



A DNN-Based Optical Aided Autonomous Navigation System for UAV Under GNSS-denied Environment

Qiang Zhang, Hua jun Zhang^(✉), Zining Lan, Wenxin Chen, and Zilong Zhang

Hubei Aerospace Vehicle Research Institute, Wuhan, Hubei 430010, China
dts.zhj@163.com

Abstract. In this paper, a novel UAV autonomous navigation system is proposed, which uses DNN to extract scene features for position matching and optical flow estimation. This method can correct position drift of IMU under GNSS-denied environment. A realistic simulation test platform is designed to validate the fusion navigation system, and we deploy the system on an airborne embedded GPU hardware. Optimization strategies for hardware are designed to significantly improve operational performance, which make it become more practical in engineering. Field experimental results further verify its effectiveness of proposed system. Compared with traditional methods, it is more robust and has higher positioning accuracy.

Keywords: Autonomous navigation · Deep neural network · GNSS-denied environment · Unmanned aerial vehicle

1 Introduction

In urban combat scene, UAV (Unmanned Aerial Vehicle) systems face complex electromagnetic interferences. Seeking continuous and stable GNSS (Global Navigation Satellite System) signal support is a luxury. Once disturbed, UAVs will not be able to accurately obtain flight position information, resulting in forced landing, falling and other accidents. Therefore, developing a reliable autonomous navigation and positioning method under GNSS-denied environment is worthwhile. Many studies have been carried out, such as terrain, geomagnetism, gravity, polarized light navigation, etc. [1–8]. Among them, image matching navigation technology is more practical in engineering, which has good anti-interference ability and low cost [9–14]. Because almost all drones wear optical reconnaissance equipment, and no additional hardware device is required. Besides, commercial remote sensing satellite maps with high resolution and high accuracy are very easy to obtain. Nowadays, people can get rich topographic and geomorphic dataset from any region for only a small fee. However, aerial image matching and positioning has some shortcomings. On one hand, environmental features, illumination changes and spectral features will affect the positioning accuracy. On the other hand, under high dynamic flight environment, efficient real-time output and the limited computing resources of platform form a mutually restrictive relationship. How to make trade-offs deserves deep consideration.

In recent years, DNN (Deep Neural Network) based feature extraction and matching technology has gradually occupied an important position in the field of computer vision. Compared with traditional feature extraction methods, it has stronger adaptability to changes in environment and light [15–18]. Under different vision angle, different seasonal conditions, even different receiving spectrum, some DNN-based methods still can accurately match the correct base map block. This high level feature abstraction ability is very suitable for position estimation. If selecting remote sensing map as base, we can get matched absolute position coordinates. Similarly, if selecting previous taken photo as base, we get matched pixel error, i.e. optical flow field, which can be converted into speed information further [19–22]. Combining these two kinds of information, we can obtain the location information of platform in real time. Based on the above idea, an optical aided autonomous navigation method is proposed. This method can effectively correct position drift of IMU under GNSS-denied environment. A simulation test bed is built to validate the fusion navigation system, and we also carry out performance optimization and evaluation on an airborne embedded GPU hardware. Field flight tests further verify the feasibility of proposed method. Compared with traditional methods, like SIFT and ORB based algorithms, our proposed method has smaller positioning deviation and stronger robustness.

2 Algorithm Design

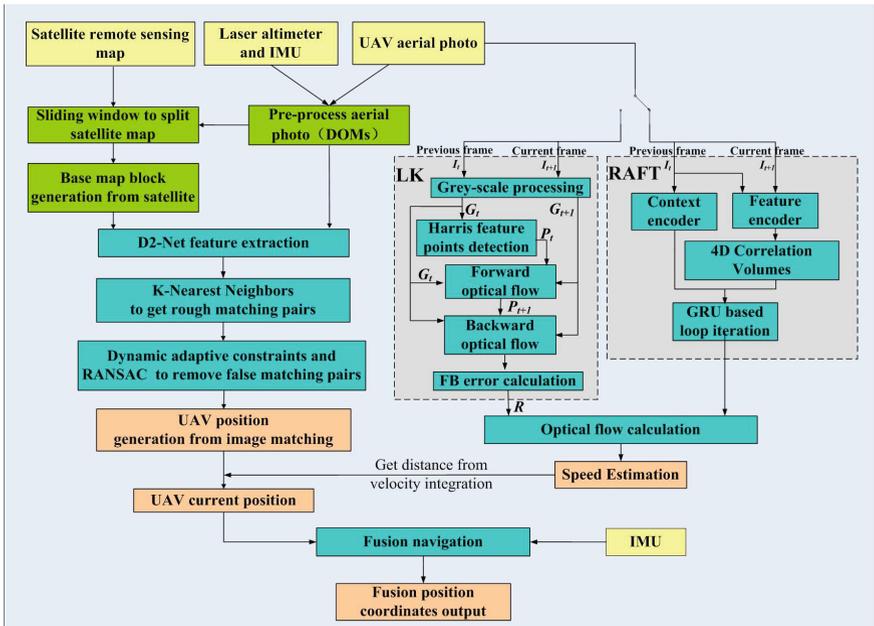


Fig. 1. Schematic diagram of autonomous navigation and positioning system

The theoretical core consists of DNN-based image matching location algorithm and speed estimation algorithm. As shown in Fig. 1, the technical implementation can be divided into following parts: (1) According to actual mission needs, the high-resolution remote sensing satellite base map blocks of a certain area are stored and loaded in advance, and meanwhile, use IMU attitude matrix transformation to get near real-time DOMs (Digital Orthoimage) of UAV. (2) Match DOM with base map block to obtain its longitude and latitude, which will be further converted into coordinates of UAV using trigonometric geometry conversion [23]. D2-net is used as backbone of feature extraction [15]. Considering its high computation complexity in embedded hardware, it usually completes a round of iterative calculation in seconds. (3) According to two consecutive frames of DOMs, the real-time velocity of UAV is obtained by optical flow algorithm, and relative displacement can be obtained by velocity integrating; (3) Superimpose delayed matching position and relative displacement to obtain real-time position of UAV; (4) Using real-time absolute coordinates to correct drift error of IMU, so as to realize the autonomous navigation and positioning function of UAV in GNSS-denied environment.

For matching location part, there are two key factors affecting the position accuracy. One is the feature extraction method, the other is the pre-processing strategies on the UAV image. First, complete DOM transformation. Considering UAV altitude, the pixel coordinates (x_g, y_g) of orthophotoed picture can be transformed from that of raw picture:

$$(x_g, y_g) = \left(f \frac{r_{11}x + r_{12}y + r_{13}f}{r_{31}x + r_{32}y + r_{33}f}, f \frac{r_{21}x + r_{22}y + r_{23}f}{r_{31}x + r_{32}y + r_{33}f} \right) \quad (1)$$

where f represents camera focal length and r_{ij} is obtained from transformation matrix R^{-1} which is determined by UAV real-time attitude. Assuming (ω, θ, ϕ) indicate pitch, yaw and roll angle respectively, the matrix R can be denoted as:

$$R = \begin{bmatrix} \cos\omega\cos\theta & \cos\omega\sin\theta & -\sin\omega \\ \sin\phi\sin\omega\cos\theta - \cos\phi\sin\theta & \sin\phi\sin\omega\sin\theta + \cos\phi\cos\theta & \sin\phi\cos\omega \\ \cos\phi\sin\omega\cos\theta + \sin\phi\sin\theta & \cos\phi\sin\omega\sin\theta - \sin\phi\cos\theta & \cos\phi\cos\omega \end{bmatrix} \quad (2)$$

Then, resize DOMs using camera field angle, altitude information to maintain consistent resolution with base block. Feature extraction module adopts D2-net as backbone to get key points and descriptors, i.e., feature vectors. Compared with classical methods, such as SIFT and HOG [24, 25], it has better adaptability when existing obvious differences between matched picture and the base block in light, season, spectrum. Pre-processing module achieve two function.

D2-net outputs the n feature vectors of aerial photo and m feature vectors of base block at the same time, which can be respectively represented as:

$$G_{\text{uav}} = (\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_n), \quad G_{\text{base}} = (\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_m) \quad (3)$$

For each vector \mathbf{p}_i , there are m matching pairs with \mathbf{q}_i . Euclidean distance of matching pairs can be denoted as $(dis_{i:1}, dis_{i:2}, \dots, dis_{i:m})$. Using k-nearest

neighbor search algorithm to roughly filter out matching point pairs whose distance are too far away, where $K = 2$. Then, choose 2 nearest distance value dis_i and dis'_i from $(dis_{i:1}, dis_{i:2}, \dots, dis_{i:m})$ to calculate the distance deviation, where $dis'_i > dis_i$. For G_{uav} , average distance deviation of all feature vectors can be denoted as:

$$dis_{avg} = \left[\sum_{i=1}^n |dis_i - dis'_i| \right] / n \quad (4)$$

If $dis_i < dis'_i - dis_{avg}$ is satisfied, the corresponding matching points pair is reserved, otherwise, it will be rejected. Then classic RANSAC (Random Sample Consensus) is applied to further delete redundant mismatching pairs [26]. Select base map block with the most matching pairs to calculate homography matrix which is used for matching and positioning [23]. Sliding window overlap rate and maximum iterations of RANSAC are two hyperparameters which can affect matching accuracy. After screening test, they are set to 80% and 100 respectively. For speed estimation part, both DNN-based RAFT (Recurrent All-Pairs Field Transforms) and classic LK (Lucas-Kanade) are verified [20, 27]. In order to keep the resolution of adjacent pictures consistent, the flight altitude change needs to be taken into account. Assuming that UAV adopts ‘‘North East earth’’ coordinate system, connect camera and UAV body as a rigid body, which therefore, X, Y and Z axes of the camera coordinate system will coincide with the that of the body coordinate system respectively. For the same feature point located at p and p' in two consecutive images respectively, the optical flow can be represented as:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} p_u - p'_u \\ p_v - p'_v \end{bmatrix} \quad (5)$$

where $[u, v]^T$ is optical flow in two vertical directions between time T and $T + 1$. Take the derivative of $[u, v]^T$ and get pixel speed:

$$\begin{bmatrix} v_{fow-u} \\ v_{fow-v} \end{bmatrix} = \begin{bmatrix} \dot{u} \\ \dot{v} \end{bmatrix} \quad (6)$$

Further considering imaging model of camera, the velocity of UAV body can be obtained through the following transformation formula:

$$\begin{bmatrix} v_x \\ v_y \end{bmatrix} = \begin{bmatrix} -h/f & 0 \\ 0 & -h/f \end{bmatrix} \begin{bmatrix} v_{fow-u} \\ v_{fow-v} \end{bmatrix} + \begin{bmatrix} \omega_y f \\ -\omega_x f \end{bmatrix} \quad (7)$$

where h represents flight altitude and (ω_x, ω_y) is angular velocity of UAV body.

3 Simulation Platform Design and Validation

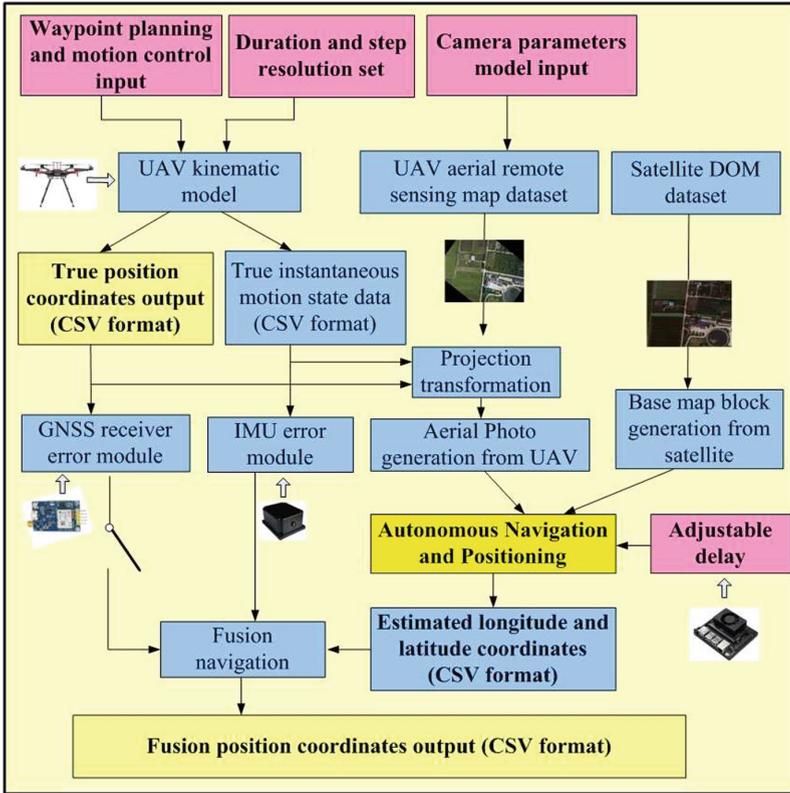


Fig. 2. Schematic diagram of simulation platform

As shown in Fig. 2, a realistic simulation test platform is designed for carrying out various ablation trials, which can help screen out optimal hyperparameters. Waypoint planning and motion control input module defines a rough flight path. After being processed by UAV motion model, it is further converted into a fine motion trajectory, which can output real-time position and motion status, such as attitude, velocity vector, GNSS coordinates, etc. according to set duration and step resolution. On one hand, these true datastream will be transferred to error model of IMU and GNSS receiver. On the other hand, UAV aerial dataset will use them to generate real continuous aerial photos which superimpose flight state disturbance and camera parameter limits. UAV aerial dataset is obtained from open source GIS website [28], which has 0.075metre/pixel resolution. Satellite remote sensing dataset is obtained from commercial GIS software: BIGEMAP, which has 0.51 m/pixel resolution. Autonomous navigation algorithm receives simulated UAV photo and satellite map block to output estimated GNSS position which will replace output of GNSS receiver under denial environment. Simulation

platform runs on a workstation which equips with 2080Ti graphics card and 32GB memory. In order to simulate the embedded system environment more realistically, an adjustable delay is added into matching positioning module.

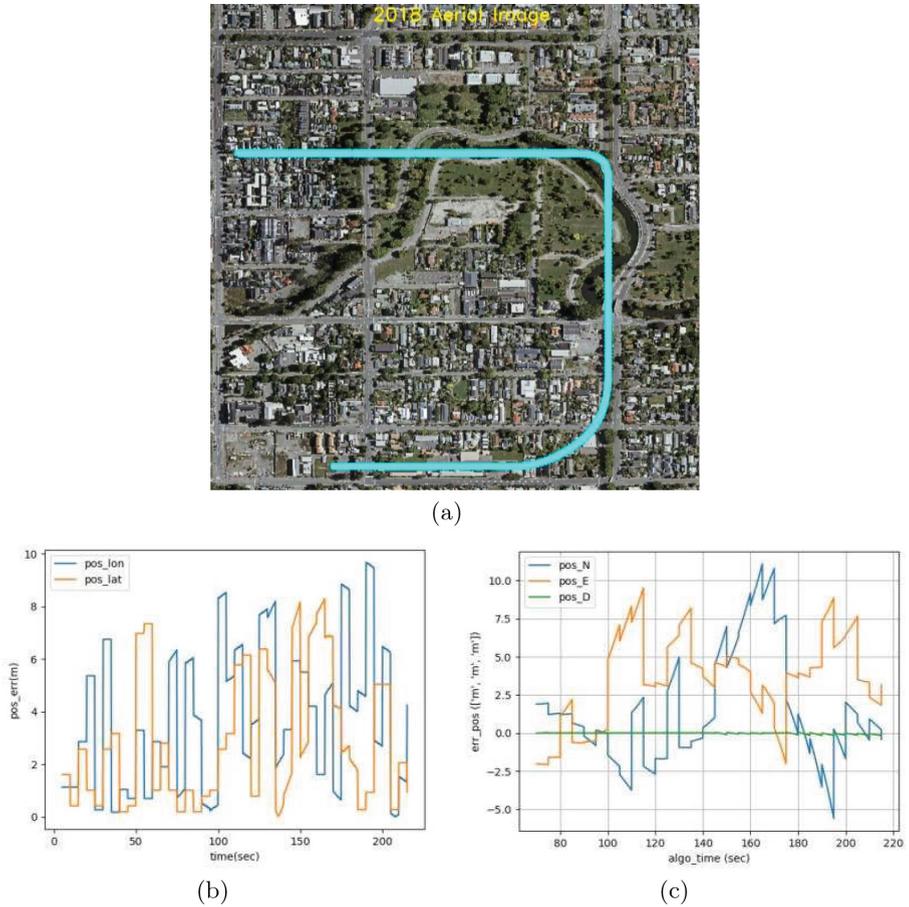


Fig. 3. Simulation example. (a)Planned waypoints. (b) Error outputs by image matching & optical flow. (c) Error outputs by image matching & optical flow & IMU.

Based on the above simulation platform, test results are shown in Fig. 3. Simulation conditions are as follows: flight altitude is set to 200 m, flight duration is no more than 220s, max speed limit is 20m/s. Considering the embedded hardware output delay in real scene, adjustable delay is set to 5s, which represents the total time of one correct image matching positioning. The planned flight path is shown is Fig. 3(a). Figure 3(b) represents the error curve between the estimated position coordinates and true GNSS positions, which has no IMU

involvement. Figure 3(c) represents the error curve between IMU-fused position coordinates and true GNSS positions, which is relatively smoother and more continuous.

4 Hardware Deployment and Evaluation

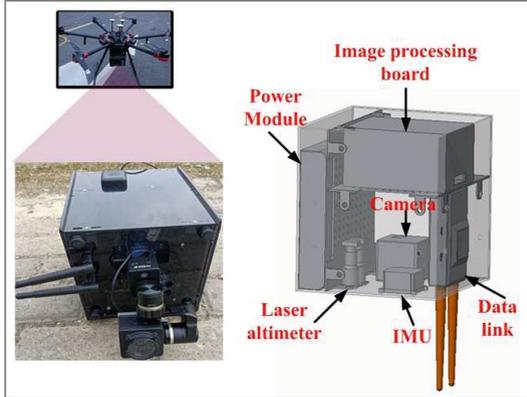


Fig. 4. Embedded hardware platform

As shown in Fig. 4, we deploy the software into an airborne embedded hardware platform, which consists of image processing board, data link module, camera, IMU, laser altimeter and power module. The core chip of image processing board is based on Nvidia Xavier SoC, which has 21 TOPS for AI computation. In order to find performance bottlenecks and accelerate hardware implementation. At first, non-optimized software is directly deployed and runs on embedded environment, which consumes 334.64s to complete 247 sliding base blocks searchings and matching computations. In a single matching processing, feature extraction module takes the longest time, about 1.103s. KNN(K-nearest neighbor) research and RANSAC module takes the second longest time, about 0.084s. The low computational efficiency is unacceptable in real scene. Therefore, several optimization strategies are applied:

- (1) Transform PyTorch model of D2-net into TensorRT model to accelerate feature extraction processing, which can make full use of embedded DLAs (deep learning accelerators).
- (2) All the features of base map blocks are extracted in advance and reloaded into cache before running.
- (3) Discard the base map blocks with inconsistent sizes when sliding, which will reduce the number of matches by 40%.

After the above optimization, total time cost is reduced to 5.5s, which becomes more practical in engineering. Detailed running time statistics on embedded hardware platform are listed in Table 1

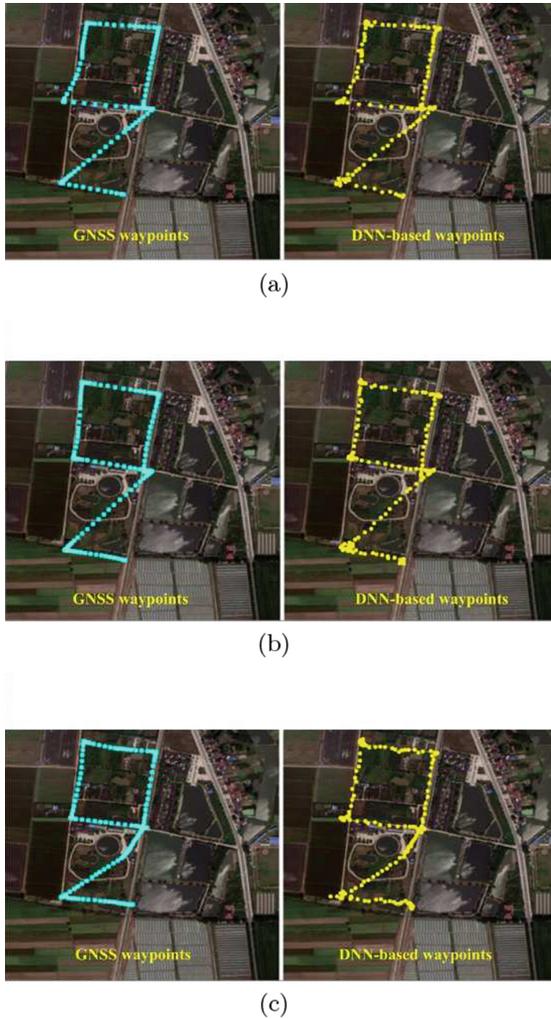
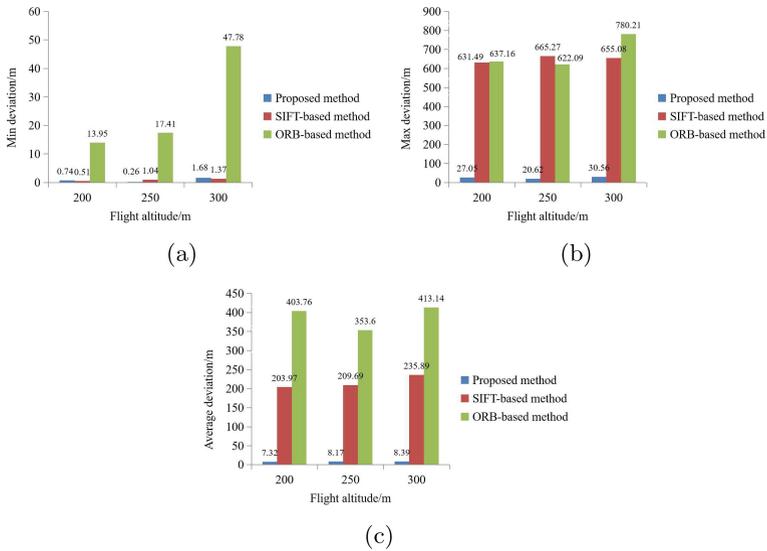


Fig. 6. Field experiment. (a) Waypoints at 200 m height. (b) Waypoints at 250 m height. (c) Waypoints at 300 m height.

In order to further verify the superiority of proposed method, we replace the DNN-based backbone with traditional algorithm, such as SIFT and ORB, which is used to compare their positioning accuracy and robustness. As shown in Fig. 7, SIFT-based method acquires an average deviation of about 200 m and ORB-based method has an average deviation of about 400 m. Besides, their variances of the minimum and maximum deviation are too large, which is poor in engineering feasibility. Conversely, DNN-based method is more robust and its location deviation is stably distributed in a smaller interval, which verifies its stronger environmental adaptability.

Table 2. Position deviation statistics of GNSS and DNN-based system

Flight altitude (unit:meter)	Amount of photos	Average deviation (unit:meter)	Max deviation (unit:meter)	Min deviation (unit:meter)
200	160	7.32	27.05	0.74
250	171	8.17	20.62	0.26
300	199	8.39	30.56	1.68

**Fig. 7.** Accuracy comparison. (a) Minimum deviation. (b) Maximum deviation. (c) Average deviation.

6 Conclusion

In this paper, an optical aided UAV autonomous navigation system is proposed, which uses DNN-based module to extract scene features for position matching and optical flow estimation. The image matching algorithm intermittently obtains absolute position information of platform, and the optical flow estimation algorithm obtains the speed information and integrates it to improve the position update rate in real time. This approach can effectively correct position drift of IMU under GNSS-denied environment. A realistic simulation test system is designed to validate the fusion navigation system. Compared with classic matching location method, it is more adaptable to the environment, light and seasonal changes. We also carry out performance optimization and evaluation on an airborne embedded GPU hardware, which makes it more practical in engineering. Field experimental results show that, compared with SIFT and ORB based methods, DNN-based method is more robust and its location deviation is

stably distributed in a smaller interval, which verifies its stronger environmental adaptability.

Acknowledgement. This work is supported by the Key R&D Program of Hubei Province under grant 2020BAB024. The authors would also like to thank the reviewers for many helpful comments and suggestions, which have enhanced the quality and readability of this paper.

References

1. Jiang, W, Li, Q., Fei, W.: A novel location method based on 2D laser scanning sensor terrain matching for UAV autonomous flight. In: IET International Conference on Information Science & Control Engineering, IET (2014)
2. Long, Z., Nan, G., Huang, B., et al.: A novel terrain-aided navigation algorithm combined with the TERCOM algorithm and particle filter.: IEEE Sensors J. 1124–1131 (2014)
3. Lu, Q.H., Ying, H., Huang, X.L.: Polarized-Light/Geomagnetism/GPS/SINS Integrated Navigatio. J. Astronaut. 897–902 (2007)
4. Cheng, J.J., Xiong, Z., Jian-Xin, X.U., et al.: Study of geomagnetic matching aided seamless INS/GPS integrated navigation. Aeronaut. Comput. Technique (2014)
5. Ying, L., Wu, M., Xie, H.: Iterative multi-level magnetic matching for UAV navigation. In: The 2010 IEEE International Conference on Information and Automation (2010)
6. Sambo, A.U., Li, S.X.: Feature selection criterion for gravity matching navigation. Int. J. Sci. Res. 2319–7064 (2016)
7. Liu B, Fan Z, Wang X.: Solar position acquisition method for polarized light navigation based on ∞ characteristic model of polarized skylight pattern. IEEE Access, 99 (2020)
8. Xia, L., Liu, R., Zhang, D., et al.: Polarized light-aided visual-inertial navigation system: global heading measurements and graph optimization-based multi-sensor fusion. Measur. Sci. Technol. **33**(5), 55111 (2022)
9. Ji, X., Han, J.W., Liang, N., et al.: UAV positioning simulation method based on scene matching. J. Syst. Simul. **19**, 2775 (2014)
10. Zheng, M., Wu, C., Chen, D., et al.: Rotation and affine-invariant SIFT descriptor for matching UAV images with satellite images. In: Proceedings of 2014 IEEE Chinese Guidance, Navigation and Control Conference. IEEE (2015)
11. Braga, J., Velho, H., Conte, G., et al.: An image matching system for autonomous UAV navigation based on neural network. In: International Conference on Control. IEEE (2016)
12. Nassar, A., Amer, K., Elhakim, R., et al.: A deep CNN-based framework for enhanced aerial imagery registration with applications to UAV Geolocalization. In: IEEE/CVF Conference on Computer Vision & Pattern Recognition Workshops. IEEE (2018)
13. Chen, S., Chen, H., Chang, C.W., et al.: Multilayer mapping kit for autonomous UAV navigation. IEEE Access **PP**(99), 1–1 (2021)
14. Mughal M H, Khokhar M J, Shahzad M.: Assisting UAV localization via deep contextual image matching. IEEE J. Sel. Topics Appl. Earth Obser. Remote Sens. **PP**, 99 (2021)

15. Dusmanu, M., Rocco, I., Pajdla, T., et al.: D2-Net: a trainable CNN for joint description and detection of local features. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE (2019)
16. Sarlin, P.E., Detone, D., Malisiewicz, T., et al.: SuperGlue: learning feature matching with graph neural networks. arXiv (2019)
17. Pautrat, R., Lin, J.T., Larsson, V., et al.: SOLD2: self-supervised occlusion-aware line description and detection (2021)
18. Sarlin, P.E., Unagar, A., Larsson, M., et al.: Back to the feature: learning robust camera localization from pixels to pose (2021)
19. Fischer, P., Dosovitskiy, A., Ilg, E., et al.: FlowNet: learning optical flow with convolutional networks. In: 2015 IEEE International Conference on Computer Vision (ICCV). IEEE (2016)
20. Teed Z, Deng J.: RAFT: recurrent all-pairs field transforms for optical flow (2020)
21. Efe, U., Ince, K.G., Alatan, A.A.: DFM: a performance baseline for deep feature matching. IEEE (2021)
22. Zheng, Z., Nie, N., Ling, Z., et al.: DIP: deep inverse patchmatch for high-resolution optical flow (2022)
23. Lin, Z.G., Li, Y.J., Pan, Q., et al.: Scene matching aided navigation technology for UAV. Northwestern Polytechnic University Press, Xi'an (2016)
24. August. TPA 11: Comparative study of various feature extractors for matching/recognition, under degradation (SIFT, SURF, HOG, Top points). (2005)
25. Wang, Y., Zhao, R., Liang, L., et al.: Block-based image matching for image retrieval. *J. Visual Commun. Image Represent.* **74**(6), 102998. (2022)
26. Fischler, M.A., Bolles, R.C.: Random sample consensus. *Commun. ACM* (1981)
27. Yedjour, H.: Optical Flow Based on Lucas-Kanade Method for Motion Estimation. In: Hatti, M. (eds.) *Artificial Intelligence and Renewables Towards an Energy Transition. ICAIRES 2020. Lecture Notes in Networks and Systems*, vol. 174. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-63846-7_92
28. <https://data.linz.govt.nz/layer/99193-christchurch-0075m-urban-aerial-photos-2018> (2018)