Hindawi
Complexity
Volume 2018, Article ID 7358491, 9 pages
https://doi.org/10.1155/2018/7358491

Research Article

An RGB-D-Based Cross-Field of View Pose Estimation System for a Free Flight Target in a Wind Tunnel

Sheng Liu, Yuan Feng, Kang Shen, Yangqing Wang, and Shengyong Chen

College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou, 310023, China

Correspondence should be addressed to Sheng Liu; edliu@zjut.edu.cn

Received 17 August 2018; Accepted 28 October 2018; Published 2 December 2018

Guest Editor: Andy Annamalai

Copyright © 2018 Sheng Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Estimating the real-time pose of a free flight aircraft in a complex wind tunnel environment is extremely difficult. Due to the high dynamic testing environment, complicated illumination condition, and the unpredictable motion of target, most general pose estimating methods will fail. In this paper, we introduce a cross-field of view (FOV) real-time pose estimation system, which provides high precision pose estimation of the free flight aircraft in the wind tunnel environment. Multiview live RGB-D streams are used in the system as input to ensure the measurement area can be fully covered. First, a multimodal initialization method is developed to measure the spatial relationship between the RGB-D camera and the aircraft. Based on all the input multimodal information, a so-called cross-FOV model is proposed to recognize the dominating sensor and accurately extract the foreground region in an automatic manner. Second, we develop an RGB-D-based pose estimation method for a single target, by which the 3D sparse points and the pose of the target can be simultaneously obtained in real time. Many experiments have been conducted, and an RGB-D image simulation based on 3D modeling is implemented to verify the effectiveness of our algorithm. Both the real scene's and simulation scene's experimental results demonstrate the effectiveness of our method.

1. Introduction

Aircraft attitude estimation plays a crucial role in aircraft control systems of the wind tunnel. During the flight of the aircraft, it is essential to adjust the flight parameter according to the real-time attitude of the aircraft [1]. And while verifying the flight performance of aircraft, it is also necessary to check the performance of the aircraft in different attitudes. Attitude estimation is an important part of this [2]. In computer vision task, aircraft attitude estimation can be regarded as an object pose estimation task. Vision system is the most widely used technique for measurements in the wind tunnel, which can provide crucial data that can be compared with computational fluid dynamics (CFD) predictions to assist validating design geometries.

However, it is hard to get satisfying precision measurement results in a low-speed wind tunnel when the target is flying freely. The main reason is that the wind tunnel is a high dynamic testing environment and there often exist a complex illumination condition and unpredictable motion of the target. These will heavily decrease the measurement accuracy of the estimation system.

The size of a low-speed wind tunnel is eight meters long, six meters wide, and six meters high. Thus, multiview live RGB-D streams are needed in the system as input to ensure the measurement area can be fully covered. Furthermore, a multimodal initialization method is developed to measure the spatial relationship between the RGB-D camera and the aircraft. Based on all the input multimodal information, our cross-FOV model is proposed to recognize the dominating sensor and accurately extract the foreground region in an automatic manner.

The object pose estimation task has been extensively studied [3]. Traditional methods of object attitude measurement are mainly divided into template matching and feature matching [4]. Template matching methods [5] are usually applied to weakly textured scenes. Such methods need to reconstruct 3D objects and then match real scenes with 3D models to find the best pose. The classic ICP algorithm [3] and RANSAC algorithm solve the current pose by minimizing the

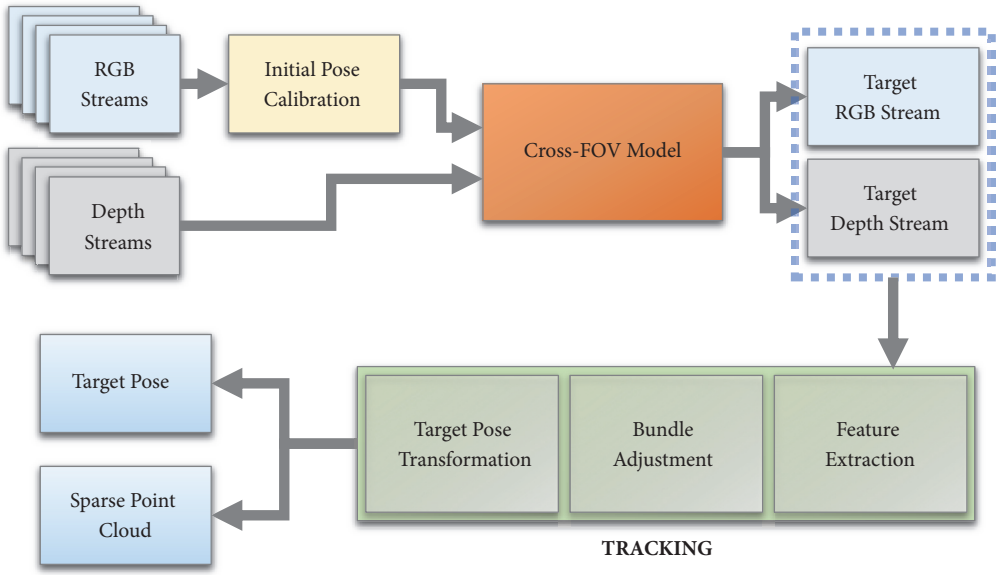


Figure 1: Overview of our method showing the data-flow starting from RGB-D streams. RGB streams are used to calibrate the initial pose of the aircraft. Our cross-FOV model can process multiple RGB-D streams and output the target stream. By tracking the target, our method can get the pose and point cloud of the target.

distance between corresponding points of the actual scene and the model [6]. Many people believe that, in computer vision applications [7–11], the contour of an object is the most reliable information, because feature-based recognition methods [12–15] are likely to fail when recognizing poses of weakly textured objects.

In this paper, to overcome those problems, we propose a cross-field pose estimation framework based on local features to estimate the pose of an aircraft in real time and deal with cross-field problems. By acquiring the relative positional relationship between the camera and the aircraft, we transform the relative pose of the camera into the relative pose of the aircraft. According to the experiment results, the pose estimation system with the cross-FOV model can get accurate measurement results in a wind tunnel. And we run our pose estimation system to measure the pose of an aircraft model. We will apply our system on the low-speed wind tunnel of China Aerodynamics Research and Development Centre (CARDC).

2. Overview of Our Method

We proposed a cross-FOV RGB-D pose estimation system that processes each new frame in real time. Also maintaining high precision pose estimation our system reconstructs the sparse point cloud for the object in the scene and can track the target motion continuously when the target moves across the different field of view. Figure 1 illustrates the frame-to-frame operation of our system and Figure 2 illustrates the structure diagram of our system in a real scene.

2.1. Pose Initialization. In this section, a relative attitude measurement module is utilized to obtain the relative attitude between the aircraft and initial camera. We set two tags at the X-axis of the aircraft shown in Figure 3. The center of the tags is the center of the aircraft. Once the tag's location is detected, the center of the aircraft is localized. Then the system can transform it into relative attitude between the aircraft and initial camera.

2.1.1. Tag Recognition. This module is used to detect the position of tags in Figure 3. We use a tag detector to detect tags following the proposed method AprilTag [5]. The first step is adopting an adaptive thresholding approach to threshold the input grayscale image into a black-and-white image. The next step is segmenting the edges based on the characteristic of the black-and-white components from which they arise to find edges which might form the boundary of a tag. Finally, the method computes a proximate partition by searching for a small number of corner points and then iterates through all possible combinations of corner points to find all fitting quad. After this whole operation, a tag is localized in the image coordinate system and the center of two tags represents the center of the aircraft.

2.1.2. Aircraft Center Localization. We transform the coordinate (u, v) of aircraft center in image coordinate with its corresponding depth value d into a 3D coordinate so that relative attitude between the aircraft and initial camera is obtained:

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = \begin{bmatrix} (u - u_0)\dfrac{d}{f_x} \\ (v - v_0)\dfrac{d}{f_y} \\ d \end{bmatrix}, \quad (1)$$

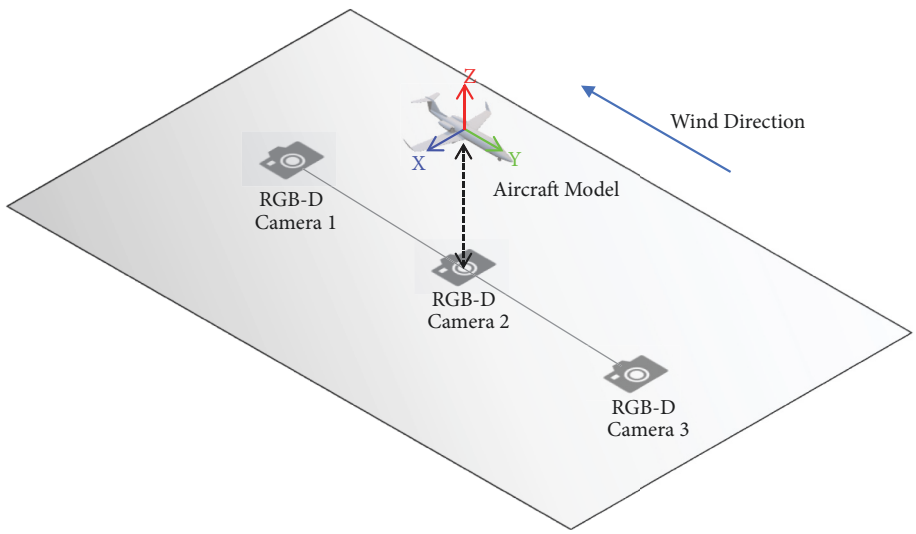


FIGURE 2: The structure diagram of our system.

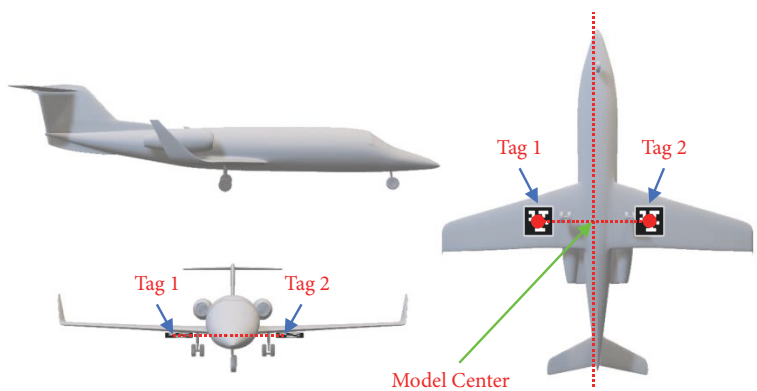


FIGURE 3: The location of visual tag on the aircraft model.

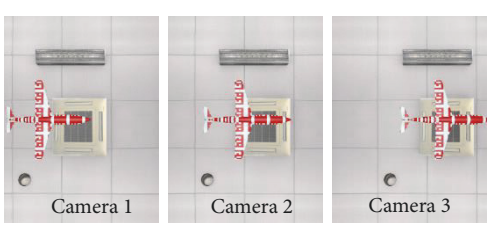


FIGURE 4: Multiple RGB stream inputs.

where (f_x, f_y) is the focal length and (u_0, v_0) is the principal point and all can be known from calibration. (X_c, Y_c, Z_c) is the 3D coordinate of aircraft center.

2.2. Cross-FOV Model. In the wind tunnel, as shown in Figure 4, the cross-field of view measurement is needed. Thus, we designed a cross-FOV model. For camera n in camera set N at the time i, we have the color frame f_i^n and depth frame d_i^n. To choose the best input camera n_i, we need to backtrack M frames to get the maximum frame score camera set $\mathcal{N}_{\mathrm{M}}$. The frame score can be expressed by

$$S\left(d_i^n\right) = \sum_{x=1}^{\mathbb{X}} \sum_{y=1}^{\mathbb{Y}} p\left(x, y\right), \tag{2}$$

$$p\left(x, y\right) = \begin{cases} 1, & d_i^n\left(x, y\right) < \mathcal{H}, \\ 0, & d_i^n\left(x, y\right) \geq \mathcal{H}, \end{cases} \tag{3}$$

where $p(x, y)$ is the depth score at $d_i^n(x, y)$ that fit the depth constraint $\mathcal{H}$. $\mathbb{X}$ and $\mathbb{Y}$ are, respectively, the frame width and height.

The best input camera n_i is chosen from the camera set N to find the maximum score

$$n_i = \arg\max_{n \in N} \left\{ S\left(d_i^n\right) \cdot \frac{\psi\left(n\right)}{M} \right\}, \tag{4}$$

Table 1: RMSE of estimate translation and rotation for synthetic sequences.

Sequence	Ours		Co-fusion [3]	
	E_{trans} (m)	E_{rotate} (°)	E_{trans} (m)	E_{rotate} (°)
rotate1	0.0016	0.4205	-	-
rotate2	0.0017	0.3391	-	-
rotate3	0.0027	1.5700	-	-
translate1	**0.0032**	0.3941	0.0263	-
translate2	**0.0075**	1.3766	0.0560	-
translate3	**0.0091**	0.6858	0.0412	-
translate&rotate	0.0245	0.4706	-	-

Table 2: Parameters of the 6-Axis Hexapod.

Parameter Name	Value	Unit
Travel range X, Y	±50	mm
Travel range Z	±25	mm
Travel range θ_X, θ_Y	±15	°
Travel range θ_Z	±30	°
Min. incremental motion X, Y	3	μm
Min. incremental motion Z	1	μm
Min. incremental motion $\theta_X, \theta_Y, \theta_Z$	5	urad

where $\psi(n)$ is a proportion calculate function, which will calculate the amount of camera number n in camera set $\mathcal{N}_M$.

2.3. Feature Extraction. The system utilizes a fast binary descriptor called ORB for the feature extracting task. This descriptor is rotation invariant and resistant to noise and illumination changes. Simultaneously, it is fast to extract and match which makes ORB suitable for real-time pose estimation work in a complex environment.

Our system handles RGB-D input. We extract ORB features on the RGB image for tracking and, for each feature with coordinates (u, v) and its corresponding depth value d, we transform them into a world coordinate system according to (1).

2.4. Bundle Adjustment. After the initialization operation, our system performs bundle adjustment, to minimize the reprojection error between the 3D point X^i and its corresponding 2D point x_p^i to estimate the camera's instantaneous pose $\{R, t\}$ relative to previous frame

$$\{R, t\} = \arg \min_{R,t} \sum_{i \in \chi} \rho \left(\left\| x_p^i - \phi \left(RX^i + t \right) \right\|_\Sigma^2 \right), \quad (5)$$

where ρ is the robust Huber cost function and Σ the covariance matrix associated with the scale of the keypoint. The projection function ϕ is defined as follows:

$$\phi \left(\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \right) = \begin{bmatrix} f_x \dfrac{X}{Z} + u_0 \\ f_y \dfrac{Y}{Z} + v_0 \end{bmatrix}, \quad (6)$$

where (f_x, f_y) is the focal length and (u_0, v_0) is the principal point, both known from the calibration.

3. Experiments

In the evaluation stage, we carry out a quantitative evaluation on both synthetic and real sequences with ground truth data. Our synthetic experimental sequences which imitate a real experimental environment are specifically designed for this work. In the synthetic scene, we set up ambient light and multiple light sources to simulate the real complex lighting conditions. As for camera settings of synthetic data, we set it up with settings of an Asus Xtion camera with resolution of 640◊480 pixels and field of view in 58°H, 45°V, 70°D.

3.1. Synthetic Experiments. Appropriate synthetic sequences were specifically created for this work. In Figure 5, the left image is a synthetic color image, middle image is the corresponding depth image, and right image is the output of our system. Point cloud is sparsely reconstructed from the model, and the coordinate axis at the center of the model is the current pose of the target. We set three kinds of translate sequences.

For each synthetic scene, we compare the estimated and ground truth trajectories of the aircraft by computing the root-mean-square errors (RMSE). Results on synthetic sequences are shown in Table 1 and Figure 6. We also make a comparison with Co-fusion [3] in translate sequences. Co-fusion [3] only supports translation output; as Table 1 shows, rotation result of Co-fusion [3] is not given. Co-fusion [3] failed in sequences *rotate1*, *rotate2*, *rotate3*, and *translate&rotate*.

As shown in Figures 6–8, the estimated trajectories of the proposed method fit well with ground truth in all scenes. In rotate sequences, the proposed method can also track the object stably. As we can see from Table 1, the proposed method performs better than Co-fusion [3]. In some scenes, Co-fusion [3] may fail to track the object which leads to a huge estimating error. But our method can achieve long-term and effective tracking so long as the first frame is provided to initialize. We carried out a quantitative evaluation on both synthetic and real sequences with ground truth data.

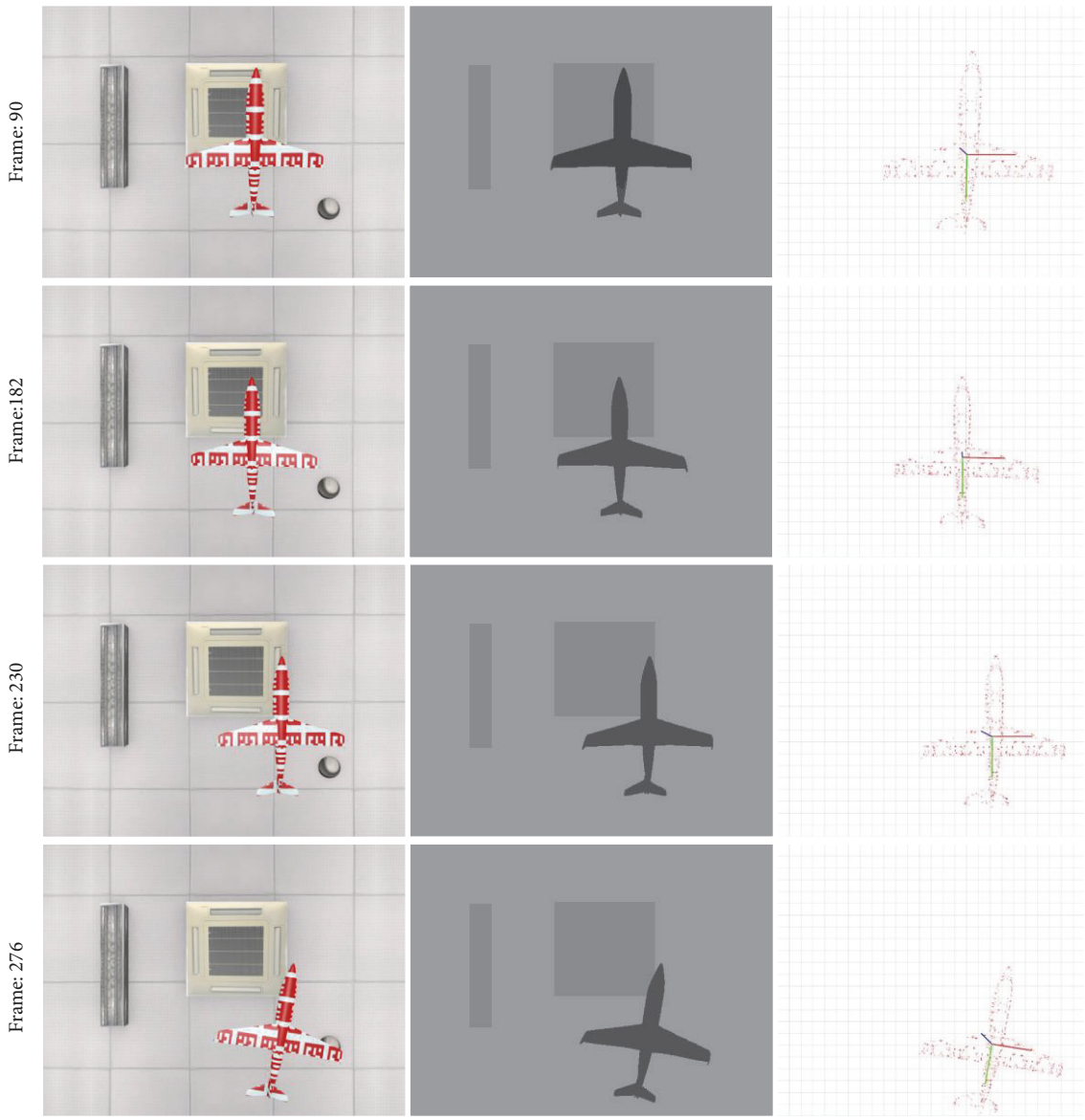
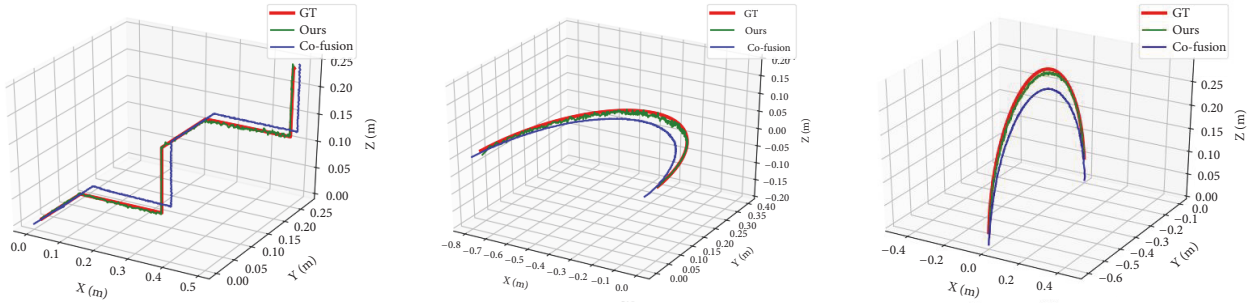
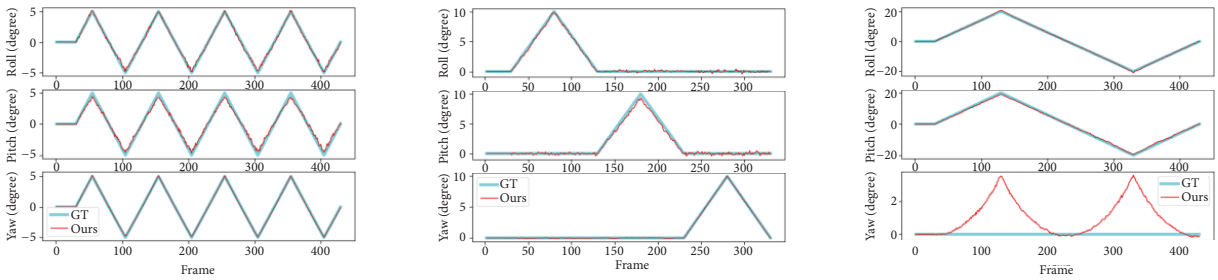


Figure 5: The synthetic RGB-D input and the estimated point cloud result of our system.



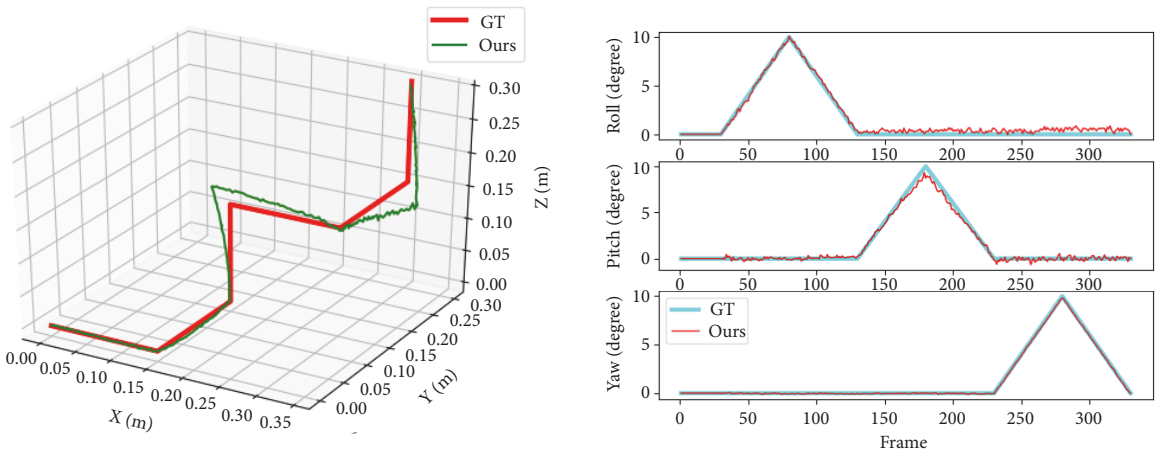
(a) translate1 (b) translate2 (c) translate3

Figure 6: Comparison between the ground truth, our estimated trajectory, and Co-fusion's estimated trajectory on the translate sequence.



(a) rotate1 (b) rotate2 (c) rotate3

Figure 7: Comparison between the ground truth and our estimated pose on rotate sequences.



(a) Trajectory (b) Pose

Figure 8: Comparison between the ground truth and our result on translate&rotate sequences.

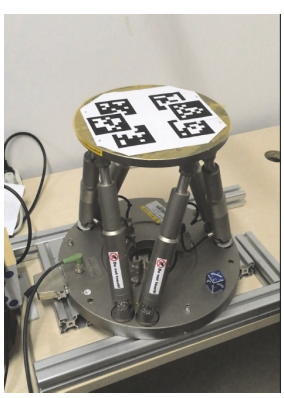


Figure 9: H-840 6-Axis Hexapod. An ultra-high precision cloud terrace with flexible control.

3.2. Experimental Verification with Hexapod and Real Scene. For real sequences, we set a series of experiments on a high precision Hexapod as Figure 9 shows; the accuracy of Hexapod can reach micron level (Table 2). For each axis, a corresponding experiment is set up. After aligning cameras with cloud terrace, we separately set the platform to move uniformly along each axis to test the accuracy of translation and rotation. Results on real sequences are shown in Figure 10.

Experiments on the high precision cloud terrace show that our proposed method also performs well in real sequences. In rotation experiments, the detection of yaw angle is the most accurate which illustrates that the proposed method can reach the highest accuracy without change of depth.

We have performed a series of qualitative experiments to demonstrate the capabilities of our method. The comparison to Co-fusion [3] indicates that our method achieves extremely high accuracy. As mentioned earlier in our paper, our method can adapt to a complex lighting environment and achieve high precision tracking and pose estimation. And particularly, it can satisfy the need of cross FOV, which means it can achieve reliable pose estimation in a wide range of environments as demonstrated in the experiments. The real

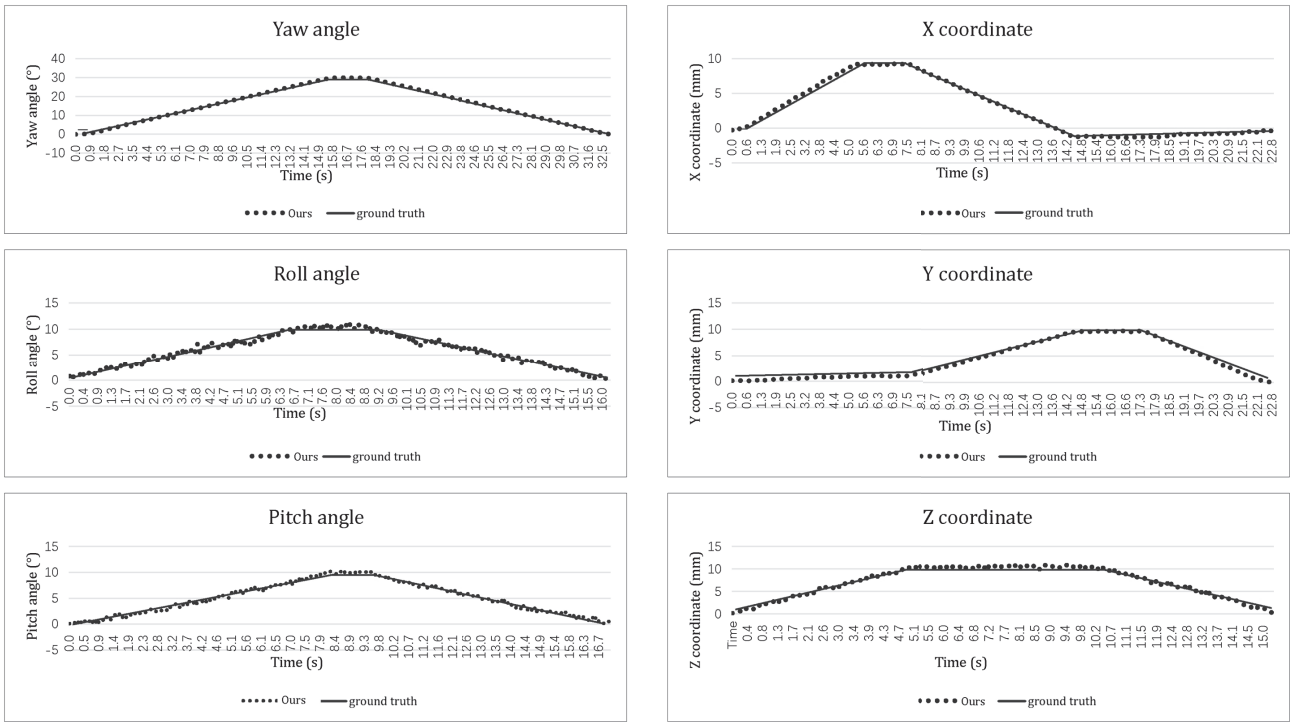


Figure 10: H-840 6-Axis Hexapod. An ultra-high precision cloud terrace with flexible control.

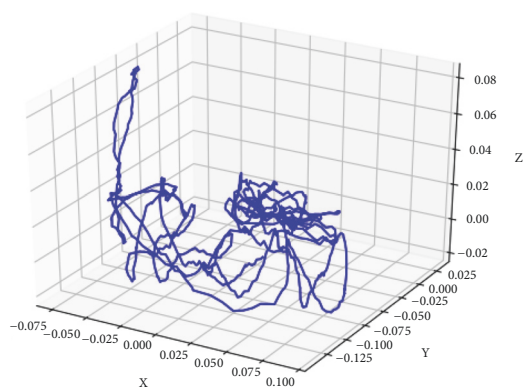


Figure 11: The estimated trajectory of the real scene sequence; the unit in the figure is m.

scene experiment is shown in Figure 12, and the estimated trajectory of the sequence is shown in Figure 11; the trajectory is accurate and concise.

4. Conclusions

We introduced a cross-field of view (FOV) real-time pose estimation system which provides high precision pose estimation of the free flight aircraft in a wind tunnel environment. Multiview live RGB-D streams are used in the system as input to ensure the measurement area can be fully covered. First, a multimodal initialization method is developed to measure the spatial relationship between the RGB-D camera and the aircraft. Based on all the input multimodal information, a so-called cross-FOV model is proposed to recognize the dominating sensor and accurately extract the foreground region in an automatic manner. Second, we develop an RGB-D-based pose estimating method for a single target by which the 3D sparse points and the pose of the target can be simultaneously reconstructed in real time. Many experiments have been conducted and an RGB-D image simulation based on 3D modeling is implemented to verify the effectiveness of our algorithm. Both the real scene's and simulation scene's

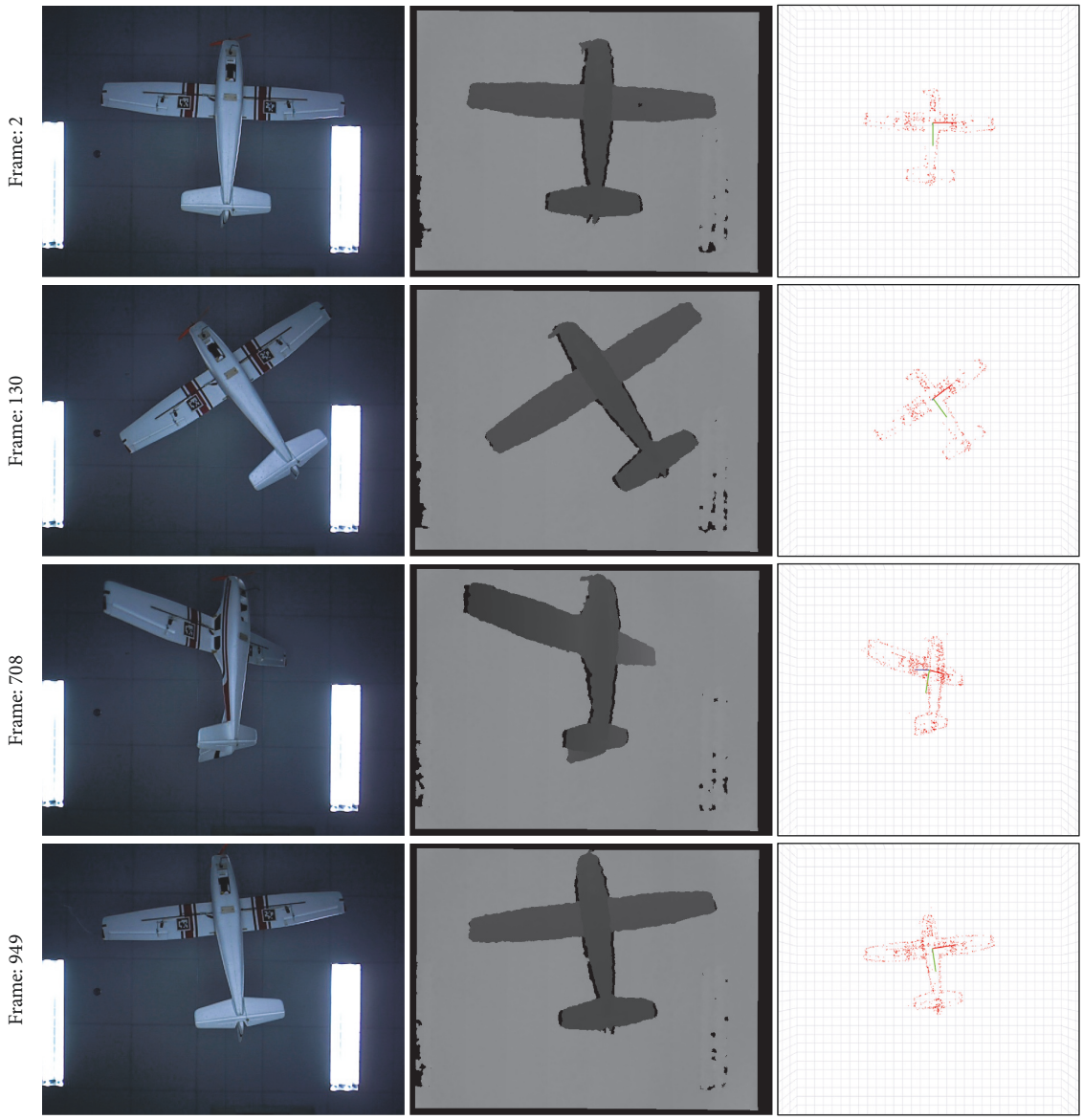


Figure 12: The real scene RGB-D input and the estimated point cloud result of our system.

experimental results demonstrate the effectiveness of our method.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the Zhejiang Provincial Natural Science Foundation of China under Grant numbers LY15F020031 and LQ16F030007 and the National Natural Science Foundation of China (NSFC) under Grant numbers 11302195 and 61401397.

References

[1] D. Dusha, W. Boles, and R. Walker, "Attitude estimation for a fixed-wing aircraft using horizon detection and optical flow," in *Proceedings of the Australian Pattern Recognition Society (APRS)*, pp. 485–492, Australia, December 2007.

[2] P. Bauer and J. Bokor, "Multi-Mode Extended Kalman Filter for Aircraft Attitude Estimation," *IFAC Proceedings Volumes*, vol. 44, no. 1, pp. 7244–7249, 2011.

[3] M. Runz and L. Agapito, "Co-fusion: Real-time segmentation, tracking and fusion of multiple objects," in *Proceedings of the 2017 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 4471–4478, Singapore, Singapore, May 2017.

[4] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos, "ORB-SLAM: a versatile and accurate monocular SLAM system," *IEEE Transactions on Robotics*, vol. 31, no. 5, pp. 1147–1163, 2015.

[5] J. Wang and E. Olson, "AprilTag 2: Efficient and robust fiducial detection," in *Proceedings of the 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2016*, pp. 4193–4198, Republic of Korea, October 2016.

[6] P. Rudol, M. Wzorek, and P. Doherty, "Vision-based pose estimation for autonomous indoor navigation of micro-scale unmanned aircraft systems," in *Proceedings of the 2010 IEEE International Conference on Robotics and Automation, ICRA 2010*, pp. 1913–1920, USA, May 2010.

[7] M. G. Breuers, F. A. Sadjadi, and N. de Reus, "Image-based aircraft pose estimation: a comparison of simulations and real-world data," in *Proceedings of the Aerospace/Defense Sensing, Simulation, and Controls*, p. 472, Orlando, FL.

[8] S. Chen, Y. Li, and N. M. Kwok, "Active vision in robotic systems: a survey of recent developments," *International Journal of Robotics Research*, vol. 30, no. 11, pp. 1343–1377, 2011.

[9] S. Chen, J. Zhang, Y. Li, and J. Zhang, "A hierarchical model incorporating segmented regions and pixel descriptors for video background subtraction," *IEEE Transactions on Industrial Informatics*, vol. 8, no. 1, pp. 118–127, 2012.

[10] C. Shengyong, *Active Sensor Planning for Multiview Vision Tasks*, Springer Science & Business Media, 2008.

[11] H. Liu, S. Chen, and N. Kubota, "Intelligent video systems and analytics: a survey," *IEEE Transactions on Industrial Informatics*, vol. 9, no. 3, pp. 1222–1233, 2013.

[12] L. Liu, Y. Wang, F. Xie, and J. Gao, "Legendre Cooperative PSO Strategies for Trajectory Optimization," *Complexity*, vol. 2018, Article ID 5036791, 13 pages, 2018.

[13] M. Wang, H. Ye, and Z. Chen, "Neural Learning Control of Flexible Joint Manipulator with Predefined Tracking Performance and Application to Baxter Robot," *Complexity*, vol. 2017, Article ID 7683785, 14 pages, 2017.

[14] M. Zollhöfer, M. Nießner, S. Izadi et al., "Real-time non-rigid reconstruction using an RGB-D camera," *ACM Transactions on Graphics*, vol. 33, no. 4, 2014.

[15] D. Tzionas and J. Gall, "Reconstructing Articulated Rigged Models from RGB-D Videos," in *Computer Vision – ECCV 2016 Workshops. ECCV 2016*, G. Hua and H. Jégou, Eds., vol. 9915, Springer, Cham, 2016, Lecture Notes in Computer Science.

Advances in Operations Research
Advances in Decision Sciences
Journal of Applied Mathematics
The Scientific World Journal
Journal of Probability and Statistics
International Journal of Mathematics and Mathematical Sciences
Hindawi
Submit your manuscripts at
www.hindawi.com
Journal of Optimization
International Journal of Engineering Mathematics
International Journal of Analysis
Journal of Complex Analysis
Advances in Numerical Analysis
Mathematical Problems in Engineering
International Journal of Differential Equations
Discrete Dynamics in Nature and Society
International Journal of Stochastic Analysis
Journal of Mathematics
Journal of Function Spaces
Abstract and Applied Analysis
Advances in Mathematical Physics