannot deal well with non-linear measurement equation. UKF incorporates higher order moments in the estimation process, which shows better estimation result than EKF.

Finally, the experimental result demonstrate that the algorithm proposed in this paper yields a promising results in the KITTI dataset with 0.76% and 1.55% position error. The outcome also shows that a turn of a moving car will have a great impact on motion estimation, which also will be a future work for further improving vision-based navigation.

References

[1] Yang Cheng; Maimone, M.; Matthies, L., "Visual odometry on the Mars Exploration Rovers," in Systems, Man and Cybernetics, 2005 IEEE International Conference on , vol.1, no., pp.903-910 Vol. 1, 10-12 Oct. 2005.

[2] Indelman V, Gurfil P, Rivlin E, et al. Real-Time Vision-Aided Localization and Navigation Based on Three-View Geometry[J]. Aerospace & Electronic Systems IEEE Transactions on, 2012, 48(3):2239-2259.

[3] Feng G, Wu W, Wang J. Observability Analysis of a Matrix Kalman Filter-Based Navigation System Using Visual/Inertial/Magnetic Sensors[J]. Sensors, 2012, 12(7):8877-8894.

[4] Zhang F, Clarke D, Knoll A. Visual odometry based on a Bernoulli filter[J]. International Journal of Control Automation & Systems, 2015, 13(3):530-538.

[5] Wei Lu; Zhiyu Xiang; Jilin Liu, "High-performance visual odometry with two-stage local binocular BA and GPU," in Intelligent Vehicles Symposium (IV), 2013 IEEE , vol., no., pp.1107-1112, 23-26 June 2013.

[6] Kong X, Wu W, Zhang L, et al. Tightly-Coupled Stereo Visual-Inertial Navigation Using Point and Line Features[J]. Sensors, 2015, 15(6):12816-12833.

[7] Perlin A V E, Johnson D B, Rohde M M, et al. Fusion of visual odometry and inertial data for enhanced real-time egomotion estimation[J]. Proc Spie, 2011, 8045(23):2155-2157.

[8] Hesch J A, Kottas D G, Bowman S L, et al. Camera-IMU-based localization: Observability analysis and consistency improvement[J]. International Journal of Robotics Research, 2014, 33(1):182-201.

[9] Kitt, B.; Geiger, A.; Lategahn, H., "Visual odometry based on stereo image sequences with RANSAC-based outlier rejection scheme," in Intelligent Vehicles Symposium (IV), 2010 IEEE , vol., no., pp.486-492, 21-24 June 2010.

[10] Bay H, Tuytelaars T, Gool L V. SURF: Speeded Up Robust Features.[J]. Computer Vision & Image Understanding, 2006, 110(3):404-417.

[11] Muja M. Fast Approximate Nearest Neighbors with Automatic Algorithm Configuration.[C]// In Visapp International Conference on Computer Vision Theory & Applications. 2009:331--340.

[12] Scaramuzza D, Fraundorfer F, Siegwart R. Real-time monocular visual odometry for on-road vehicles with 1-point RANSAC[C]// IEEE International Conference on Robotics & Automation. 2009:4293-4299.

[13] Hartley R, Zisserman A. Multiple view geometry in computer vision[J]. Cambridge University Press, 2003, 30(9-10):1865 - 1872.

[14] Julier S J, Uhlmann J K. Unscented filtering and nonlinear estimation[J]. Proceedings of the IEEE, 2004, 92(3):401-422.

[15] Zhang F, Clarke D, Knoll A. Visual odometry based on a Bernoulli filter[J]. International Journal of Control Automation & Systems, 2015, 13(3):530-538.

[16] Li, Mingyang, Mourikis, Anastasios I. High-precision, consistent EKF-based visual – inertial odometry[J]. International Journal of Robotics Research, 2013, 32(6):690-711.

[17] Geiger A, Lenz P, Stiller C, et al. Vision meets robotics: The KITTI dataset[J]. International Journal of Robotics Research, 2013, 32(11):1231-1237.

[18] Julier S J, Uhlmann J K. Unscented filtering and nonlinear estimation[J]. Proceedings of the IEEE, 2004, 92(3):401-422.