Cost-effective and Robust Visual Based Localization with Consumer-level Cameras at Construction Sites

Runqiu Bao
Dept. of Precision Engineering
The University of Tokyo
Tokyo, Japan
bao@robot.t.u-tokyo.ac.jp

Ren Komatsu
Dept. of Precision Engineering
The University of Tokyo
Tokyo, Japan
komatsu@robot.t.u-tokyo.ac.jp

Renato Miyagusuku
Dept. of Mechanical and Intelligent Engineering
Utsunomiya University
Tochigi, Japan
miyagusuku@cc.utsunomiya-u.ac.jp

Masaki Chino
Construction Division
HAZAMA ANDO CORPORATION
Tokyo, Japan
chino.masaki@ad-hzm.co.jp

Atsushi Yamashita
Dept. of Precision Engineering
The University of Tokyo
Tokyo, Japan
yamashita@robot.t.u-tokyo.ac.jp

Hajime Asama
Dept. of Precision Engineering
The University of Tokyo
Tokyo, Japan
asama@robot.t.u-tokyo.ac.jp

***Abstract*—Localization is essential for construction machines navigation. Current construction machines use RTK-GNSS for this purpose. We consider that using on-board cameras with SLAM algorithms to estimate machines' ego-motion and position change would not only be cheaper, but also more robust and widely applicable. To investigate the feasibility of this idea, we installed several consumer cameras on real construction machines and performed experiments at real construction sites. Our experiments show promising results, achieving localization accuracy of about 0.06 meters.**

***Index Terms*—consumer camera, SLAM, construction site**

Fig. 1. A typical construction site scene in our setting

I. Introduction

Navigation of autonomous construction machines requires real-time and accurate localization. Currently, RTK-GNSS (Real Time Kinematic - Global Navigation Satellite System) is widely used because of its centimeter-level accuracy. But it has shortcomings of being costly and unstable due to losing satellite signals in underground, between tall buildings, and near mountains and trees. In this paper, we investigate the feasibility of replacing expensive RTK-GNSS by a more cost-effective visual localization system composed of several on-board cameras and a Simultaneous Localization and Mapping (SLAM) algorithm.

We propose the use of visual SLAM as it has recently experienced great developments in terms of accuracy and computational complexity. Visual SLAM can not achieve absolute localization, but most construction machines only have simple moving patterns within a limited area, making absolute localization unnecessary. Moreover, we use consumer cameras as the visual sensor, since recent improvements in consumer cameras: high frame rate, high resolution, large field of view (FoV), low price; make them a sensible choice.

The remaining of this work is organized as follows: In Sec. II we detail our proposed visual localization system, which only uses consumer cameras. In Sec. III, we test whether this system can replace RTK-GNSS, by performing experiments at real construction sites using construction machines performing their routine tasks. Finally, in Sec. IV we present our conclusions and future work.

II. Method

Our goal is to build a visual localization system for construction machines. Two difficulties arise from our setting: First, feature extraction and tracking may be difficult at construction sites, as they are typically flat and poorly-textured (Fig. 1). Second, the severe vibration of some construction machines leads to motion blur and rolling shutter deformations in the captured image data, which could affect the accuracy of motion estimation.

To verify if visual SLAM is feasible, we compare a commonly used feature-based formulation, ORB-SLAM monocular [1], against DSO (Direct Sparse Odometry) [2], a widely used formulation which does not use features (direct method). We wonder which approach will have higher motion estimation accuracy in our setting. For fair comparison, we do not use loop-closure and relocalization modules originally included in ORB-SLAM when comparing to DSO. Since these are monocular vSLAM formulations, the scale was adjusted using prior information. And for the purpose of automatically getting real-world scale, ORB-SLAM stereo [3] is applied in further evaluation.

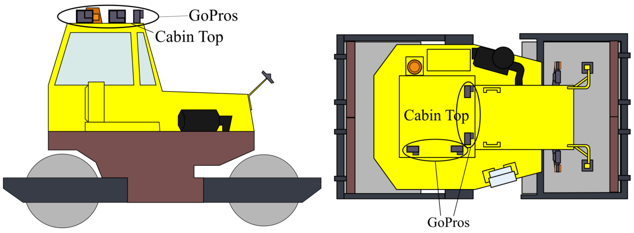

Fig. 2. Vibration roller with GoPro array, side view and top view

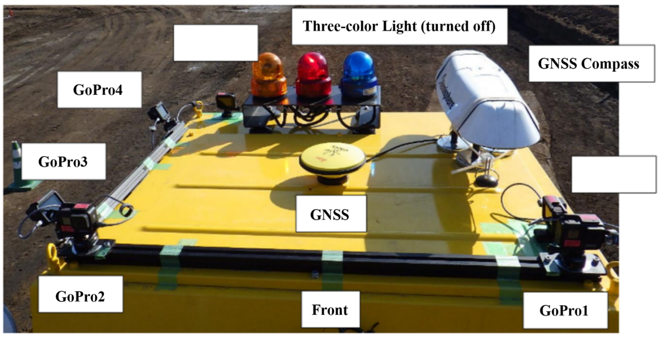

Fig. 3. Devices on the cabin top of the vibration roller

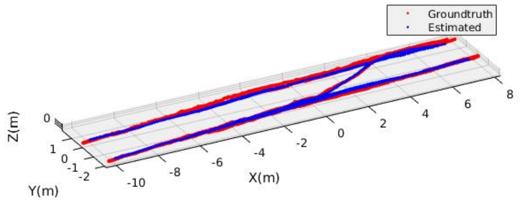

Fig. 4. Estimated trajectory with ground truth in an experiment (Loop closure and re-localization enabled)

TABLE I
TRAJECTORY ACCURACY RMSE (ROOT MEAN SQUARE ERROR)

Experiment conditions	**Result trajectory RMSE (m)**		
	Side mounted camera		**Front mounted camera**
Normal vibration	**Feature based method (stereo)**		
	0.114		0.057
Severe vibration	**Direct method**†	**Feature based method***†	
	1.027	0.059	

*No loop closure nor relocalization triggered
†Distance scale was adjusted using prior information, because monocular vSLAM formulation can not provide real-world scale.

For our visual localization system, we choose the following camera configurations:

- Cameras are set to their largest FoV and are mounted on the top of the machine cabin to expand the horizon of vision.
- Since construction machines typically move along straight lanes, we mount the cameras on the side perpendicular to the moving direction and on the front side facing the moving direction.
- Two cameras are combined into a stereo pair to enable distance measurement [3].

A. *Camera calibration and synchronization*

For the wide FoV cameras we use, standard camera models are no longer valid. Instead, a fish-eye camera model is used for calibration as proposed in [4].

For the stereo pair, another issue is synchronization between the two cameras used. Our approach is to show a millisecond accuracy timer to the stereo pair to roughly synchronize the videos. For consumer cameras, usually there will be no dedicated synchronization interface. The USB data transmission ports may be used for synchronization, if an official API is provided.

B. *Offline result evaluation*

To evaluate localization accuracy, we first use ICP (Iterative Closest Point) [5] to align the estimated trajectory with ground truth obtained by RTK-GNSS. Then we calculate RMSE (Root Mean Square Error) between the estimated trajectory and the corresponding subset in ground truth.

RTK-GNSS that we are using can achieve average localization accuracy of 3cm – 5cm. Thus if the result RMSE is also centimeter-level, replacing RTK-GNSS with our visual localization system should be a viable idea.

III. EXPERIMENT

A. *Experiment equipment*

We conducted experiments at real construction sites using a vibration roller, shown in Fig. 2. Vibration rollers move along straight lanes while vibrating, to flatten the ground. In experiments, a vibration roller would move forward for about 20 meters, then move back to the start point and change lanes. After finishing two lanes, one round finished and we saved all sensors data. There will be more than five rounds in each experiment to collect enough data.

For the cameras, we choose GoPro HERO7 black version. Cameras were set to 3840×2160 resolution (downsized to 818×478 when processing the frame, for increased processing speed), 60 fps frame rate, 120 degree horizontal FoV.

This camera was chosen due to its wide range of available camera settings, micro-SD card port which supports up to 256 GB, numerous available mounting accessories and GoPro Smart Remote which enables simultaneous control of multiple GoPros.

B. *Experiment results*

Although poorly-textured environment and severe vibration of the machine, feature extraction and tracking is working well. This is, on one hand, because of the high resolution and frame rate setting in GoPro. On the other hand, surprisingly even in such environment as shown in Fig. 1, there are still enough features can be found. Thus, feature-based method for camera ego-motion estimation is considered applicable in our setting.

As shown in Table 1, off-line evaluation has indicated that the estimated machine trajectory by feature-based method can achieve centimeter-level accuracy, which is compatible with RTK-GNSS localization. Figure 3 shows the best result of the estimated trajectory in an experiment when the camera was in normal vibration and the data captured by the front-mounted cameras were used. The resulting RMSE is about 0.057 m.

Direct method shows worse accuracy, especially when the machine's vibration was severe. In one experiment with severe machine vibration, ORB feature-based method (with neither loop closure nor re-localization) remained robust and RMSE was about 0.059 m (scale adjusted). However, direct method accuracy dropped to 1.027 m (scale adjusted). According to Raul et al. [3], performance of direct method can be severely degraded by unmodeled effects like rolling shutter deformations caused by machine's vibration.

About whether real-time computation is feasible, current processing speed is 23 fps using a computer of Intel Core i7-6800K CPU. Compared with 10 Hz output rate of the RTK-GNSS localization device, it can be regarded as real-time.

Besides, for on-line localization, we need to directly process video stream captured by GoPros. Fortunately, developers have made an available programming interface[1] for GoPro video stream transmission, which may allow us to do this. Utilizing the micro HDMI port on GoPro (HERO7) is also worth a try.

IV. Conclusion and Future Work

This research aims to investigate the feasibility of replacing RTK-GNSS by on-board consumer cameras for construction machines localization. According to our experiments and result evaluation, we found this is a viable idea. In future work, we will implement an on-line localization system. Moreover, we will also look into dynamic objects problem as there are lots of persons, vehicles and machines moving in a construction site which can cause problems in feature-based camera ego-motion estimation.

References

[1] Mur-Artal, Raul, Jose Maria Martinez Montiel, and Juan D. Tardos, "ORB-SLAM: a versatile and accurate monocular SLAM system," IEEE Transactions on Robotics 31.5 (2015): 1147-1163.

[2] Engel, Jakob, Vladlen Koltun, and Daniel Cremers, "Direct sparse odometry," IEEE Transactions on Pattern Analysis and Machine Intelligence 40.3 (2018): 611-625.

[3] Mur-Artal, Raul, and Juan D. Tards, "Orb-slam2: An open-source slam system for monocular, stereo, and RGB-D cameras," IEEE Transactions on Robotics 33.5 (2017): 1255-1262.

[4] D. Scaramuzza, A. Martinelli, and R. Siegwart "A Toolbox for Easy Calibrating Omindirectional Cameras," 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE, 2006.

[5] Chen, Y. and G. Medioni "Object Modelling by Registration of Multiple Range Images," Image and Vision Computing 10.3 (1992): 145-155.

[1]Konrad Iturbe, Unofficial GoPro API Library for Python, https://github.com/KonradIT/gopro-py-api