Depth Enhanced Visual-Inertial Odometry Based on Multi-State Constraint Kalman Filter

Pang Fumin[1,2], Chen Zichong[1], Pu li[1], and Wang Tianmiao[2]

Segway Robotics Inc.[1]

Beihang University[2]

2017.9.25

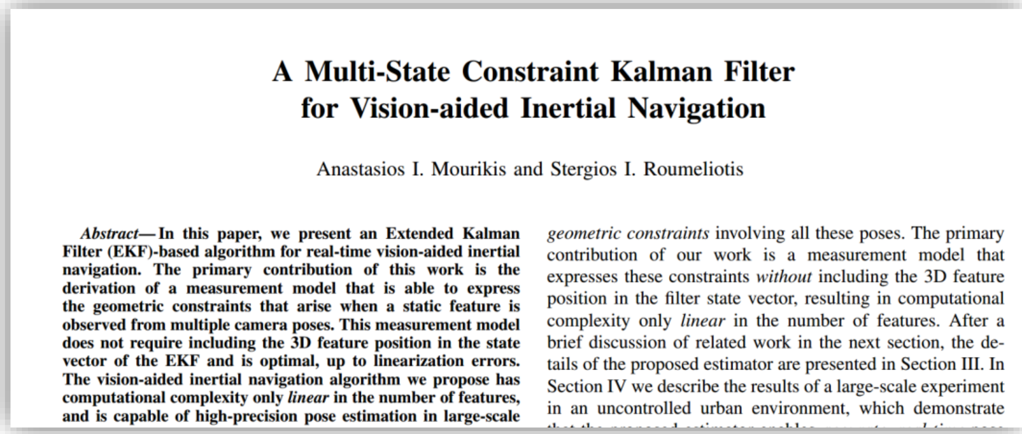
2017 IEEE/RSJ International Conference on Intelligent Robots and Systems. Vancouver, Canada, September 24–28, 2017

Content

1. A simple review of MSCKF (Multi-State Constraint Kalman Filter)
2. Our previous work and the proposed Depth enhanced MSCKF
3. Experiment
4. Conclusion

Review of MSCKF

1.Multi-State Constraint Kalman Filter is a Extended Kalman filter (EKF)-based methods for **VIO(Visual-Inetial Odometry).**
2. Methods utilizing iterative minimization over a window of states.



A Multi-State Constraint Kalman Filter for Vision-aided Inertial Navigation

Anastasios I. Mourikis and Stergios I. Roumeliotis

Abstract—**In this paper, we present an Extended Kalman Filter (EKF)-based algorithm for real-time vision-aided inertial navigation. The primary contribution of this work is the derivation of a measurement model that is able to express the geometric constraints that arise when a static feature is observed from multiple camera poses. This measurement model does not require including the 3D feature position in the state vector of the EKF and is optimal, up to linearization errors. The vision-aided inertial navigation algorithm we propose has computational complexity only *linear* in the number of features, and is capable of high-precision pose estimation in large-scale**

geometric constraints involving all these poses. The primary contribution of our work is a measurement model that expresses these constraints *without* including the 3D feature position in the filter state vector, resulting in computational complexity only *linear* in the number of features. After a brief discussion of related work in the next section, the details of the proposed estimator are presented in Section III. In Section IV we describe the results of a large-scale experiment in an uncontrolled urban environment, which demonstrate

$$\widetilde{\mathbf{X}}_{\mathrm{IMU}} = \begin{bmatrix} \delta\boldsymbol{\theta}_I^T & \widetilde{\mathbf{b}}_g^T & {}^G\widetilde{\mathbf{v}}_I^T & \widetilde{\mathbf{b}}_a^T & {}^G\widetilde{\mathbf{p}}_I^T \end{bmatrix}^T$$

Assuming that N camera poses are included in the EKF state vector at time-step k, this vector has the following form

$$\hat{\mathbf{X}}_k = \begin{bmatrix} \hat{\mathbf{X}}_{\mathrm{IMU}_k}^T & \boxed{{}^{C_1}_{G}\hat{\bar{q}}^T \quad {}^G\hat{\mathbf{p}}_{C_1}^T} & \dots & {}^{C_N}_{G}\hat{\bar{q}}^T & {}^G\hat{\mathbf{p}}_{C_N}^T \end{bmatrix}^T$$

R & T of camera pose

${}^{C_i}_{G}\hat{\bar{q}}$ and ${}^G\hat{\mathbf{p}}_{C_i}$, $i = 1 \dots N$ are the estimates of the camera attitude and position, respectively.

mourikis2007multi

Review of MSCKF

Pipeline of MSCKF

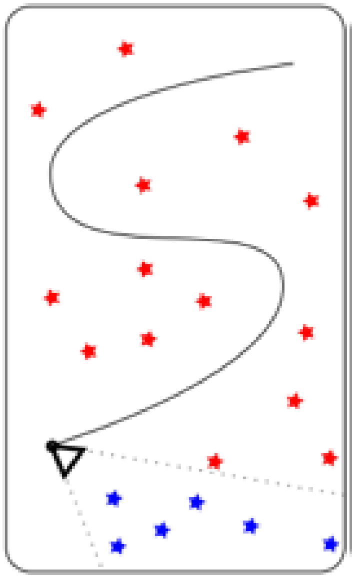
Algorithm 1 Multi-State Constraint Filter

Propagation: For each IMU measurement received, propagate the filter state and covariance (cf. Section III-B).

Image registration: Every time a new image is recorded,

- augment the state and covariance matrix with a copy of the current camera pose estimate (cf. Section III-C).
- image processing module begins operation.

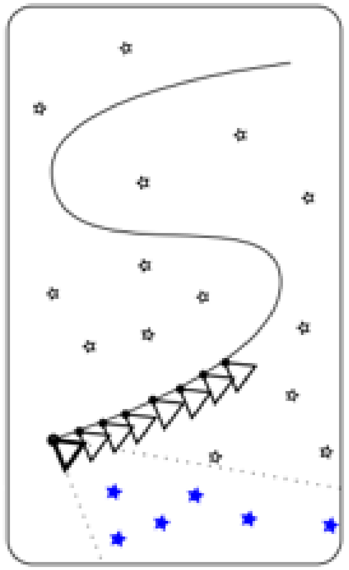
Update: When the feature measurements of a given image become available, perform an EKF update (cf. Sections III-D and III-E).



EKF-SLAM
(traditional)

$O(n^3)$

VS.



MSCKF

$O(n^2 l)$

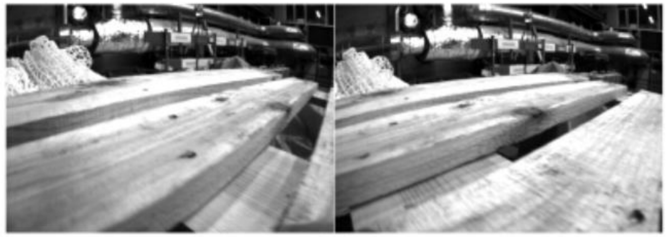
n: # landmarks
l: # iamge frames
MSCKF is more accurate and faster, due to ***l<<n.***

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

Content

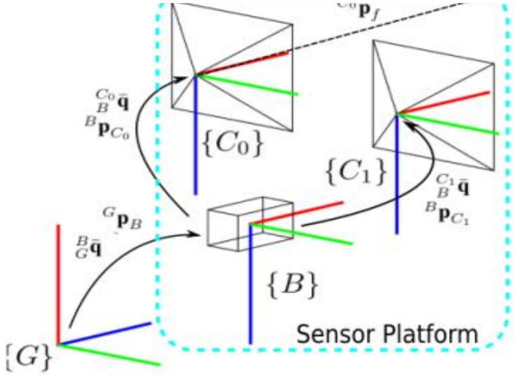
Previous work

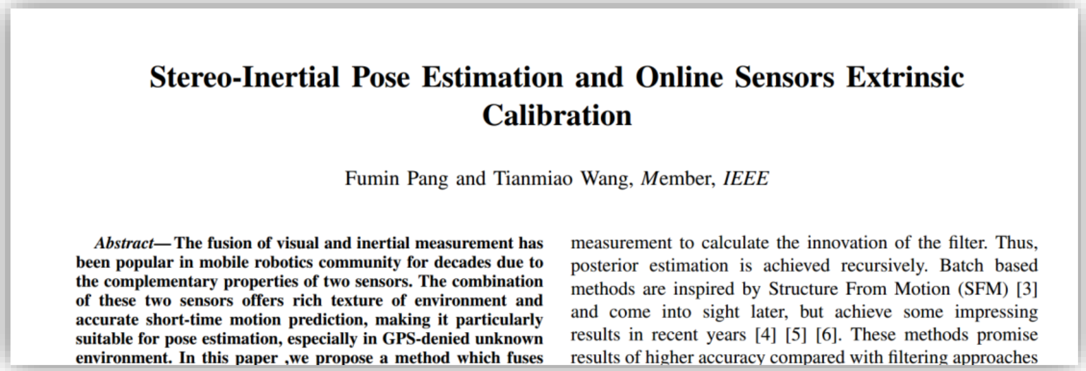
Stereo-Inertial MSCKF



(a)

(b)





Stereo-Inertial Pose Estimation and Online Sensors Extrinsic Calibration

Fumin Pang and Tianmiao Wang, *Member, IEEE*

***Abstract*—The fusion of visual and inertial measurement has been popular in mobile robotics community for decades due to the complementary properties of two sensors. The combination of these two sensors offers rich texture of environment and accurate short-time motion prediction, making it particularly suitable for pose estimation, especially in GPS-denied unknown environment. In this paper ,we propose a method which fuses**

measurement to calculate the innovation of the filter. Thus, posterior estimation is achieved recursively. Batch based methods are inspired by Structure From Motion (SFM) [3] and come into sight later, but achieve some impressing results in recent years [4] [5] [6]. These methods promise results of higher accuracy compared with filtering approaches

The visual measurements are from two cameras, the measurement model is augmented with two camera projection process. Depth information of features are used to estimate the states.

$$H_{x_{B_l}}^{(j)} = \begin{bmatrix} 0_{4\times15} & \Pi_l & 0_{4\times9(l-1)} & H_{B_l} & 0_{4\times9(N-l)} \end{bmatrix} \tag{28}$$

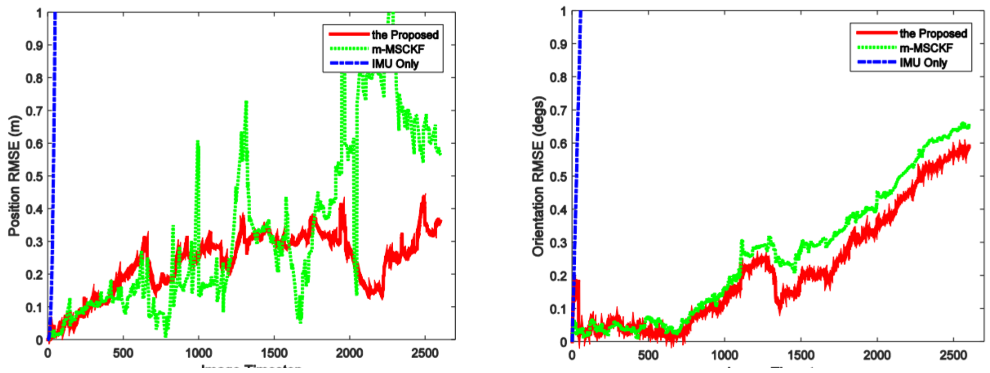
where

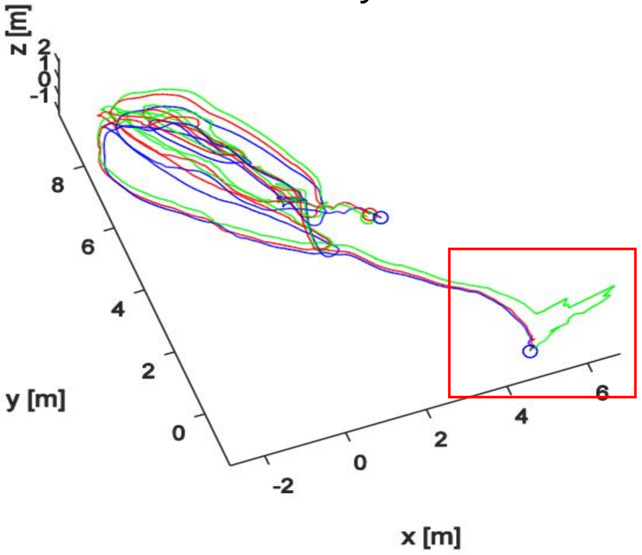
$$\Pi_l = \begin{bmatrix} \Pi_{\theta_{l,0}} & \Pi_{p_{l,0}} & 0_{2\times3} & 0_{2\times3} \\ 0_{2\times3} & 0_{2\times3} & \Pi_{\theta_{l,1}} & \Pi_{p_{l,1}} \end{bmatrix} \tag{29}$$

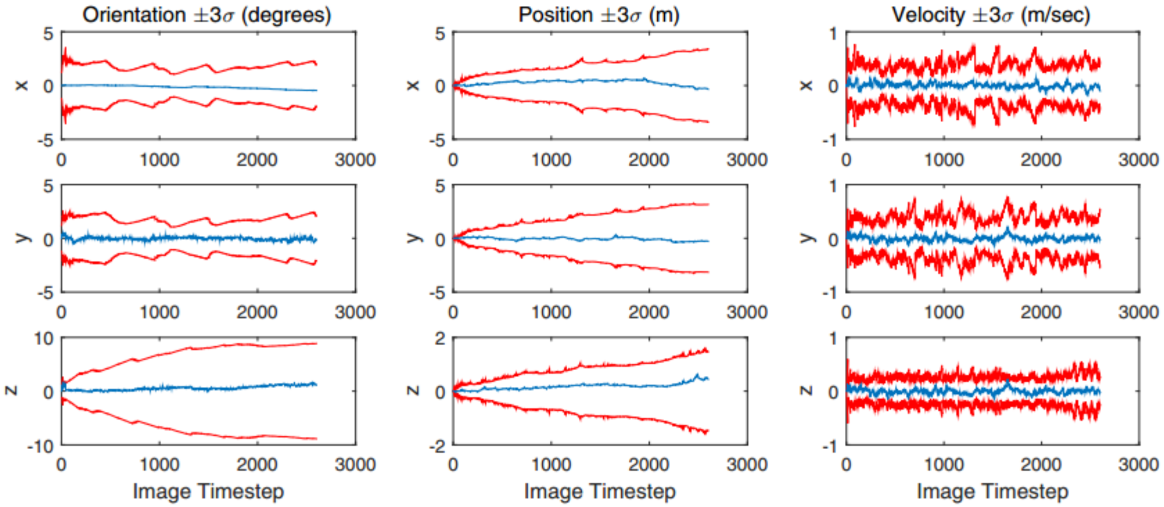
$$H_{B_l} = \begin{bmatrix} H_{\theta_{l,0}} & H_{p_{l,0}} & 0_{2\times3} \\ H_{\theta_{l,1}} & H_{p_{l,1}} & 0_{2\times3} \end{bmatrix} \tag{30}$$

Previous work

Introduce depth information can give:
1. More accurate pose estimate.
2. Same correct consistence.
3. More Rapid convergence and initialization for VIO system.







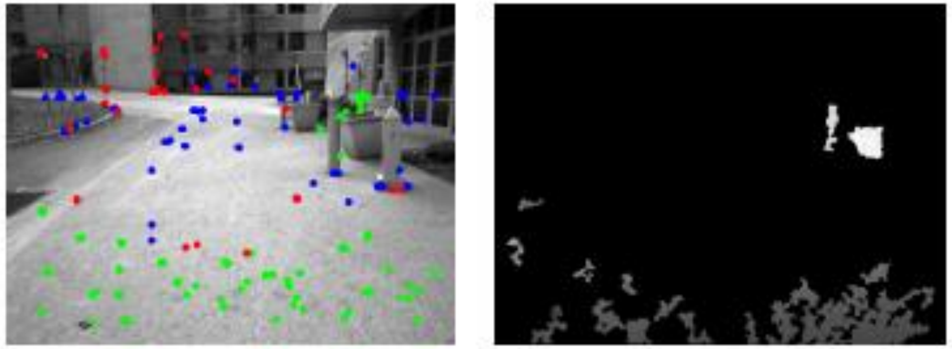
Fumin Pang,Tianmiao Wang. Stereo-Inertial Pose Estimation and Online Sensors Extrinsic Calibration[C]//Robotics and Biomimetics (ROBIO), 2016 IEEE International Conference on. IEEE, 2016:1377-1382.

Previous work

But depth information is always sparse!

1.Scenarios with large depth variation leave large area in the images where depth is unavailable **spatially**.

2. When robot moves, visual measurements of a salient points always do not have depth corresponding information throughout all poses **temporally**.



Sparse depth example from zhang2014real

Overlay image of From fisheye and depth information form Segway Robot Loomo

Depth Enhanced MSCKF

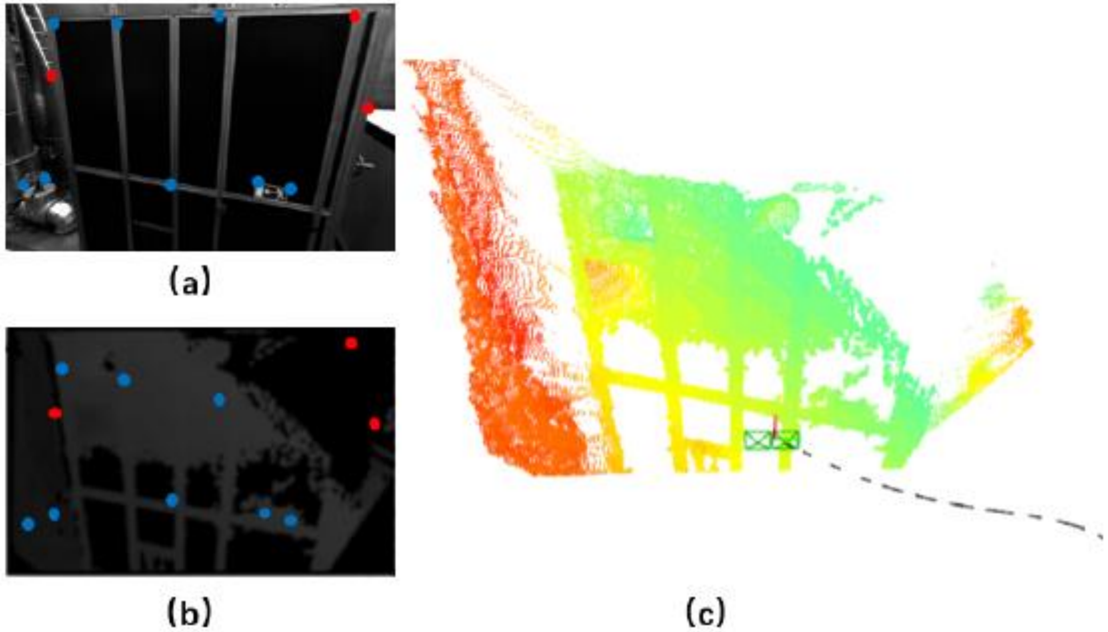
Inspired by zhang2014real, we come up with Depth enhanced MSCKF.

Contribution

We combine:

1. Features **without** available depth information to form 2D projection measurements.
2. Features **with** depth to form 3D measurements.

In the update step of this depth enhanced MSCKF, 2D reprojection measurement errors for landmarks and 3D position measurement errors are jointed to correct the pose estimate.



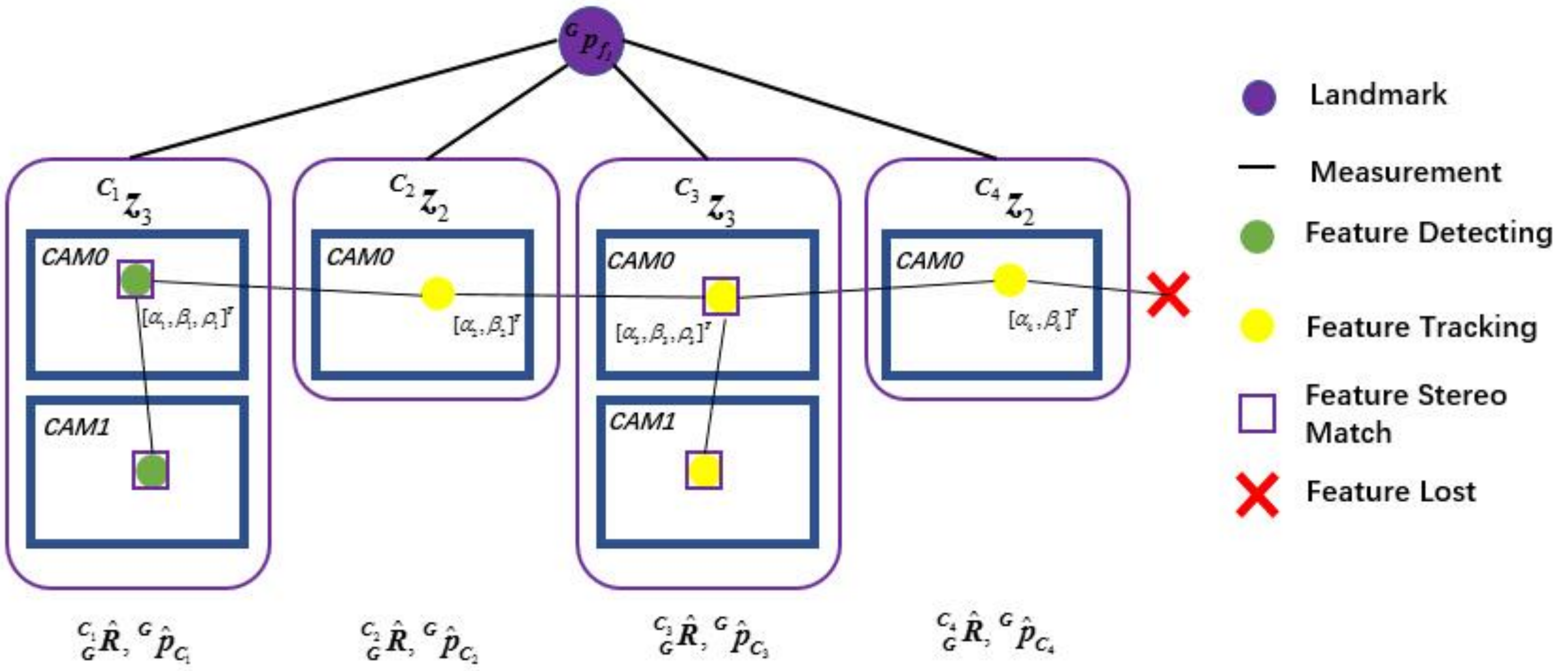
(a) (b) (c)

The blue dots encode the features with depth
The red dots encode the features without depth.

Zhang J, Kaess M, Singh S. Real-time depth enhanced monocular odometry[C]//Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on. IEEE, 2014: 4973-4980.

Depth Enhanced MSCKF

A toy example of a feature track in MSCKF pipeline.

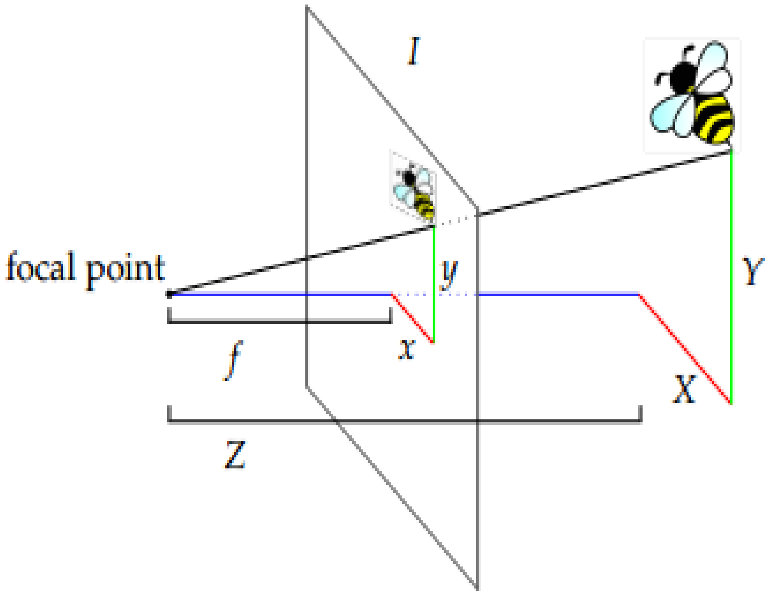


The feature is with two 2D measurements and 2 3D measurements at 4 different pose before lost. We combine these two type measurements together to update the filter.

Depth Enhanced MSCKF

2D Measurement Model:

Mean and Covariance of 2D measurement



Mean:

$$ {}^{C_i}\hat{z}^{(j)}_{2,i} = \boldsymbol{\pi}_2\left({}^{C_i}_{G}\hat{\boldsymbol{R}}\left({}^{G}\hat{\mathbf{p}}_{f_j} - {}^{G}\hat{\mathbf{p}}_{C_i}\right)\right) $$

$$ \boldsymbol{\pi}_2(\boldsymbol{h}) = \begin{bmatrix} \frac{h(1)}{h(3)} \\ \frac{h(2)}{h(3)} \end{bmatrix} $$

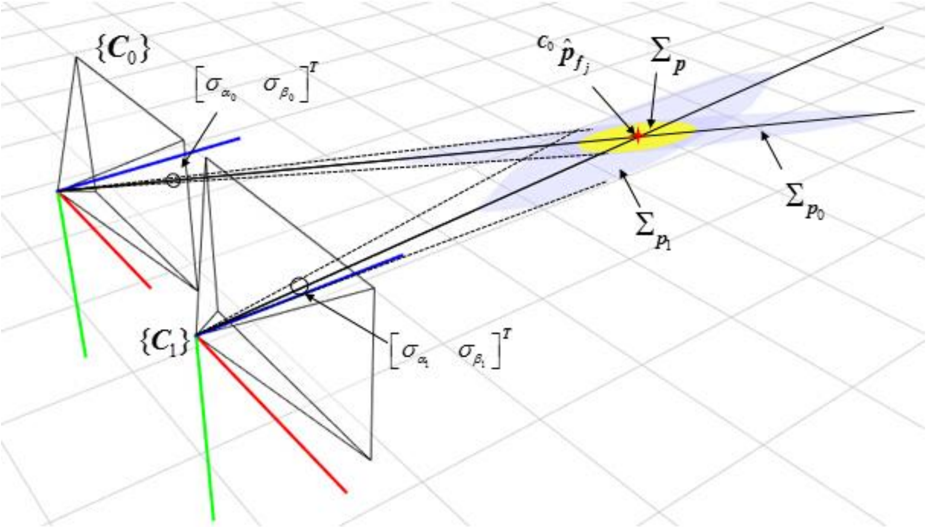
Measurement variance：

$$ {}^{C_i}\boldsymbol{\Sigma}_2 = R_2 = \begin{bmatrix} \sigma_\sigma^2 & 0 \\ 0 & \sigma_\beta^2 \end{bmatrix} $$

Depth Enhanced MSCKF

3D Measurement Model:

Mean and Covariance of 3D measurement



Two individual 3d measurement means：

$$ {}^{C_i^0}\boldsymbol{z}_3 = \begin{bmatrix} {}^{C_i^0}\alpha \\ {}^{C_i^0}\beta \\ {}^{C_i^0}\rho \end{bmatrix} = \boldsymbol{\pi}_3 \begin{pmatrix} {}^{C_i^0}X \\ {}^{C_i^0}Y \\ {}^{C_i^0}Z \end{pmatrix} = \begin{bmatrix} \frac{{}^{C_i^0}X}{{}^{C_i^0}Z} \\ \frac{{}^{C_i^0}Y}{{}^{C_i^0}Z} \\ \frac{1}{{}^{C_i^0}Z} \end{bmatrix} $$

$$ {}^{C_i^1}\boldsymbol{z}_3 = \begin{bmatrix} {}^{C_i^1}\alpha \\ {}^{C_i^1}\beta \\ {}^{C_i^1}\rho \end{bmatrix} = \boldsymbol{\pi}_3 \left(\boxed{{}^{C^1}_{C^0}\boldsymbol{R}} \begin{pmatrix} {}^{C_i^0}X \\ {}^{C_i^0}Y \\ {}^{C_i^0}Z \end{pmatrix} + \boxed{{}^{C^1}\boldsymbol{p}_{C^0}} \right) $$

Sensor extrinsic pose

Two individual 3d measurement variance：

$$ \boldsymbol{R}^j{}_3 = \begin{bmatrix} \sigma_\sigma^2 & 0 & 0 \\ 0 & \sigma_\beta^2 & 0 \\ 0 & 0 & \sigma_\rho^2 \end{bmatrix} $$

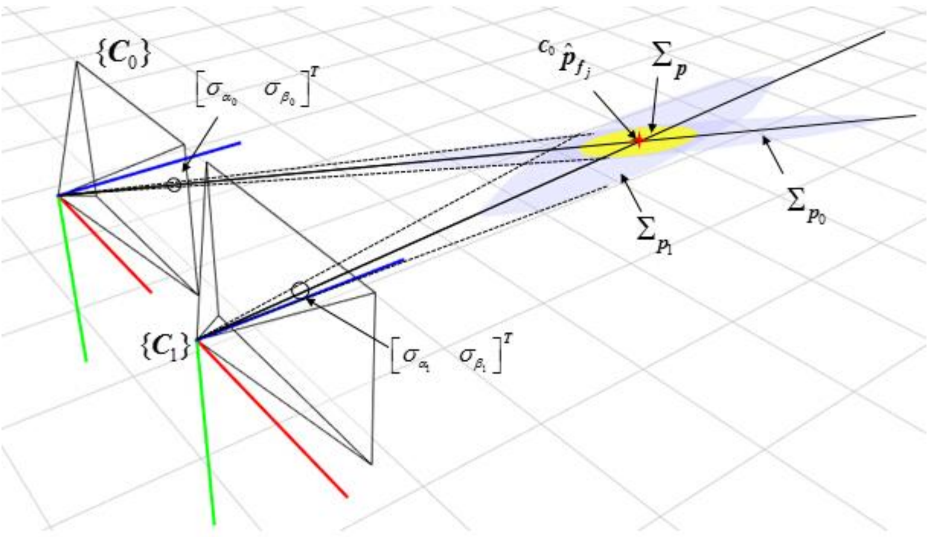
$$ \rho \sim \mathcal{N}\left(\rho, \sigma_\rho^2\right) \qquad j = 0, 1 $$

We set the inverse depth std. to a large value ,because we do not have any information about depth in individual camera before fusion.

SEGWAY 北京航空航天大學 BEIHANG UNIVERSITY

Depth Enhanced MSCKF

Mean and Covariance of 3D measurement



Fuse two individual 3d measurement:

Mean:

The feature point can be calculated by two view triangulation.

$$ {}^{C_i}\boldsymbol{z}_3 = \begin{bmatrix} {}^{C_i}\alpha \\ {}^{C_i}\beta \\ {}^{C_i}\rho \end{bmatrix} $$

In Cam0

3D variance:

1. Calculate 3d feature position variance in individual in C^0 frame. (two larger ellipsoid)

$$ \boldsymbol{\Sigma}_{p^j} = \boldsymbol{J}_{p^j}\boldsymbol{diag}\left(\sigma^2_{\alpha^j}, \sigma^2_{\beta^j}, \sigma^2_{\rho^j}\right)\boldsymbol{J}_{p^j}{}^T $$

$$ \boldsymbol{J}_{p^j} = \frac{\partial^{C^0}\boldsymbol{p}_f}{\partial^{C_i^j}\boldsymbol{z}_3} $$

2. Fusing two 3-dimensional Gaussian distributions reduces the uncertainty of the distribution .(the smaller ellipsoid)

$$ {}^{C_i}\boldsymbol{\Sigma}_p = \boldsymbol{\Sigma}_{p^0}\left(\boldsymbol{\Sigma}_{p^0} + \boldsymbol{\Sigma}_{p^1}\right)^{-1}\boldsymbol{\Sigma}_{p^1} $$

2. Derivate the final 3D covariance.

$$ {}^{C_i}\boldsymbol{\Sigma}_3 = \boldsymbol{J}_{\boldsymbol{\pi}_3}{}^{C_i}\boldsymbol{\Sigma}_p\boldsymbol{J}_{\boldsymbol{\pi}_3}{}^T $$

$$ \boldsymbol{J}_{\boldsymbol{\pi}_3} = \frac{\partial^{C_{0,i}}\boldsymbol{z}_3}{\partial^{C_0}\boldsymbol{p}_f} $$

Depth Enhanced MSCKF

Depth Enhanced Feature Position Estimation

Combine 2D and 3D measurement to estimate feature position using least-squares method.

2D measurement error: $$\boldsymbol{e}_{2,i}({}^{C_1}\hat{\boldsymbol{p}}_{f_j}) = {}^{C_i}\boldsymbol{z}_{2,i} - \boldsymbol{\pi}_2\left({}^{C_i}_{C_1}\hat{\boldsymbol{R}}^{C_1}\hat{\boldsymbol{p}}_{f_j} + {}^{C_i}\hat{\boldsymbol{p}}_{C_1}\right)$$

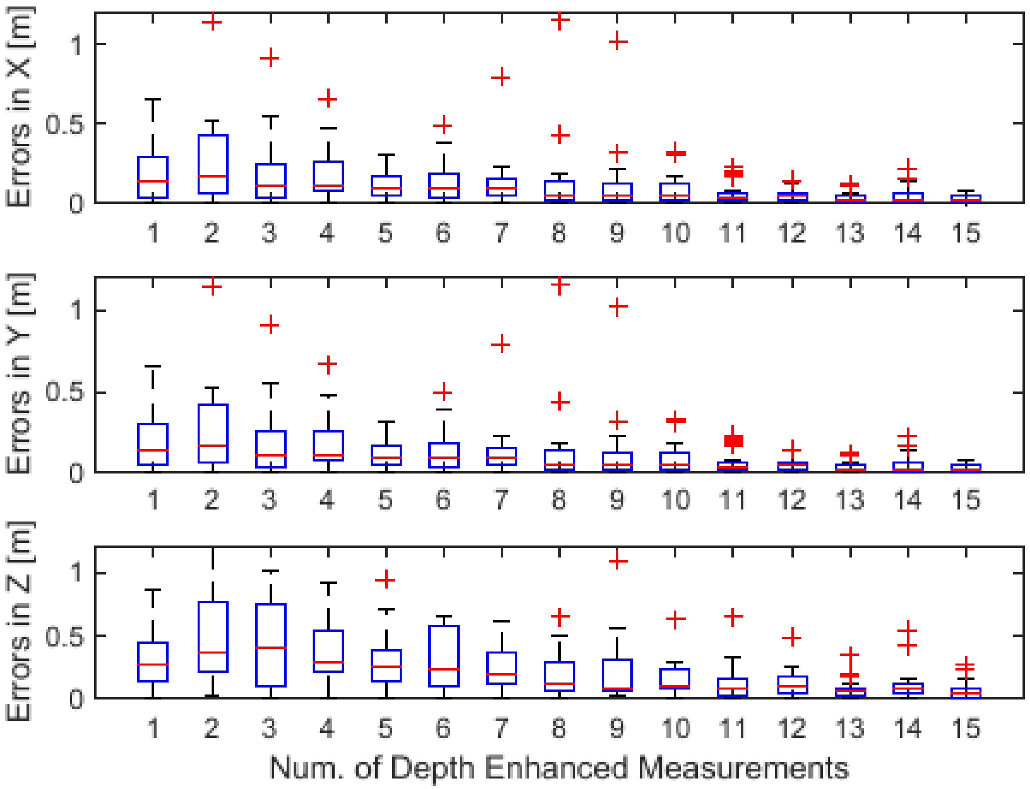
3D measurement error: $$\boldsymbol{e}_{3,i}({}^{C_1}\hat{\boldsymbol{p}}_{f_j}) = {}^{C_i}\boldsymbol{z}_{3,i} - \boldsymbol{\pi}_3\left({}^{C_i}_{C_1}\hat{\boldsymbol{R}}^{C_1}\hat{\boldsymbol{p}}_{f_j} + {}^{C_i}\hat{\boldsymbol{p}}_{C_1}\right)$$

Linearized equation: $$\left(\boldsymbol{J}^T\boldsymbol{W}^{-1}\boldsymbol{J}\right)\boldsymbol{\delta}^{C_1}\boldsymbol{p}^{\star}_{f_j} = -\boldsymbol{J}^T\boldsymbol{W}^{-1}\boldsymbol{e}\left({}^{C_1}\hat{\boldsymbol{p}}_{f_j}\right)$$

2D and 3D measurement weight: $$\boldsymbol{W} = \boldsymbol{diag}\left\{{}^{C_1}\boldsymbol{\Sigma}_3 \quad {}^{C_2}\boldsymbol{\Sigma}_2 \quad {}^{C_3}\boldsymbol{\Sigma}_3 \quad {}^{C_4}\boldsymbol{\Sigma}_2\right\}$$

Depth Enhanced MSCKF

Simulation on
Depth Enhanced Feature Position Estimation



Combine more 3D measurements will make landmark position estimation more accurate, **especially in Z-axis**.

Depth Enhanced MSCKF

Depth Enhanced Filter Update:

1. Stack all 2D and 3D residuals together

$$\boldsymbol{r}_i^{(j)} = {}^{C_i}\boldsymbol{z}_i^{(j)} - {}^{C_i}\hat{\boldsymbol{z}}_i^{(j)}$$

$$\boldsymbol{r}_i^{(j)} = \boldsymbol{H}_{\boldsymbol{x}_{\boldsymbol{B}_i}}^{(j)}\tilde{\boldsymbol{x}} + \boldsymbol{H}_{\boldsymbol{f}_i}^{(j)}\,{}^{G}\tilde{\boldsymbol{p}}_{f_j} + \boldsymbol{n}_i^{(j)}$$

2. Feature Error Marginalization

Multiply both sides with the left nullspace of ***Hf*** **(*j*).**

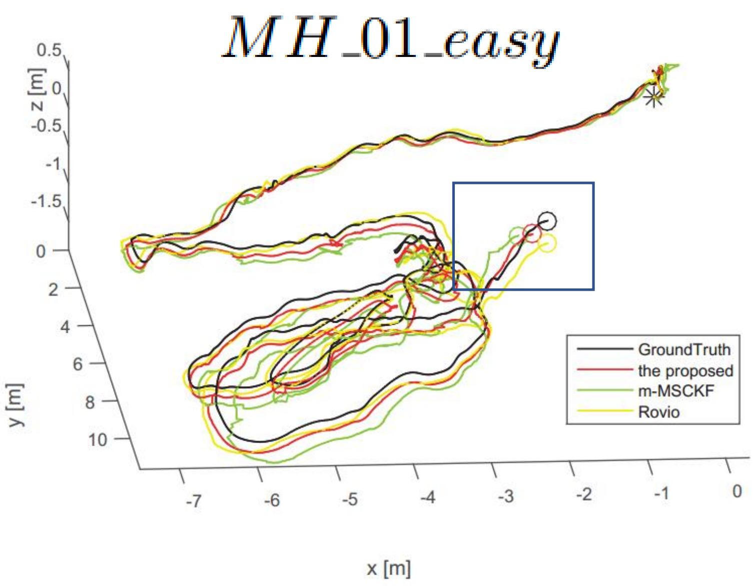
$$\boldsymbol{r}_o = \boldsymbol{H}_o\tilde{\boldsymbol{x}} + \boldsymbol{n}_o$$

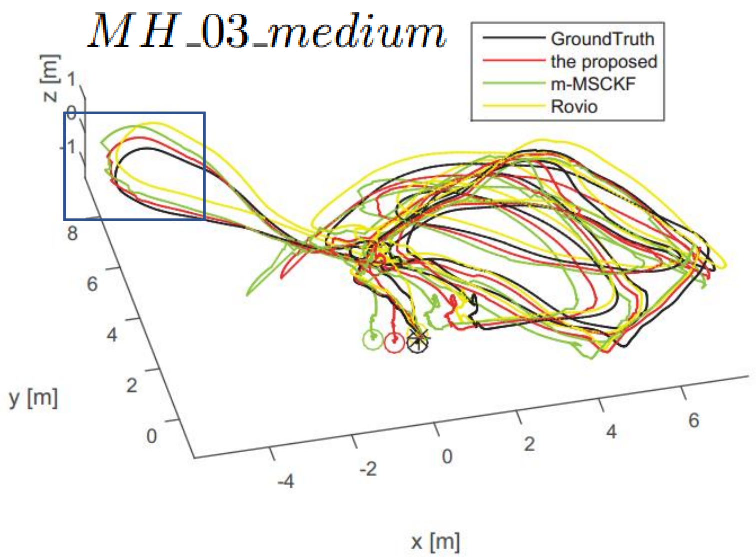
3. To reduce the computational complexity of the MSCKF update, a ***QR****-decomposition of* ***Ho*** *is employed.*

Content

Experiment

A. *Pose Estimation* **Evaluated on EuRoC MAV dataset**





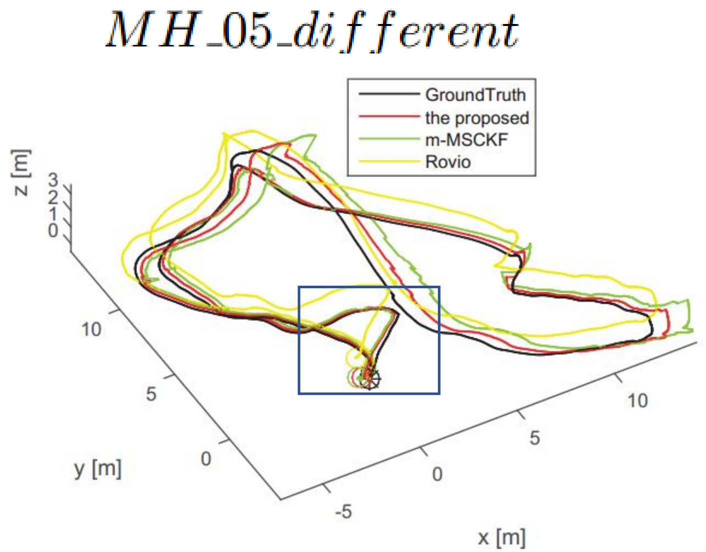
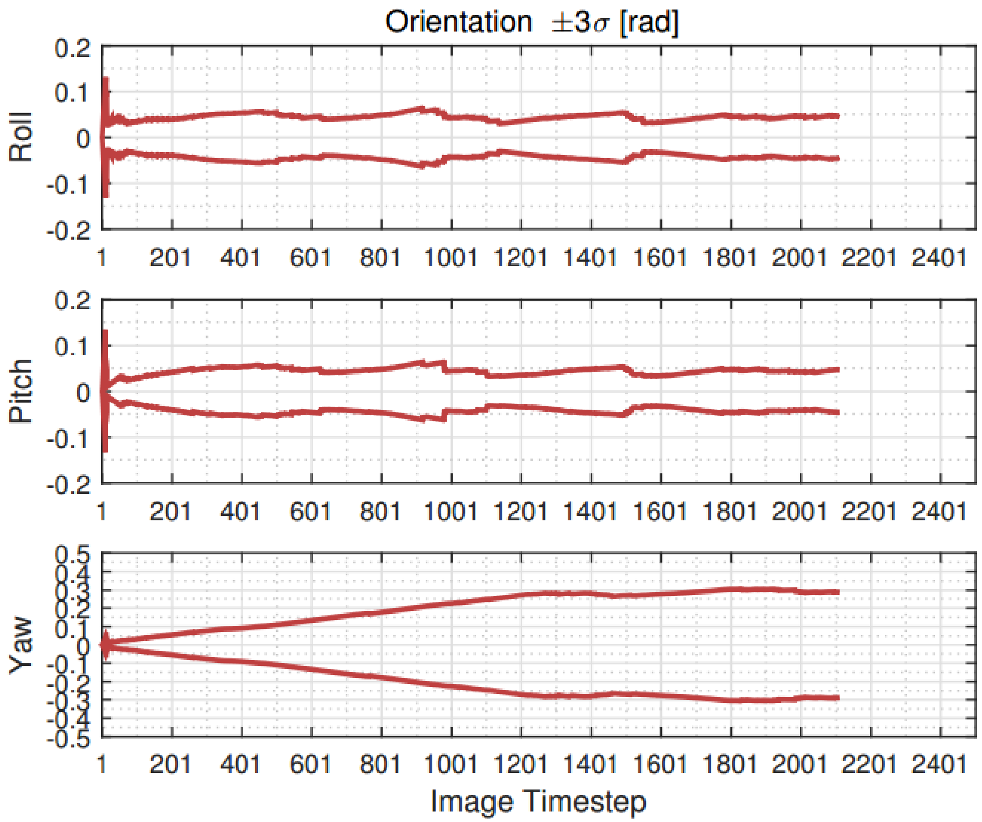


TABLE I: ATE on EuRoC / ASL Dataset

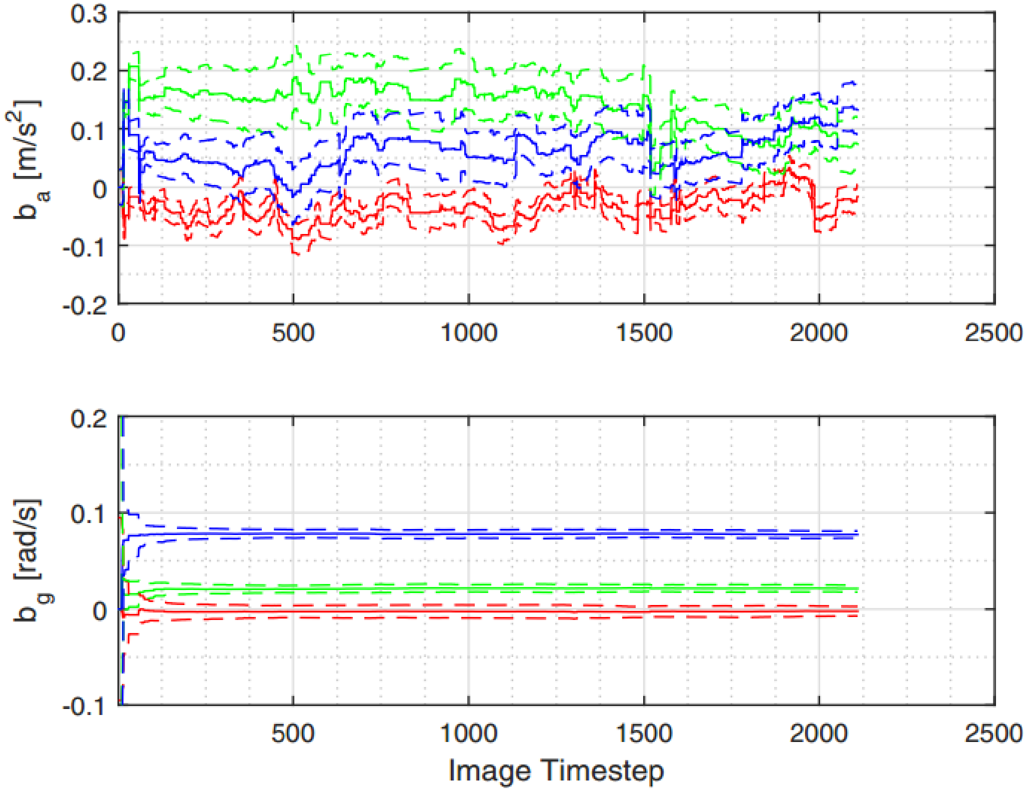
	mono-MSCKF			Rovio			The Proposed		
Dataset	Mean	Median	RMSE[m]	Mean	Median	RMSE[m]	Mean	Median	RMSE[m]
MH_01_easy	0.3536	0.3433	0.3875	0.2517	0.2560	0.2761	**0.1716**	**0.1768**	**0.1936**
MH_03_medium	0.7778	0.8026	0.8511	0.4053	**0.3839**	0.4524	**0.4002**	0.4134	**0.4439**
$MH\_05\_difficult$	0.8943	0.6438	1.1091	1.0490	1.0891	1.1250	**0.4292**	**0.2984**	**0.5262**
$V2\_01\_easy$	0.2500	0.1963	0.2998	0.2383	0.2437	0.2564	**0.0959**	**0.0633**	**0.1370**
$V2\_02\_medium$	0.4825	0.4576	0.5329	0.3660	**0.2948**	**0.4296**	**0.1643**	0.4525	0.4576

Experiment

B. Orientation Observability



C. Bias Estimation



Content

Conclusion

1. In this paper, we have presented a depth enhanced visual inertial odometry system based on MSCKF.

2. A 3D measurement model of landmark position is derivated, which reduces the estimation uncertainty.

3. Compared to mono-MSCKF and other popular open-source method, the proposed gets competitive accuracy.

4. In this work, we use stereo camera to obtain depth information. But it can be extended and adapted to various sensors which can provide bearing and range information.

Thank you!

For more detail:

Please contact Pang Fumin : *fuminpang@buaa.edu.cn*
pangfumin@gmail.com