# Visual SLAM of Unmanned Aerial Vehicles: A Survey

CMVIT2022
Paper C018

*Yikun Tian[1], Binchao Yang[2], Hong Yue[1]*, Jinchang Ren[3]*

[1]Dept. Electronic and Electrical Engineering, University of Strathclyde, Glasgow, G1 1XW, UK (*hong.yue@strath.ac.uk)
[2]College of Electrical and Power Engineering, Taiyuan University of Technology, Taiyuan 030024, China
[3]National Subsea Centre, Robert Gordon University, Aberdeen AB21 0BH, UK

## Introduction

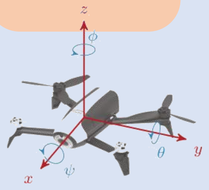
- **Simultaneous Localization and Mapping (SLAM)** aims to realize environmental perception and understanding in an unfamiliar environment to complete self-positioning and path planning of robotics [1]. Localization and mapping are the basic needs of humans and mobile devices, where humans can perceive their movements and the environments through multimodal sensing, relying on the awareness of the location to navigate in a complex three-dimensional space.
- A complete SLAM system consists of four parts: (i) the front-end tracking, (ii) the back-end optimization, (iii) the loop detection, and (iv) the map reconstruction, where visual odometry is one of the challenging and open topics in the vSLAM system for determining the position and orientation of robots by analyzing the captured images from the associated cameras.
- There are a huge number of applications with various sensing equipment, single or binocular cameras based on SLAM. Benefiting from new visual-sensing equipment, powerful data processing and high flexibility, SLAM can now be implemented in a simpler and low cost system structure [2].

## Conventional vSLAM

- Traditional vSLAM solutions can be classified into feature point based methods and direct methods. The former is mainly composed of feature detector, descriptor, and motion estimation. The feature detector can be divided into the point-feature detector and the blob detector. The motion estimation is divided into the 3D-to-3D method, the 3D-to-2D method and the 2D-to-2D method. On the contrary, the direct method needs not to extract feature points, which can directly restore the camera's posture and map the structure through the photometric error, without calculating feature points and descriptors.

**Three steps in feature point method**

- Extract stable characteristic points from each picture frame. These features are usually immutable descriptions. The matching of adjacent frames is done through the descriptor.
- Restore the camera pose and the coordinates of the mapping point through the epipolar geometry.
- Fine-tune the camera pose and map the structure by minimizing the projection error. Feature points extracted from each frame are cyclically detected or relocated through operations such as clustering.



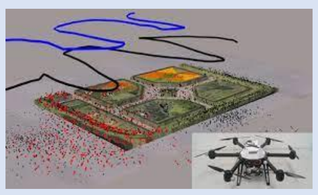

**Analysis of limitations**

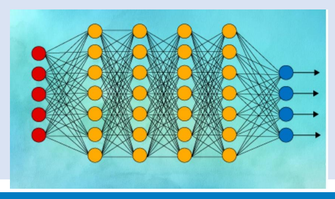
- The extraction of feature points and the computation of descriptors are very time-consuming.
- All other information is ignored except for the feature points.
- The camera sometimes moves to places without enough features.

The traditional vSLAM has the following limitations: (i) Under adverse conditions e.g. poor lighting conditions or large changes of lighting, the robustness is not good. (ii) In the case of large movement or rapid shaking, they are prone to unmatched feature points. (iii) They cannot identify the foreground objects, that is, moving objects in the scene can only be considered as "bad pixels".

## Data-Driven Visual Odometry

- Combining deep learning and vSLAM overcomes the limitations of hand-crafted algorithms in visual odometry and scene recognition [3], which improves the learning ability and intelligence of robotic platforms. Based on multi-layer neural networks to learn hierarchical feature representations and automatically discover tasks-related features, deep learning algorithms are more in line with the laws of human cognition and the interaction of the environment [4].





## UAV Image Stabilization

- Videos captured from unmanned aerial vehicles (UAV) is often susceptible to unexpected sensor movement, which can severely interfere with the detection and tracking of the targets of interest. Before 2010, there were two main video stabilization methods, i.e. the mechanical video stabilization (MVS) and the digital video stabilization (DVS). MVS stabilizes video frames via a special equipment, it cannot remove all video vibration, along with a high hardware cost. On the contrary, DVS stabilizes video frames by image processing techniques, which is effective with a relatively low cost hence is preferred.
- DVS consists of three stages.

**Global motion estimation**

Global motion estimation (GLE), which determines the global motion between two consecutive frames, is achieved by global intensity alignment for feature-based approaches [5]. Compared with global intensity alignment approaches, feature-based methods are generally faster [6].

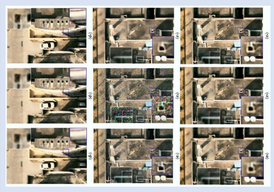
**Motion filtering & compensation**

After motion capture, motion filtering methods, e.g. random sample consensus (RANSAC), least mean square (LMS), and M-estimate of sample consensus (MSAC), are used to filter the outliers to avoid misleading movement. Motion correction then uses motion vector integration, low-pass filtering, Kalman filter, etc. to distinguish intentional and unintentional motions.

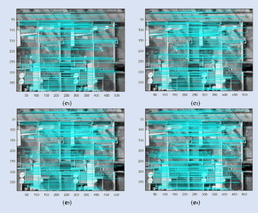
**Video completion**

Finally, frame compensation moves the image frame out of the correction vector from the filtered result. The purpose of video completion is to fill in missing frames in a video.

## UAV Image Denoising



As a low-altitude remote sensing platform, UAV images are affected by lightning, ground electromagnetic waves, light changes, and the UAV mechanical noise. Thus, it is necessary to apply algorithms to remove UAV image noise.



- To reduce speckle noise from SAR images, a clustering-based method is used in [7]. A noisy image is first divided into several disjoint regions, each of which is then denoised by the Wiener filter by minimize a linear mean square error in a linear discriminant analysis domain. A denoised SAR image is eventually aggregately produced by the denoised patches.
- Kim et al. adopted an adaptive noise filtering method based on background registration to deal with unfixed pattern noise in UAV infrared images, which relies on background registration processing and robust principal component analysis [8].
- To improve both the automation level and the (corresponding) denoising effect, the unmixing-based denoising method is improved by adjusting the contribution of each spectral band to the final result as well as its automation degree [1].
- To further improve the efficacy of denoising and feature extraction, a number of new techniques can be further applied, such as sparse learning and genetic feature selection for effective signal detection and classification [9].

## Conclusions & Discussions

- We summarize the research on UAV based path planning using SLAM with environmental perception and understanding. It is worth noting that the videos captured by UAVs is usually susceptible to unexpected sensor movements, which can seriously interfere with the detection and tracking of the targets of interest. Replacing the homography transformation with a deep-learning-based reconstruction method can improve UAV effectiveness of the captured image.
- As a low-altitude remote sensing platform, UAV images are affected by lightning, ground electromagnetic waves, light changes, and the mechanical noise. It is necessary to denoise UAV images in a more efficient way i.e. by accelerating with a GPU.

## References

[1] Cadena C *et al*. 2016 *IEEE Trans. Robot.* 32 1309-32
[2] Sünderhauf N *et al*. *2018 Int J. Rob. Res.* 37 405-20
[3] Ma L *et al*. *2017 Proc. Int. Conf. IROS* (Vancouver, Canada) 598-605
[4] Brendan McMahan H *et al*. *2013 ACM SIGKDD Int. Conf. KDD* (Chicago, USA) 1222-30
[5] Buyukyazi T, Bayraktar S and Lazoglu I, *2013 Proc. RAST* (Istanbul, Turkey) 121-6
[6 ] Yu J *et al*. *2015 Proc. Int. Conf. CCCV* (Berlin, Germany) 121-6
[7] Wang H *et al*. *2016 EURASIP J Adv Signal Process* 2016 10
[8] Kim B, Kim M and Chae Y 2018 *Sensors* 18 1
[9] Padfield N *et al*. 2021 *Neurocomputing* 463 566-70