# Real-Time UAV 3D Image Point Clouds Mapping

Shangzhe Sun[1,2,3], Chi Chen[1,2,3,*], Zhiye Wang[1,2,3], Jian Zhou[1], Liuchun Li[4], Bisheng Yang[1,2,3], Yangzi Cong[1,2,3], Haoyu Wang[1,2,3]

[1] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China - (SSZ, chichen, zhiye.wang, jianzhou, bshyang, yzcong, spacewang)@whu.edu.cn
[2] Engineering Research Centre for Spatio-Temporal Data Acquisition and Smart Application(STSA), Ministry of Education in China, Wuhan, China
[3] Institute of Artificial Intelligence in Geomatics, Wuhan University, Wuhan, China
[4] Institute of Artificial Intelligence, School of Computer Science, Wuhan University, Wuhan, China - (LiuC.Lee)@whu.edu.cn

**KEY WORDS:** Real-time, SLAM, VIO, UAV, Mapping

**ABSTRACT:**

This paper proposes a real-time 3D image point clouds mapping algorithm for UAVs that is capable of mapping effectively in weak GNSS environments. And a UAV mapping system is integrated with a RGB camera, an inertial measurement unit (IMU), a GNSS receiver, data transmission devices, and a DJI M300 flight platform. To achieve real-time and robust mapping, the system utilizes a visual-inertial odometry (VIO) that tightly couples GNSS, RGB image, and IMU, which provides stable state estimation information for mapping. Subsequently, a dense matching algorithm based on key frames is adopted to recover 3D mapping information with low-computational cost. Extensive experiments are conducted on our test site, demonstrating the system's ability to build maps stably, even under the effect of wind. The results compared with the trajectory reconstructed by Pix4D show that the system achieves competitive accuracy of pose estimation and is capable of real-time mapping.

## 1. INTRODUCTIONS

In recent years, the advancement of UAV technology has driven its growing utilization in civil applications (Guan et al., 2022), including the power industry where airborne point cloud data is used for related applications(C. Chen et al., 2018, 2022). In the field of computer vision and remote sensing, numerous algorithms for UAVs have been developed (Al-Kaff et al., 2018), typically utilizing multimodal data, such as both airborne point clouds and images for alignment(Yang & Chen, 2015). However, fewer algorithms are proposed for real-time mapping based solely on images. In the remote sensing field, aerial mapping is commonly performed by flying a UAV equipped with a GNSS receiver over an area, taking overlapping photos, and processing them with 3D modelling software like Pix4D. However, the processing time of the method varies depending on the quality and quantity of the images, which is time-consuming and unsuitable for fast map building scenarios.

Most UAV mapping systems (Li et al., 2019; Lin et al., 2019) currently use Light Detection And Ranging (LiDAR) sensors for simultaneous localization and mapping(SLAM), resulting in accurate 3D reconstruction of target areas, which is however cost-consuming. Camera-based mapping systems rely on visual and inertial guidance information(Qin et al., 2018), which can lead to state estimation drift in the weakly textured regions or during significant system movement. Dense mapping algorithms require abundant computational resources, making real-time mapping a challenging task without powerful computing units.

This paper proposes a real-time UAV 3D image point clouds mapping algorithm for outdoor environments with weak GNSS and low texture and integrates a UAV mapping system with an RGB camera as the main sensor. The system includes a calibrated sensor suite with a visual-inertial odometry, tightly coupling GNSS, visual and inertial information, and a mapping algorithm on the on-board computer. The algorithm uses the estimated pose and GNSS data to recover scale information, acquired in real-time by computing key frame images. The details of the algorithm are[1] presented in the Section 3 of this paper.

The main contributions of this paper can be summarized in two points as follows:

1. A robust visual-inertial odometry with tightly coupled GNSS data, vision, and inertial data is proposed, achieving competitive accuracy (0.778m) in position estimation. And a real-time mapping algorithm based on block matching is proposed, which utilizes GNSS data to recover scale information in the optimization process, enabling real-time mapping with a 0.447 m accuracy.

2. A low-cost real-time mapping system for UAVs is integrated and the experimental results demonstrate a great potential for practical applications .

The remainder of this paper is structured as follows: Section 2 introduces the related algorithms. Section 3 outlines the proposed system and algorithm. Section 4 presents the experiments and evaluation of the proposed algorithm. Finally, Section 5 concludes the paper.

## 2. RELATED WORK

Current visual-inertial methods for estimating state, as well as mapping, have demonstrated efficacy. However, in certain scenarios, such as in weak GNSS or low texture environments, they may exhibit some shortcomings.

### 2.1 VIO

Visual-inertial odometry is a real-time state estimation technique that utilizes the camera and IMU But it can drift easily, leading to unstable pose estimation, especially in the low-texture and dynamic environments. To address this issue, three main solutions are used: 1) introducing auxiliary exogenous data as constraints, 2) employing multiple primitive features during feature extraction and tracking, and 3) filtering out dynamic features in the environment to retain only static features.

[*] Corresponding author

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-1-W1-2023-1097-2023 | © Author(s) 2023. CC BY 4.0 License.

VINS series(Qin, Cao, et al., 2019; Qin et al., 2018; Qin, Pan, et al., 2019; Qin & Shen, 2018) and ORB-SLAM3(Campos et al., 2021) are popular VIO methods. VINS-Mono (Qin et al., 2018) tightly couples visual and inertial measurements for real-time pose estimation, while VINS-Fusion supports multiple sensor suites(Qin, Pan, et al., 2019), online temporal offset calibration (Qin & Shen, 2018) and GNSS for global coordinate access (Qin, Cao, et al., 2019). ORB-SLAM3 (Campos et al., 2021) track ORB feature points to improve pose estimation.

In low-texture scenes, VIO using only visual information and IMU suffers from drift. To tackle this issue, the data from multi-source sensors are incorporated to increase VIO's robustness. GVINS (Cao et al., 2022) and VINS-RGBD (Shan et al., 2019) respectively combine GNSS data and depth information to improve VIO performance. Alternatively, front-end feature extraction can include other primitive features in the visual scene to solve the few-point-feature problem in low-texture scenario. PLD-VINS (Zhu et al., 2021) uses point features, line features, and depth information, while PL-VINS (Q. Fu et al., 2022) is based on VINS-Mono and uses point and line features, both with improved accuracy over VINS-Mono.

Moreover, to adapt to dynamic scenes in VIO, DS-SLAM (Yu et al., 2018) uses semantic segmentation and movement consistency to reduce the impact of dynamic objects and improve localization accuracy. VINS-Dimc (D. Fu et al., 2022) introduces epipolar constraints on the IMU-derived motion models and uses flow vector bound constraints to filter out deviating features, enhancing system stability in dynamic environments. DynaVINS (Song et al., 2022) improves pose estimation accuracy by discarding the features associated with dynamic and temporary static targets.

Overall, mapping with UAVs faces several challenges such as wind-induced oscillations and low-texture features in aerial images that make it difficult to perform stable flight and pose estimation. To overcome these challenges, we tightly couple GNSS, visual, and IMU data in the VIO framework to provide more stable pose estimation. In addition, GNSS information is also optional to discard in case of poor signal quality.

### 2.2 Mapping

The current image-based mapping algorithms can be categorized into two types based on the mapping results: the one produces 2D image stitching results while the other generates 3D results such as point clouds, Mesh, Grid Map, etc. Map2DFusion (Bu et al., 2016) achieves real-time stitching of aerial images by using monocular SLAM to estimate the position and orientation of camera, but it only generates 2D results without 3D information. TerrainFusion (Wang et al., 2019) generates local DSMs with 3D information in real-time by processing key frames generated by monocular SLAM, which are subsequently accessed into the global DSM. DenseFusion (L. Chen et al., 2020) uses a novel DSM fusion method to generate point clouds, DOM, and DSM in real-time from aerial imagery. OpenREALM (Kern et al., 2020) can acquire mosaics or 3D surface information with different modes of operation. Additionally, Miller et al. propose a mapping algorithm ASOOM (Miller et al., 2022) that can generate maps in the GridMap format in real-time for collaborative air-ground missions.

In our proposed UAV mapping system, we also utilize an image-based dense mapping algorithm based on keyframes from VIO to output point cloud maps and meshes, reducing computational costs and achieving real-time mapping.

Most real-time mapping systems for unmanned aerial vehicles (UAVs) currently rely on LiDAR technology. Zhou et al. proposes a UAV mapping system (Zhou, 2021) that tightly integrates camera and LiDAR data to generate point clouds and uses GNSS information for bundle adjustment. However, this system is computationally expensive. Similarly, Qian et al. proposes a robust mapping system (Qian et al., 2022) that combines vision and LiDAR data to acquire 3D maps with texture information, but it still requires costly LiDAR technology.

In contrast, this paper introduces a low-cost UAV mapping system that uses only image data from an affordable RGB camera to generate 3D maps with RGB information, making it a more accessible and cost-effective option for UAV mapping applications.

## 3. METHDOLOGY

This paper proposes a real-time 3D image point clouds mapping algorithm and integrates a real-time UAV 3D image point clouds mapping system called DCSI LUOJIA Explorer that employs the DJI 300 as the flight platform. The mapping suite comprises a RGB camera, an IMU, a GNSS receiver, a digital transmission module, an image transmission module, and a Nvidia AGX Xavier. Figure 1 illustrates the algorithm framework, which is composed of a pose estimation module and a mapping module. The pose estimation module includes four phases: data input, data pre-processing, initialization, and non-linear optimization. The mapping module uses the pose obtained from the pose estimation module to identify the key frames, which adopts intensive stereo matching to obtain stereo information. The module then assigns the image colour information to the corresponding point cloud and generates a map via an optimization process.

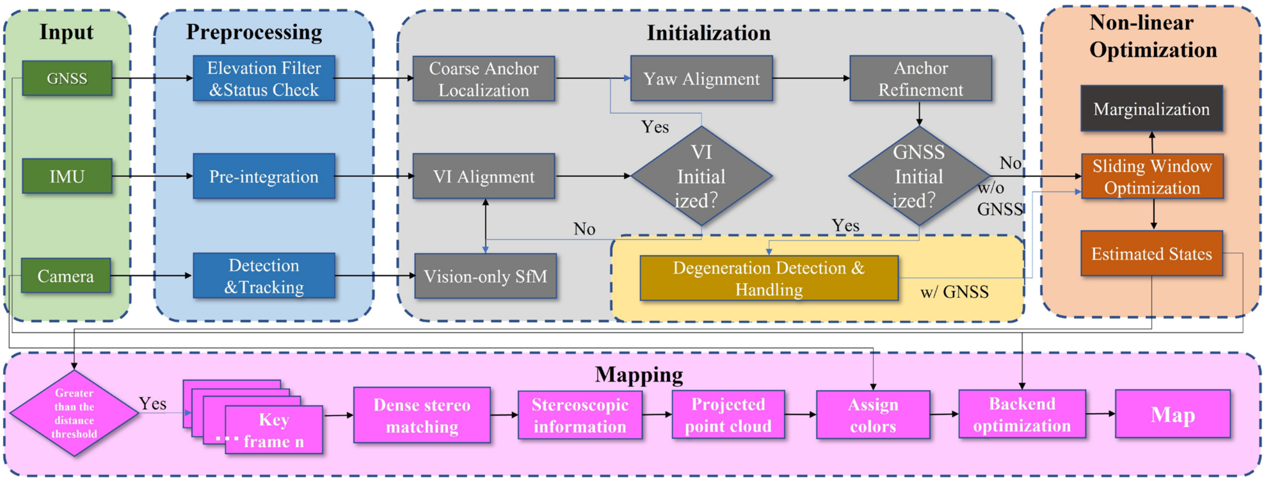


**Figure 1.** Framework of the proposed algorithm

### 3.1 DCSI LUOJIA Explorer system hardware composition

As depicted in Figure 2, the proposed low-cost real-time UAV mapping system called DCSI LUOJIA Explorer comprises of a multisensory suite and an unmanned aircraft platform. The UAV platform chosen for this system is the DJI 300 that has a maximum load capacity of approximately 2.5 kg and can endure a maximum of 30 minutes while carrying the maximum load. The multi-sensor suite includes a GNSS receiver, a digital transmission device, a low-cost IMU, an RGB camera, and a long-range image transmission device.

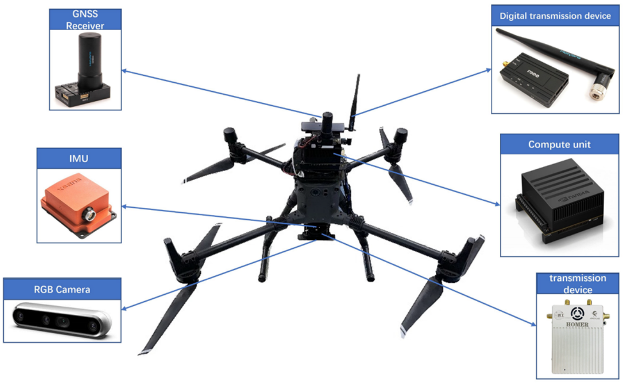


**Figure 2.** DCSI LUOJIA Explorer hardware design

The GNSS receiver used in the system is the Ublox F9P GNSS receiver, which has a GNSS data accuracy of 1.5 m and operates at a reception frequency of 10 Hz. The used low-cost IMU is Xsens Mti300, with an in-run bias of 0.015 mg for the accelerometer and 10/h for the gyroscope during actual operation. The output frequency of the IMU is set to 200 Hz. The camera utilized in the mapping system is Intel Realsense D455, with a camera resolution of 1280*720 and an output frequency of 30 Hz. Notably, only the RGB image from the camera is utilized in the entire system. Table 1 presents the information for the relevant sensors utilized in the system.

| Sensors | descriptions |
|---|---|
| GNSS receiver | Supporting BDS/GPS/GLONASS/Galileo |
| digital transmission device | transmit power from 100mW to 1W;902-928 MHz frequency band; Range up to 60km |
| IMU | accelerometer in-run bias is 0.015mg; gyroscope in-run bias is 10 /h |
| Camera | 1280 ×720; 30Hz |
| Video transmission devices | 100Mbps; 5.1~5.9GHz; Range up to 2-4km; Delay 150ms |

**Table 1.** Sensor Information.

### 3.2 Coordinate systems involved in the experiment

There are three coordinate systems involved in the experiment.

1. Sensor coordinate system. The IMU coordinate system is the coordinate system of Xsens Mti-300 in our system. The camera coordinate system is a Cartesian 3D coordinate system with the RGB camera Intel Realsense D455 as the origin, and the relationship between the two coordinate systems is determined by advanced calibration.

2. Local coordinate system. This coordinate system uses the position of the UAV mapping system at take-off as the origin, with the X-axis facing forward, the Y-axis facing left, and the Z-axis facing up, and the visual odometry output takes this coordinate system as the reference system.

3.ENU coordinate system, with a point on the earth reference ellipsoid as the origin, the XYZ axes of this coordinate system point to the east, north and gravity inverse direction respectively.

### 3.3 Tightly coupled GNSS-visual-inertial Fusion

To achieve a stable and robust mapping algorithm, a reliable visual inertial odometry (VIO) system that can operate smoothly during UAV flight is necessary. However, VIOs used for estimating system states often face the issue of instability and drift, as shown in popular VIO frameworks like the VINS series and ORB-SLAM3. In practical scenarios, high-speed UAV flights at high altitudes or system jitter due to control can cause image magnitudes to shake and overall system accelerations to be significant, leading to VIO drift. To address this drift issue, this paper proposes a self-positioning method that integrates GNSS data.

The proposed visual-inertial odometry in this paper addresses the issue of drift in VIO by incorporating GNSS data in a tightly-coupled manner, resulting in better mapping performance and increasing robustness. The self-localization framework, as illustrated in Figure 1, consists of a data pre-processing module, initialization module, and non-linear optimization module, with raw GNSS observation data, IMU data, and image data from the camera serving as inputs. In the pre-processing phase, GNSS data are altitude-filtered to remove lower altitude data, and only stable satellite signals from satellites that have been continuously locked for a period are accessed. IMU data are pre-integrated, and feature points are extracted and tracked from the image data. The three types of data are pre-processed and then used to initialize the system state. During initialization, visual-only SfM operations are first performed to obtain an initial motion estimate, and the trajectory obtained from the IMU are aligned with the visual-only SfM results to recover velocity, scale, and gravity. If the visual inertial alignment is successful, a coarse-to-fine initialization process is conducted for GNSS data. This process includes obtaining a coarse anchor position using the SPP algorithm, aligning the local and global coordinate systems by using local velocity obtained from the visual inertial alignment process and GNSS Doppler measurements to achieve yaw alignment, and refining the global position of the anchor using precise local trajectory and clock constraints. Once initialization is completed, the system can continuously monitor and process the GNSS signal to prevent signal degradation.

If the GNSS signal becomes weak or absent, the GNSS initialization will fail, and the self-localization module will switch to VIO-only mode, using only visual inertial data for state estimation. In the non-linear optimization stage, a two-way marginalization method is added after sliding window optimization for actions that may cause visual inertial drift, such as rotation. This method removes certain frames based on the parallax test, ensuring real-time operation and system robustness, and finally outputs the estimated pose.

The state estimation problem is formulated and derived using a probabilistic framework. This will allow us to model the uncertainties and correlations involved in the system and measurement processes, and to arrive at a more accurate and reliable estimate of the system state. All the problem is structured as a factor graph, where the measurements from sensors are represented as a series of interconnected factors. These factors, in turn, constrain the system states and help to determine the overall behavior of the system. To define the optimum system state, we aim to maximize the posterior probability based on all the measurements. Assuming that each measurement is independent of the others and has a Gaussian-distributed noise with a zero mean, we can transform the MAP problem into a cost minimization problem where each cost corresponds to a specific measurement.

$$\begin{aligned}\mathcal{X}^\star &= \arg\max_{\mathcal{X}} p(\mathcal{X} \mid \mathbf{a}) \\ &= \arg\max_{\mathcal{X}} p(\mathcal{X}) p(\mathbf{a} \mid \mathcal{X}) \\ &= \arg\max_{\mathcal{X}} p(\mathcal{X}) \prod_{i=1}^{n} p(\mathbf{a}_i \mid \mathcal{X}) \\ &= \arg\min_{\mathcal{X}} \left\{ \|\mathbf{r}_p - \mathbf{H}_p \mathcal{X}\|^2 + \sum_{i=1}^{n} \|\mathbf{r}(\mathbf{a}_i, \mathcal{X})\|^2_{\mathbf{P}_i} \right\},\end{aligned} \tag{1}$$

where $\mathcal{X}$ is the system state and $\mathcal{X}^\star$ is the optimum system state. $\{\mathbf{r}_p, \mathbf{H}_p\}$ represents the prior information about the system state, while $\mathbf{a}$ is the aggregation of $n$ independent sensor measurements. $\mathbf{r}(\cdot)$ denotes the residual function of each measurement, and $\|\cdot\|_{\mathbf{P}}$ is the Mahalanobis norm used to evaluate the similarity between the measurements and the prior state information.

The experimental process utilizes a soft synchronization method for time synchronization, where GNSS time is obtained through UbloxDriver and local time is adjusted to GNSS time.

In this paper, experimental testing of the proposed VIO framework shows that it can accurately estimate the attitude of the UAV from take-off to landing without any drift, demonstrating its robustness and effectiveness during UAV flight.

### 3.4 Real-time 3D mapping based on RGB image

In the experimental tests, the mapping algorithm proposed in this paper has demonstrated its ability to perform real-time mapping with reliable accuracy. The algorithm utilizes the odometry from the real-time output of the state estimator in Section 3.3 as an exogenous data input and creates a new keyframe with recorded data such as keyframe image, odometry, and GNSS data once the UAV has moved a threshold distance. Dense stereo data is computed by block matching. In equation (2), the depth value for a given pixel location is determined by minimizing the sum of the differences between the grayscale values at corresponding locations in two images. The size of the matching window is denoted by $w$ and $d$ represents the displacement amount within the search window. The grayscale value is denoted by $I$.

$$D_L(u,v) = \arg, \min_{d} \sum_{x=u-w}^{u+w} \sum_{y=v-w}^{v+w} \big(I_L(x,y) - I_R(x-d,y)\big)^2 \tag{2}$$

And then dense stereo data is then projected into the point cloud with each point aligned to the corresponding cell in the 2D plane. Additionally, the colour of the corresponding pixel closest to the image center is assigned to each cell.

To recover the scale information of the map and solve the scale problem using monocular visual odometry, pose estimates and GNSS data are used by an optimizer.

## 4. EXPERIMENTS

### 4.1 Experimental site and flight setup

To verify the effectiveness of the proposed mapping system and evaluate the mapping results, an experiment was conducted in an abandoned playground. The old playground was chosen as the test site, and a satellite imagery of the area is shown in Figure 4.

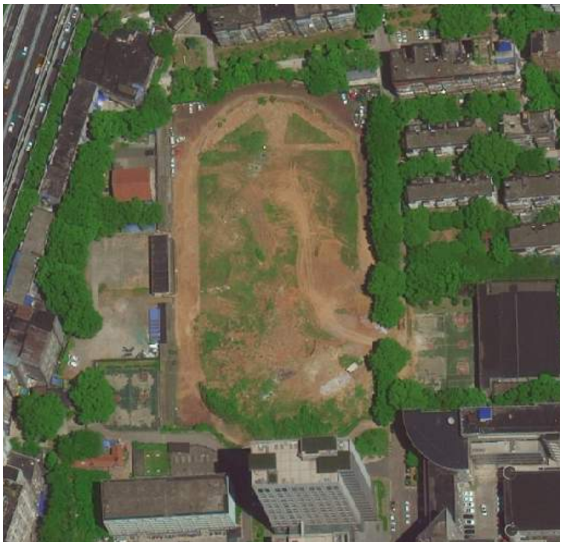


**Figure 4.** The test site.

The UAV was flown at an altitude of approximately 40 meters above the ground, with a speed of 3 meters per second for data collection. The IMU frequency was set to 200 Hz, while the RGB camera had a resolution of 1280 * 720 pixels and operated at a frequency of 30 Hz. The GNSS data was set to a frequency of 10 Hz, but during the actual experiment, the GNSS frequency was between 7 to 8 Hz and the UAV is controlled to fly longitudinally and horizontally over the test site for approximately 30 minutes. The details of the experimental setup are provided in Table 2.

| Experimental setup | Settings |
|---|---|
| Site | The old playground |
| Flight Altitude (From the Ground) | 40m |
| IMU Frequency | 200Hz |
| Camera Resolution | 1920 * 720 |
| Image Frequency | 30Hz/10Hz |
| GNSS Frequency | 10Hz(7~8Hz) |
| Duration of the flight | 30min |

**Table 2.** Experimental setup.

### 4.2 Ground truth acquisition

The 3D ground truth for the experiment was reconstructed by Pix4D with the images captured by the UAV. In the reconstruction process, one frame was extracted every 5 seconds from the data packet acquired during a section of the UAV's trajectory, and the corresponding latitude, longitude, and altitude of the GNSS receiver data were recorded. A total of 110 frames were extracted and input into the PiX4D software for 3D reconstruction.

However, due to the 1.5m accuracy of the GNSS data used during the UAV flight, there may be some error in the ground truth. To minimize this error, 23 control points with precise geographic coordinates were collected at the experimental site using the same receiver before the reconstruction process. During the acquisition of the control points, the precise RTK measurements of control points were obtained. A total of 23 control points were acquired using this method in the experiment.

The images obtained by frame extraction and the corresponding geographic coordinates file were imported into Pix4D along with the control point files. Control point punctures were conducted in Pix4D for the corresponding locations in the images as constraints, and the Pix4D software was then started for 3D reconstruction. The initialization process took 4 minutes and 41 seconds, point cloud generation took 47 seconds, 3D texture generation took 34 seconds, and DSM generation took 16 seconds, for a total of 6 minutes and 18 seconds. The accuracy of the reconstructed 3D map using PiX4D is shown in Table 3, and the accuracy in the X, Y, Z axes reaches 0.01m, 0.014m, and 0.006m, respectively. The indicators thus demonstrate that the constructed map can be used as the ground truth. The 3D model constructed by PiX4D is shown in Figure 5, and it is observed that the constructed 3D map is highly accurate.。

| RPE-X(m) | RPE-Y(m) | RPE-Z(m) | Duration |
|---|---|---|---|
| 0.010 | 0.014 | 0.006 | 6min18s |

**Table 3.** Pix4D 3D reconstruction accuracy.

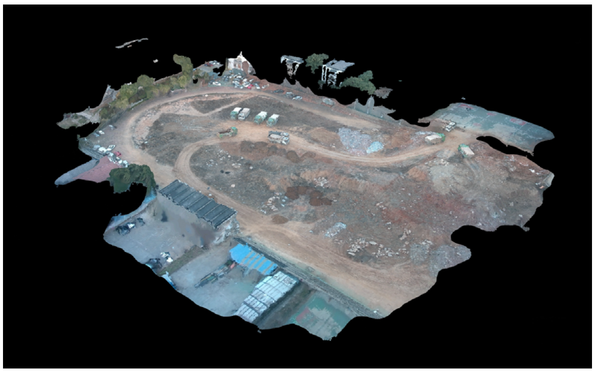
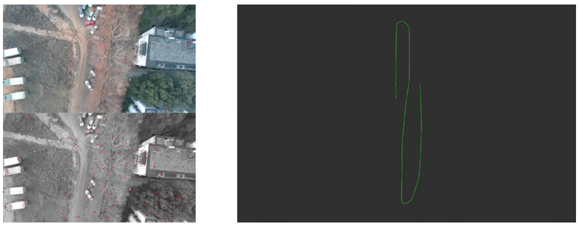


**Figure 5.** Mesh built by Pix4D.

### 4.3 Visual-inertial Odometry result

Experiments were conducted with data collected during a flight over the old playground. Figure 6 illustrates the real-time attitude estimation process using the low-cost UAV real-time mapping system proposed in this paper. In Figure 6 (a), there is an RGB image and an image with sparse feature points extracted, respectively while in Figure 6 (b) the pose estimator proposed in this paper performs pose estimation of the UAV system to obtain the estimated odometry. It can be qualitatively judged that the attitude estimator proposed in this paper can estimate the system state of the UAV flying at high altitude in a more stable manner, while also achieving real-time mapping.



(a) (b)

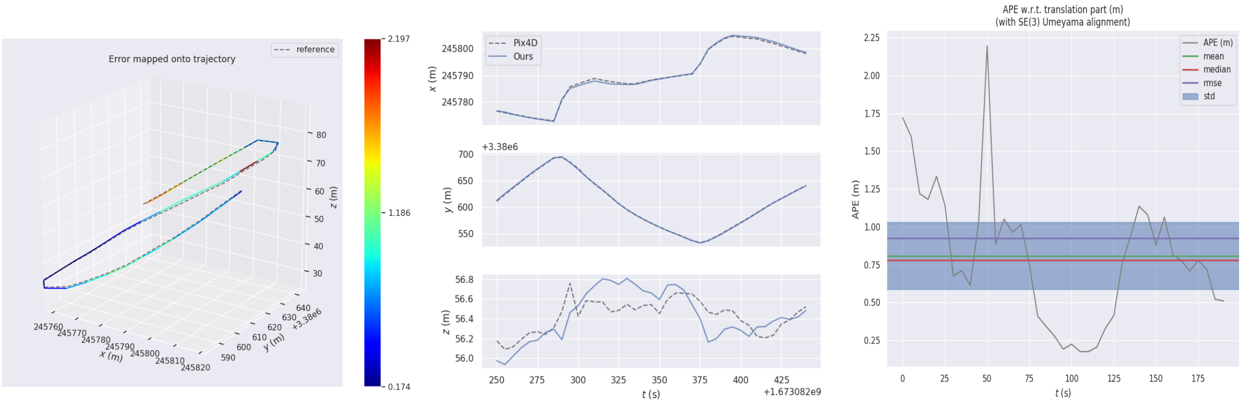
**Figure 6.** Real-time pose estimation.

To evaluate the accuracy of the proposed attitude estimation method, the second half of the UAV flight data was selected since the GNSS data in the initial phase of the flight were unstable. After trajectory alignment, the accuracy evaluation results are presented in Table 4. The results demonstrate that the maximum error is 2.197m, the average error is 0.808m, the median error is 0.778m, the minimum error is 0.174m, and the root mean square error is 0.922m. It is evident that the visual inertial odometry with tightly coupled GNSS proposed in this paper can achieve decimeter-level accuracy, surpassing the 1.5m accuracy of the used GNSS. These results suggest that the proposed pose estimation method can offer stable and precise pose estimation for the mapping system and meet the mapping requirements of the UAV real-time mapping system.

| Max | Mean | Median | Min | RMSE | STD |
|---|---|---|---|---|---|
| 2.197 | 0.808 | 0.778 | 0.174 | 0.922 | 0.445 |

**Table 4.** Accuracy of the proposed VIO method.

The accuracy evaluation results obtained using the EVO tool are presented in Figure 7. In Fig. 7(a), the experimental trajectory is compared with the reference true trajectory, and in Fig. 7(b), the differences between the X, Y, and Z coordinates and the true values are plotted, respectively. It can be observed that the experimental results are in good agreement with the true values in the X and Y directions, while some instability is observed in the Z direction, which may be attributed to the lower accuracy of the GNSS data used in this direction. The statistical plot of the experimental trajectory accuracy variation over time is shown in Figure 7 (c).



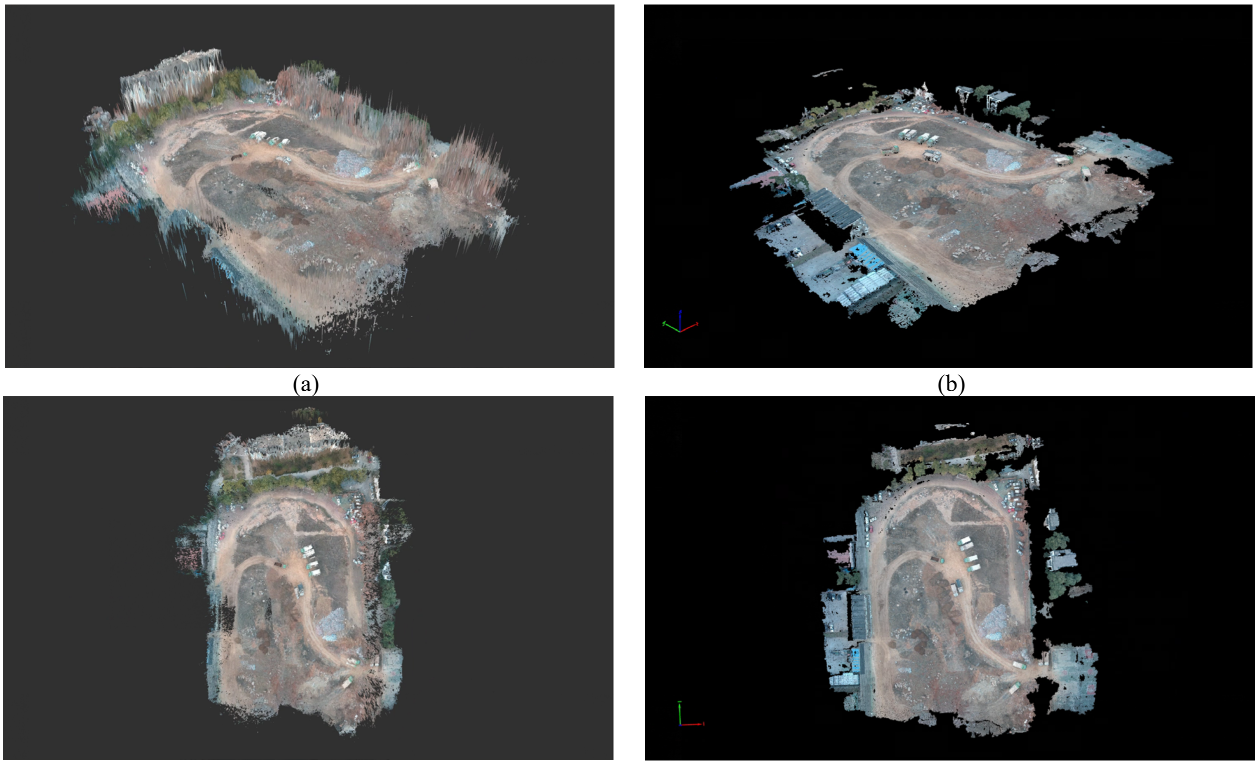
(a) (b) (c)

**Figure 7.** Visualization of trajectory accuracy evaluation.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-1-W1-2023-1097-2023 | © Author(s) 2023. CC BY 4.0 License.

### 4.4 Mapping result

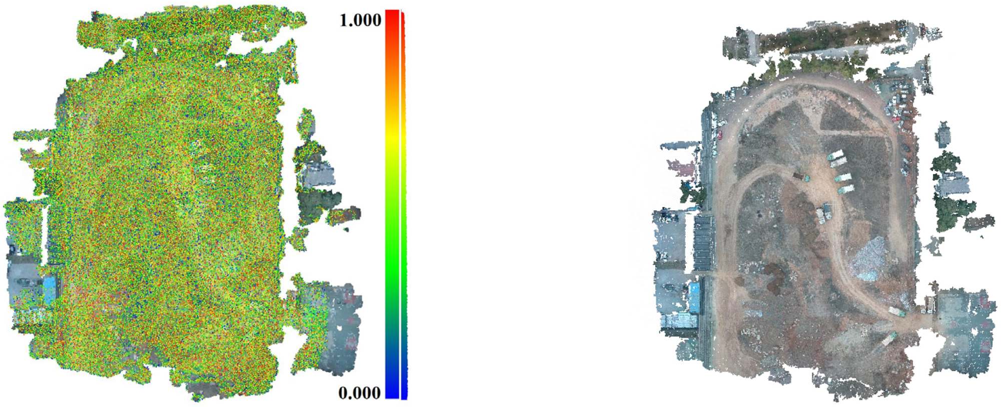
The experimentally generated 3D map was first qualitatively compared with the ground truth generated using Pix4D, as shown in Figure 8. The 3D map reconstructed by our algorithm is shown in Figure 8 (a) and Figure 8 (c), while Figure 8 (b) and Figure 8 (d) represent the map obtained using Pix4D, which is considered as ground truth. The comparison revealed that the map obtained by our mapping algorithm is almost as good as the one reconstructed using Pix4D. However, it does not work as well as Pix4D in processing the images of the edge part of the flight route. Specifically, the map built by our algorithm will have some sharp raised parts in the edge part, which is due to the less overlapping part of the edge image. As a result, the dense matching algorithm cannot estimate the elevation of this part. Unlike Pix4D, our algorithm does not utilize filtering process. Nonetheless, our proposed mapping algorithm is real-time and can update the map once per second, making it far more efficient than Pix4D.



(a) (b) (c) (d)

**Figure 8.** Comparison of the built map with the ground truth.

Besides, the point cloud data generated by our mapping algorithm was saved and compared with the point cloud data generated by Pix4D. However, due to the large number of points generated by our algorithm, the whole dense point cloud has a large file size. To improve the efficiency of the experiment, we first performed a thinning operation and then calculated the cloud to cloud distances using Cloud Compare with the point cloud generated by Pix4D. The results obtained are shown in Figure 9, where the average distance was found to be 0.447, and the standard deviation was 0.279. Specifically, Figure 9 (a) displays the result of comparing the distance between the experimentally generated point cloud and the point cloud generated by Pix4D and the distribution of the number of points in different distance ranges, and Figure 9 (b) displays the point cloud generated by Pix4D.



(a) (b)

**Figure 9.** Point cloud distances calculation.

## 5. CONCLUSIONS AND FUTURE WORK

In this paper, we present a real-time UAV 3D image point clouds mapping algorithm and integrate a low-cost UAV mapping system, which includes an RGB camera, a GNSS receiver, an IMU, and image and digital data transmission devices. The system utilizes a visual inertial odometry with tightly coupled GNSS, visual, and inertial data to perform robust real-time state estimation of the UAV. Furthermore, a block matching-based mapping algorithm is employed to achieve real-time mapping. The experimental results demonstrate that the proposed system can accomplish accurate and efficient real-time mapping. Notably, the accuracy of the system's visual inertial odometry can reach the decimetre level, showcasing the significant potential of the proposed system for 3D mapping applications. However, to obtain stable mapping and pose estimation, the proposed method still requires GNSS data. In a GNSS-denied environment, the proposed method might be limited. Introducing point cloud data at this time can help solve this problem. Furthermore, real-time semantic mapping will be explored in our future work.

## ACKNOWLEDGEMENTS

This research was funded by the National Natural Science Foundation of China (No.U22A20568, No.42071451), the National Key RESEARCH and Development Program (No.2022YFB3904101), the National Key RESEARCH and Development Program-Key Special Projects for International Cooperation in Science and Technology Innovation between Governments (No.2019YFE0123300), Natural Science Foundation of Hubei, China (No.2022CFB007), the National Natural Science Foundation of China (No.42130105), the National Science Fund for Distinguished Young Scholars (No.41725005), the European Union’s Horizon 2020 Research and Innovation Program (No.871149).

## REFERENCES

Al-Kaff, A., Martín, D., García, F., Escalera, A. de la, & María Armingol, J. (2018). Survey of computer vision algorithms and applications for unmanned aerial vehicles. *Expert Systems with Applications*, *92*, 447–463. https://doi.org/10.1016/j.eswa.2017.09.033

Bu, S., Zhao, Y., Wan, G., & Liu, Z. (2016). Map2DFusion: Real-time incremental UAV image mosaicing based on monocular SLAM. *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 4564–4571. https://doi.org/10.1109/IROS.2016.7759672

Campos, C., Elvira, R., Rodríguez, J. J. G., Montiel, J. M. M., & Tardós, J. D. (2021). ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual-Inertial and Multi-Map SLAM. *IEEE Transactions on Robotics*, *37*(6), 1874–1890. https://doi.org/10.1109/TRO.2021.3075644

Cao, S., Lu, X., & Shen, S. (2022). GVINS: Tightly Coupled GNSS–Visual–Inertial Fusion for Smooth and Consistent State Estimation. *IEEE Transactions on Robotics*, 1–18. https://doi.org/10.1109/TRO.2021.3133730

Chen, C., Jin, A., Yang, B., Ma, R., Sun, S., Wang, Z., Zong, Z., & Zhang, F. (2022). DCPLD-Net: A diffusion coupled convolution neural network for real-time power transmission lines detection from UAV-Borne LiDAR data. *International Journal of Applied Earth Observation and Geoinformation*, *112*, 102960. https://doi.org/10.1016/j.jag.2022.102960

Chen, C., Yang, B., Song, S., Peng, X., & Huang, R. (2018). *Automatic Clearance Anomaly Detection for Transmission Line Corridors Utilizing UAV-Borne LIDAR Data*. 21.

Chen, L., Zhao, Y., Xu, S., Bu, S., Han, P., & Wan, G. (2020). DenseFusion: Large-Scale Online Dense Pointcloud and DSM Mapping for UAVs. *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 4766–4773. https://doi.org/10.1109/IROS45743.2020.9341413

Fu, D., Xia, H., Liu, Y., & Qiao, Y. (2022). VINS-Dimc: A Visual-Inertial Navigation System for Dynamic Environment Integrating Multiple Constraints. *ISPRS International Journal of Geo-Information*, *11*(2), 95. https://doi.org/10.3390/ijgi11020095

Fu, Q., Wang, J., Yu, H., Ali, I., Guo, F., He, Y., & Zhang, H. (2022). *PL-VINS: Real-Time Monocular Visual-Inertial SLAM with Point and Line Features* (arXiv:2009.07462). arXiv. http://arxiv.org/abs/2009.07462

Guan, S., Zhu, Z., & Wang, G. (2022). A Review on UAV-Based Remote Sensing Technologies for Construction and Civil Applications. *Drones*, *6*(5), 117. https://doi.org/10.3390/drones6050117

Kern, A., Bobbe, M., Khedar, Y., & Bestmann, U. (2020). OpenREALM: Real-time Mapping for Unmanned Aerial Vehicles. *2020 International Conference on Unmanned Aircraft Systems (ICUAS)*, 902–911. https://doi.org/10.1109/ICUAS48674.2020.9213960

Li, J., Yang, B., Chen, C., & Habib, A. (2019). NRLI-UAV: Non-rigid registration of sequential raw laser scans and images for low-cost UAV LiDAR point cloud quality improvement. *ISPRS Journal of Photogrammetry and Remote Sensing*, *158*, 123–145. https://doi.org/10.1016/j.isprsjprs.2019.10.009

Lin, Y.-C., Cheng, Y.-T., Zhou, T., Ravi, R., Hasheminasab, S., Flatt, J., Troy, C., & Habib, A. (2019). Evaluation of UAV LiDAR for Mapping Coastal Environments. *Remote Sensing*, *11*(24), 2893. https://doi.org/10.3390/rs11242893

Miller, I. D., Cladera, F., Smith, T., Taylor, C. J., & Kumar, V. (2022). *Stronger Together: Air-Ground Robotic Collaboration Using Semantics* (arXiv:2206.14289). arXiv. http://arxiv.org/abs/2206.14289

Qian, J., Chen, K., Chen, Q., Yang, Y., Zhang, J., & Chen, S. (2022). Robust Visual-Lidar Simultaneous Localization and Mapping System for UAV. *IEEE Geoscience and Remote Sensing Letters*, *19*, 1–5. https://doi.org/10.1109/LGRS.2021.3099166

Qin, T., Cao, S., Pan, J., & Shen, S. (2019). *A General Optimization-based Framework for Global Pose Estimation with Multiple Sensors* (arXiv:1901.03642). arXiv. http://arxiv.org/abs/1901.03642

Qin, T., Li, P., & Shen, S. (2018). VINS-Mono: A Robust and Versatile Monocular Visual-Inertial State Estimator. *IEEE Transactions on Robotics*, *34*(4), 1004–1020. https://doi.org/10.1109/TRO.2018.2853729

Qin, T., Pan, J., Cao, S., & Shen, S. (2019). *A General Optimization-based Framework for Local Odometry Estimation with Multiple Sensors* (arXiv:1901.03638). arXiv. http://arxiv.org/abs/1901.03638

Qin, T., & Shen, S. (2018). Online Temporal Calibration for Monocular Visual-Inertial Systems. *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 3662–3669. https://doi.org/10.1109/IROS.2018.8593603

Shan, Z., Li, R., & Schwertfeger, S. (2019). RGBD-Inertial Trajectory Estimation and Mapping for Ground Robots. *Sensors*, *19*(10), 2251. https://doi.org/10.3390/s19102251

Song, S., Lim, H., Lee, A. J., & Myung, H. (2022). *DynaVINS:* A Visual-Inertial SLAM for Dynamic Environments. *IEEE Robotics and Automation Letters*, *7*(4), 11523–11530. Q2. https://doi.org/10.1109/LRA.2022.3203231

Wang, W., Zhao, Y., Han, P., Zhao, P., & Bu, S. (2019). TerrainFusion: Real-time Digital Surface Model Reconstruction based on Monocular SLAM. *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 7895–7902. https://doi.org/10.1109/IROS40897.2019.8967663

Yang, B., & Chen, C. (2015). Automatic registration of UAV-borne sequent images and LiDAR data. *ISPRS Journal of Photogrammetry and Remote Sensing*, *101*, 262–274. https://doi.org/10.1016/j.isprsjprs.2014.12.025

Yu, C., Liu, Z., Liu, X.-J., Xie, F., Yang, Y., Wei, Q., & Fei, Q. (2018). DS-SLAM: A Semantic Visual SLAM towards Dynamic Environments. *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 1168–1174. https://doi.org/10.1109/IROS.2018.8593691

Zhou, T. (2021). Tightly-coupled camera/LiDAR integration for point cloud generation from GNSS/INS-assisted UAV mapping systems. *ISPRS Journal of Photogrammetry and Remote Sensing*.

Zhu, Y., Jin, R., Lou, T., & Zhao, L. (2021). PLD-VINS: RGBD visual-inertial SLAM with point and line features. *Aerospace Science and Technology*, *119*, 107185. https://doi.org/10.1016/j.ast.2021.107185